VOLUME 123, NUMBER 10
MARCH 14, 2001

JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Carbon Monoxide Binding by de Novo Heme Proteins Derived from Designed Combinatorial Libraries

David A. Moffet, Martin A. Case, John C. House, Kathleen Vogel, Robert D. Williams, Thomas G. Spiro, George L. McLendon, and Michael H. Hecht*

Contribution from the Department of Chemistry, Princeton University, Princeton, New Jersey 08544-1009

Received October 5, 2000

Abstract: Carbon monoxide binding was studied in a collection of de novo heme proteins derived from combinatorial libraries of sequences designed to fold into 4-helix bundles. The design of the de novo sequences was based on the previously reported "binary code" strategy, in which the patterning of polar and nonpolar amino acids is specified explicitly, but the exact identities of the side chains are varied extensively.[1] The combinatorial mixture of amino acids included histidine and methionine, which ligate heme iron in natural proteins. However, no attempt was made to explicitly design a heme binding site. Nonetheless, as reported previously, approximately half of the binary code proteins bind heme.[2] This collection of novel heme proteins provides a unique opportunity for an unbiased assessment of the functional potentialities of heme proteins that have not been prejudiced either by explicit design or by evolutionary selection. To assess the capabilities of the de novo heme proteins to bind diatomic ligands, we measured the affinity for CO, the kinetics of CO binding and release, and the resonance Raman spectra of the CO complexes for eight de novo heme proteins from two combinatorial libraries. The CO binding affinities for all eight proteins were similar to that of myoglobin, with dissociation constants (K_d) in the low nanomolar range. The CO association kinetics (k_{on}) revealed that the heme environment in all eight of the de novo proteins is partially buried, and the resonance Raman studies indicated that the local environment around the bound CO is devoid of hydrogen-bonding groups. Overall, the CO binding properties of the de novo heme proteins span a narrow range of values near the center of the range observed for diverse families of natural heme proteins. The measured properties of the de novo heme proteins can be considered as a "default" range for CO binding in α-helical proteins that have neither been designed to bind heme or CO, nor subjected to genetic selections for heme or CO binding.

Introduction

We previously reported the design and construction of a combinatorial library of de novo proteins targeted to fold into 4-helix bundles.[1] The library was created using the "binary code" strategy for protein design. In this strategy, each position in the amino acid sequence is designed to be either polar or nonpolar; however, the exact identities of the polar and nonpolar residues are not specified and are varied combinatorially. Combinatorial diversity is made possible by the organization of the genetic code: Five nonpolar amino acids (Met, Leu, Ile, Val, and Phe) can be encoded by the degenerate codon NTN, and six polar amino acids (Lys, His, Glu, Gln, Asp, and Asn) can be encoded by the degenerate codon XAN. (N represents any of the DNA bases A, G, C, or T, while X represents A, G, or C.)

The binary code strategy is based on the premise that the periodicity of polar and nonpolar amino acids in the linear sequence of a protein is sufficient to direct the formation of amphiphilic secondary structures and that such structures will self-assemble into globular folds that maximize burial of hydrophobic side chains and exposure of hydrophilic side chains. For the design of de novo four-helix bundles, the binary patterning of polar and nonpolar amino acids was designed to match the structural periodicity of 3.6 residues per turn found in α-helical secondary structure.[1]

The de novo proteins were expressed from a library of

* Correspondence should be addressed to: Michael H. Hecht. Telephone: 609-258-2901. Fax: 609-258-6746. E-mail: hecht@princeton.edu.

(1) Kamtekar, S.; Schiffer, J. M.; Xiong, H.; Babik, J. M.; Hecht, M. H. *Science* **1993**, *262*, 1680−1685.

10.1021/ja0036007 CCC: $20.00
Published on Web 02/10/2001

synthetic genes in which polar and nonpolar residues were encoded by the degenerate codons XAN and NTN, respectively. In the resulting collection each protein has a different amino acid sequence. Yet, all sequences in the collection share the identical binary pattern of polar and nonpolar residues. Purification and characterization of >50 de novo proteins from the designed collection demonstrated that virtually all of the sequences fold into compact α-helical structures,[1,3] and several of them adopt structures with nativelike properties.[3−6]

Subsequently, we reported that approximately half of the proteins in the collection bind heme, typically with micromolar affinities.[2] The de novo heme proteins had the typical red color of natural heme proteins and displayed the characteristic absorbance spectra, with a Soret peak at ∼412 nm in the oxidized state and at ∼426 nm in the reduced state.[2]

It should be emphasized that the binary code proteins were not explicitly designed to bind heme. Although histidine and methionine (which ligate iron in natural heme proteins) were included in the combinatorial mix, no attempt was made to position these side chains into a heme binding site. Moreover, the library was not subjected to genetic selections for heme binding either in vivo or in vitro. Therefore, this collection of novel heme proteins provides a unique opportunity for an unbiased assessment of the functional potential of heme proteins that have neither been prejudiced by rational design nor selected by evolution.

Among the known functions of natural heme proteins are (i) catalysis of redox reactions, (ii) binding of small molecules (e.g., O_2, CO, CN, NO), and (iii) electron transfer. Each of these activities can be assessed for de novo proteins from our binary code libraries. We have already established the catalytic potential of the novel heme proteins by demonstrating that several of them have significant peroxidase activity, including one heme protein with a catalytic turnover rate only ∼3.5-fold slower than that of horseradish peroxidase.[7] In the current study we assess the capabilities of the de novo proteins to bind and release carbon monoxide.

Results

A Designed Combinatorial Library of de Novo α-Helical Proteins. The proteins used in this study are derived from two different libraries of sequences designed to form 4-helix bundles. Sequences from the initial library[1] are 74 amino acids long and were designed to contain 14 residues per α-helix (not including N-cap and C-cap residues). Several proteins from this library exhibited properties consistent with the formation of well-packed nativelike structures.[3−6] However most of the sequences in the initial 74 residue library appeared to have fluctional structures,[3−6] reminiscent of those observed in molten globule folding intermediates. For the current study, we chose five proteins from this initial library (proteins 16, 76, 86, 90, and F).

The second library of proteins was constructed recently as an iterative redesign of the first library.[8] The second generation sequences are 102 residues long and the α-helices were designed to be ∼50% longer than those in the first library (20 versus 14 residues per helix). These new sequences were generated by using a preexisting first generation sequence as the "parent". We chose sequence 86 because it (i) bound heme, (ii) remained monomeric at high concentrations, and (iii) displayed molten globule-like properties.[3,5,6] To generate the second generation library, the interhelical regions were lengthened, and the four α-helices of protein 86 were each elongated by six residues. The newly introduced residues in the elongated helices were chosen from a combinatorial mix in accordance with the binary code patterning of α-helical structure. The proteins in the second generation library conserve 14 residues per helix, but differ in the remaining six residues per helix and in some of the turn residues. Overall, the proteins from the second generation library are combinatorially diverse at 32 sites. Proteins isolated from this second generation library are not molten globules. They form stable structures that denature cooperatively and yield well-dispersed NMR spectra. These properties indicate that the second generation proteins fold into well-ordered nativelike structures. Three proteins from the second generation library (S-213, S-285, and S-824; the "S" prefix indicates second generation) were assessed for CO binding.

Carbon Monoxide Association Kinetics. To monitor carbon monoxide association kinetics we used flash photolysis to dissociate bound CO and then followed the rebinding of CO over time. Since binding of CO to the ferrous form of the heme proteins causes a blue-shift in the Soret band from ∼426 to ∼421 nm, the rate of CO association can readily be followed by measuring the change in the absorbance spectrum as a function of time. UV/visible spectra were recorded from 40 ns to 2 ms after photolysis. The spectrum of the unphotolyzed complex was subtracted from the measured spectra, and these difference spectra were plotted as a function of time. An example is shown in Figure 1 for heme protein 90. At early times (100 ns after photolyis) the difference spectrum is dominated by a negative peak at 421 nm and a positive peak at 426 nm. This corresponds to loss of the CO-bound species and concomitant appearance of the CO-free ferrous heme protein. At later times, as CO rebinds, these peaks diminish, and the difference spectrum approaches the baseline. Data were collected for all eight heme proteins in this study. The second-order rate constants for CO association are summarized in Table 1.

The CO association rate constants for our designed heme proteins range from 2.9 to 11 μM^{-1} sec^{-1}. These values can be compared to those measured for two well characterized systems, myoglobin and heme octapeptide (a proteolytic digestion product of cytochrome C containing heme covalently linked to an octapeptide that includes the histidine ligand, also known as "microperoxidase 8").[9] The association rate constants for our designed heme proteins (2.9−11 μM^{-1} sec^{-1}) are between those of myoglobin (0.76 μM^{-1} sec^{-1}) and the heme octapeptide (20 μM^{-1} sec^{-1}).[10] Thus, CO binds faster to the de novo heme proteins than to myoglobin, but not as fast as it binds to the heme octapeptide. These results suggest that for all eight of the binary code proteins, the heme is more solvent-exposed than in myoglobin, but not as exposed as in the heme octapeptide.

Carbon Monoxide Dissociation Kinetics. To monitor CO dissociation kinetics without the complications that would result from CO rebinding, it is essential to trap the dissociated heme protein in a form that cannot rebind CO.[11] This can readily be accomplished by using potassium ferricyanide to rapidly oxidize the dissociated heme protein from the Fe^{2+} to the Fe^{3+} form. In the oxidized state, the heme proteins do not rebind CO. The overall reaction—CO dissociation followed by heme oxidation—

(2) Rojas, N. R. L.; Kamtekar, S.; Simons, C. T.; Mclean, J. E.; Vogel, K. M.; Spiro, T. G.; Farid, R. S.; Hecht, M. H. *Protein Sci.* **1997**, *6*, 2512−2524.

(3) Roy, S.; Hecht, M. H. *Biochemistry* **2000**, *16*, 4603−4607.

(4) Roy, S.; Ratnaswamy, G.; Boice, J. A.; Fairman, F.; McLendon, G.; Hecht, M. H.; *J. Am. Chem. Soc.* **1997b**, 119, 5302−5306.

(5) Roy, S.; Helmer, K. J.; Hecht, M. H. *Folding Des.* **1997a**, *2*, 89−92.

(6) Rosenbaum, D. M.; Roy, S.; Hecht, M. H. *J. Am. Chem. Soc.* **1999**, *121*, 9509−9513.

(7) Moffet, D. A.; Certain, L. K.; Smith, A. J.; Kessel, A. J.; Beckwith, K. A.; Hecht, M. H. *J. Am. Chem. Soc.* **2000**, *122*, 7612−7613.

(8) Liu, T.; Wei, Y.; Pelczer, I.; Sazinsky, S. L.; Moffet, D. A.; Hecht, M. H. Manuscript submitted.

(9) Adams, P. A. In *Peroxidases in Chemistry and Biology*; Everse, J., Everse, K. E., Grisham, M. B., Eds.; CRC Press: Boston, MA, 1993; Vol II, pp 171−200.

(10) Sharma, V. J.; Schmidt, M. R.; Ranney, H. M. *J. Biol. Chem.* **1976**, *251*, 4267−4272.

(11) Cassat, J. C.; Marini, C. P. *Biochemistry* **1974**, *13*, 5323−5328.

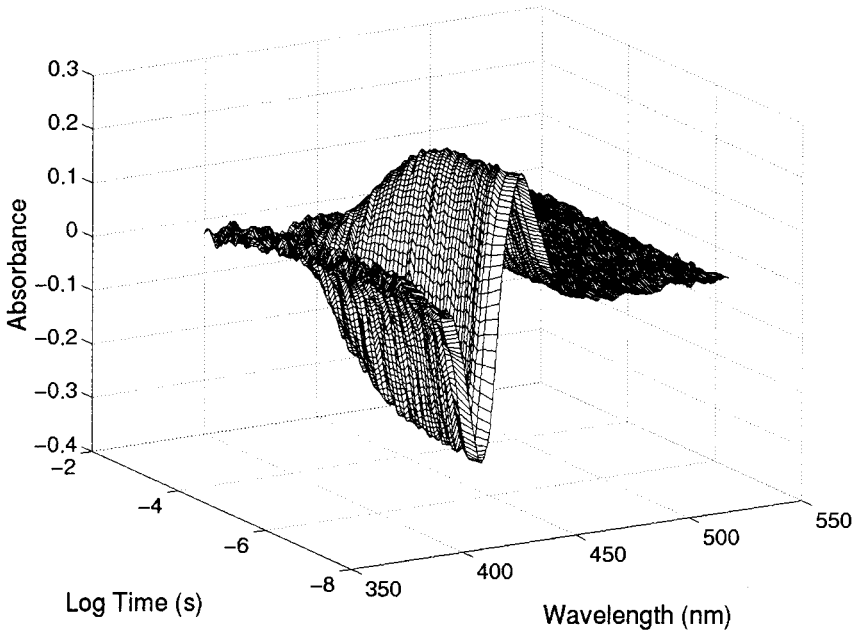

Figure 1. Time course difference spectra showing the CO association kinetics for heme protein 90. The negative peak at 421 nm corresponds to the loss of the CO bound species. The positive peak at 428 nm corresponds to the formation of the ferrous, CO-dissociated, heme protein.

Table 1: Carbon Monoxide Association Rate Constants (k_{on}), Dissociation Rate Constants (k_{off}), and Equilibrium Dissociation Constants (K_d) for de Novo Heme Proteins Compared with Natural Heme Proteins and Heme Octapeptide[a]

protein	k_{on} (μM^{-1} s^{-1})	k_{off} (s^{-1})	$K_{Dissociation}$ (nM)	references
86	5.1	0.028	5.5	this work
F	4.6	0.05	11	this work
90	4.4	0.11	25	this work
76	4.5	0.084	18	this work
16	11	0.086	7.7	this work
S-824	5.3	0.031	5.9	this work
S-213	2.9	0.046	16	this work
S-285	3.4	0.039	11	this work
Mb(equine)	0.76	0.017	22	this work
P450		6.5		12
HRP		0.000072		12
Mb(whale)	0.51	0.019	37	22
Mb(human)	0.76	0.022	29	22
Mb(pig)	0.78	0.019	24	22
octapeptide	20	0.01	0.5	10

[a] Abbreviations: proteins 86, F, 90, 76, and 16 are from the original collection of designed proteins. Proteins S-824, S-213, and S-285 are from the second generation collection of proteins with elongated helices. Mb = myoglobin, P450 = cytochrome P450, HRP = horseradish peroxidase, and octapeptide is the proteolyzed heme-containing fragment of cytochrome *c* (also called microperoxidase).

is monitored by measuring the diminution of the Soret peak at 421 nm (which corresponds to the CO-bound Fe^{2+} species).

CO dissociation kinetics were measured at several concentrations of potassium ferricyanide (see Experimental Section). The time-dependent attenuation of the Soret peak of the Fe^{2+} form of protein S-824 is shown in Figure 2 for several concentrations of potassium ferricyanide. These data were used to calculate an apparent rate constant (k_{app}) at each potassium ferricyanide concentration. A double reciprocal plot of $1/k_{app}$ versus $1/[Fe(CN)_6]^{3+}$ yields a straight line with a *Y*-intercept corresponding to $1/k_{off}$ (i.e., the reciprocal of the CO dissociation rate constant—see Experimental Section). The double reciprocal plot for protein S-824 is shown in Figure 3. Data were collected for all eight heme proteins. The first-order CO dissociation rate constants for the de novo heme proteins range from 0.028 to 0.11 s^{-1} (Table 1). These values fall near the middle of the range

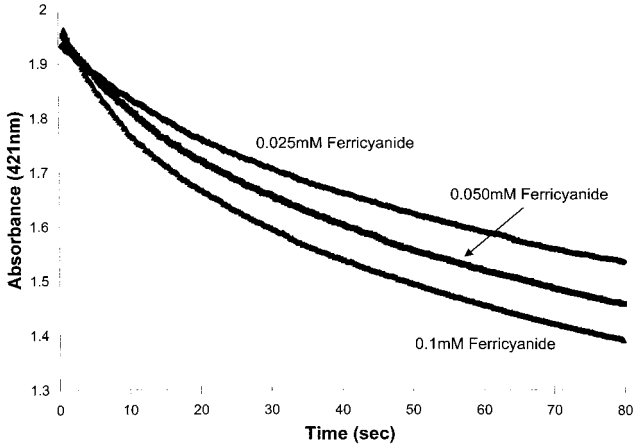

Figure 2. CO dissociation kinetics for heme protein S-824 shown at different concentrations of ferricyanide. As CO dissociates from the heme, ferricyanide oxidizes the heme iron to the ferric state, thereby preventing CO from rebinding. Rates are determined by monitoring the loss of the CO-bound ferrous heme signal at 421 nm.

observed for natural heme proteins, which span from 7.2×10^{-5} sec^{-1} for HRP to 6.5 s^{-1} for cytochrome P450.[12]

Carbon Monoxide Affinity. The equilibrium dissociation constants for CO binding to the binary code heme proteins were calculated from the rate constants ($K_D = k_{off}/k_{on}$) and are listed in Table 1. The dissociation constants are all nanomolar, ranging from 5.5 to 25 nM. For comparison, the K_D of horse myoglobin for CO is 22 nM. Thus, the affinities of the binary code proteins for CO are similar to (and in some cases several-fold greater than) that of myoglobin.

Resonance Raman Spectroscopy of CO-Bound Heme Proteins. The local environment around the bound CO was probed by resonance Raman spectroscopy. Previous studies of the CO adducts of natural heme proteins provide a solid framework for correlating spectral features with structural properties.[13] Two spectral features that are particularly informative are the stretching frequencies of the C−O and Fe−CO

(12) Scheele, J. S.; Kharitonov, V. G.; Martasek, P.; Roman, L. J.; Sharma, V. S.; Masters, B. S. S.; Magde, D. *J. Biol. Chem.* **1997**, *272*, 12523−12528.

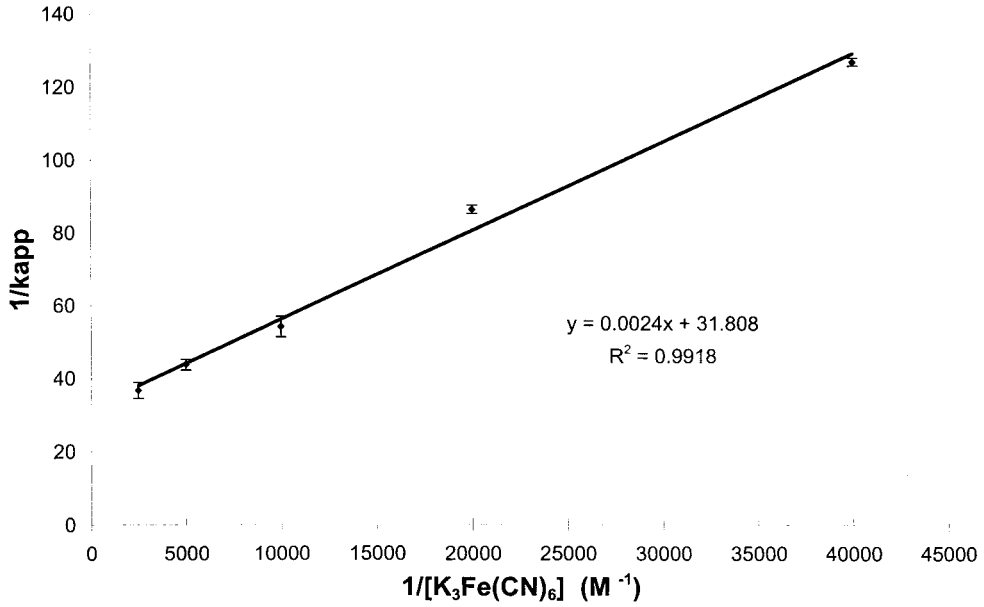

Figure 3. Double reciprocal plot used to determine k_{off} for protein S-824. *Y*-intercept has the value of $1/k_{off}$. The slope has the value of k_{on}[CO]/$k_{ox}k_{off}$. Error bars represent the standard deviation of 3−5 runs for each concentration of ferricyanide.

bonds. The resonance Raman peaks corresponding to these bond stretches can readily be assigned by measuring spectra for both ^{12}CO and ^{13}CO adducts and noting which peaks are shifted by the isotopic substitution. For each of the eight proteins described above, we recorded spectra for both isotopic forms of CO. Examples of these spectra are shown in Figure 4. The C−O and Fe−CO stretching frequencies for the eight de novo proteins and for several model compounds and natural proteins are summarized in Table 2.

Correlations between the stretching frequencies of the C−O and Fe−CO bonds, and the local structure of the CO binding site can be derived from earlier studies of model compounds and natural heme proteins.[13] These correlations are summarized in Figure 5 for several systems in which the proximal ligand is either Met or His. (Since the binary code does not encode cysteine, model data for Cys ligation are not included in the figure.) The entry on the top left of the figure shows horseradish peroxidase (HRP).[14] In the structure of HRP, the hydrogen of an arginine side chain forms a hydrogen bond with the oxygen atom of the bound CO. This weakens the C−O bond, thereby decreasing its stretching frequency. As a result, the Fe−CO bond is strengthened, and its stretching frequency is increased. The next two entries in Figure 5, cytochrome C peroxidase (CCP)[15] and myoglobin (Mb),[16] also place a hydrogen near the bound CO; however, they form only weak or indirect H-bonds with the CO oxygen, and thus exert a less dramatic effect on the stretching frequencies.

At the other extreme, on the bottom right of Figure 5, is an entry for a myoglobin mutant (H64V, V68T). In the structure of this mutant protein, a lone pair on the oxygen of Thr-68 points toward the oxygen atom of the bound CO.[17] This causes the C−O bond to be strengthened and the Fe−CO bond to be weakened. Hence, for Mb (H64V, V68T) the frequency of the C−O stretch is increased, and that of the Fe−CO stretch is decreased.[17]

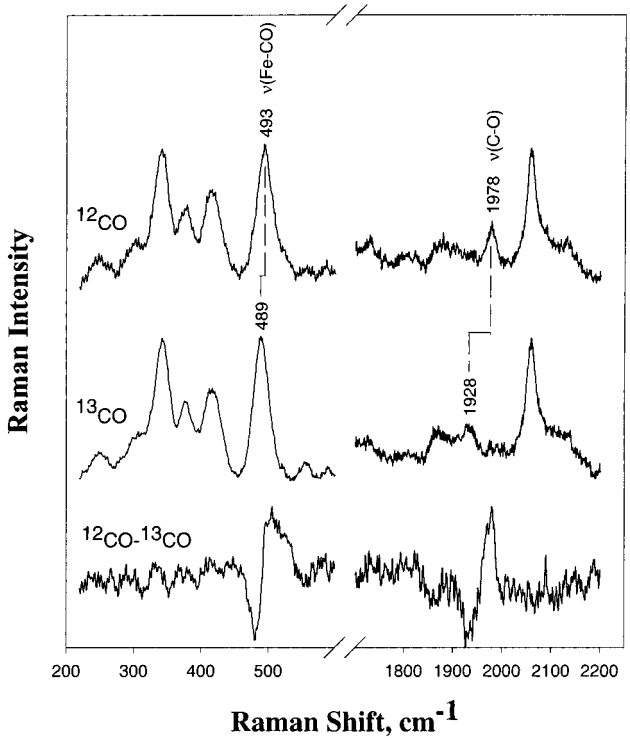

Figure 4. Low- and high-frequency resonance Raman spectra of carbonmonoxy heme protein 76. The top spectra were obtained using ^{12}CO, the middle spectra were obtained using ^{13}CO, and the bottom spectra were obtained by subtracting the ^{13}CO spectra from the ^{12}CO spectra.

Midway between myoglobin (Mb) and Mb (H64V, V68T) is an entry for myoglobin at pH 4. In this structure, the distal histidine is protonated and swings away from the CO binding site. Consequently, neither an H-bond donor nor a lone pair is available to interact with the bound CO. As a result, the stretching frequencies for the C−O and Fe−CO bonds are midway between those observed for structures containing either

(13) Ray, G. B.; Li, X.-Y.; Ibers, J. A.; Sessler, J. L.; Spiro, T. G. *J. Am. Chem. Soc.* **1994**, *116*, 162−176.
(14) Evangelisa-Kirkup, R.; Smulevich, G.; Spiro, T. G. *Biochemistry* **1986**, *25*, 4420−4425.
(15) Smulevich, G.; Evangelista-Kirkup, R.; Spiro, T. G. *Biochemistry* **1986**, *25*, 4426−4429.
(16) Tsubaki, M.; Srivastava, R. B.; Yu, N. T. *Biochemistry* **1982**, *21*, 1132.
(17) Biram, D.; Garrah, L. J.; Hester, R. E. In *Spectroscopy of Biological Molecules*; Hestser, R. E., Girling, R. B., Eds.; Royal Society of Chemistry: Cambridge, U.K., 1991; pp 433−434.

Table 2: Resonance Raman Fe−CO and C−O Stretching Frequencies for de Novo Heme Proteins Compared to Natural Heme Proteins and a Synthetic Heme Model Compound

protein/compound	proximal ligation	ν(Fe−CO)	ν(C−O)	reference
HRP, pH 7	His	537	1904	14
CCP, pH 7	His	530	1922	15
sperm whale Mb, pH 7 (A_1)	His	512	1944	16
FePPDME(ImH)/Bz	Im	495	1969	14
protein F	His	488	1967	this work
protein 86	His	492	1963	this work
protein 16	His	493	1970	this work
protein 90	His	492	1968	this work
protein 76	His	493	1978	this work
protein 213	His	496	1979	this work
protein 285	His	497	1972	this work
protein 824	His	497	1973	this work
Mb, pH 4 (A_o)	His	493	1971	18
Mb(H64V/V68T)	His	476	1984	17

H-bond donors or lone pairs.[18] Next to the entry for myoglobin at pH 4 is an entry for Fe−PPDME(ImH/Bz).[14] In this model compound (protoporphyrin IX dimethylester), the proximal ligand is imidazole, and the solvent is benzene. Therefore, as with the Mb(pH4) structure, neither a H-bond donor nor a lone pair is available for interaction with the CO.

Resonance Raman spectra of all eight of the de novo proteins yielded C−O and Fe−CO stretching frequencies similar to those observed for Mb(pH4) and PPDME(ImH/Bz). These results demonstrate that the local environment of the bound CO in these structures has neither a H-bond donor nor a lone pair available for interaction with the CO.[19]

Discussion

The carbon monoxide association rate constants for our designed heme proteins range from 2.9 to 11 μM^{-1} sec^{-1}. In contrast, the CO association rate constants for myoglobins from various different species are typically less than 1.0 μM^{-1} sec^{-1}.[21−23] The relatively slow association rate for myoglobin is generally attributed to steric hindrance; the heme is buried within the protein and therefore is less accessible to small ligands. Compared to myoglobin, the heme sites in our de novo proteins are not as well enclosed.

Another well-studied system is the heme octapeptide.[9] This model system, also known as microperoxidase-8, contains a 5-coordinate heme iron with histidine as the proximal ligand. The distal side of the heme is entirely exposed to solvent. The association rate constant for CO binding to the heme octapeptide is 20 μM^{-1} sec^{-1}.[10] Since the association constants for our de novo heme proteins are 2- to 7-fold slower than for heme octapeptide, we conclude that the heme sites in our novel proteins are (at least) partially shielded from solvent. Thus, the CO association experiments indicate that the heme sites in the binary code proteins are not as buried as in myoglobin, but neither are they as exposed as in the heme octapeptide.

Both myoglobin and heme octapeptide are 5-coordinate heme species with an open sixth ligation site. Among our combinatorial libraries of de novo heme proteins we expect some will have 5-coordinate heme sites, while others will have 6-coordinate sites. In the latter case, rebinding of CO after photolysis would require either (i) that the CO rebinds before the displaced amino acid ligand can rebind or (ii) that the amino acid ligand rebinds first and is then displaced by CO. Because of the combinatorial nature of our collections, both scenarios are possible. Therefore, we note that CO rebinding in our de novo proteins may occur via mechanisms that differ from those governing the rebinding of CO to 5-coordinate systems such as myoglobin or heme octapeptide.

The CO dissociation rate constants for natural heme proteins cover a much wider range than the association rate constants. Moreover, correlating dissociation rate constants with structural features is more difficult than for association rate constants. This is because the rate of dissociation is determined not only by the structural environment around the CO but also by the strength of the bond between the proximal ligand and the heme iron.[12] The dissociation rate constants for natural heme proteins range from 7.2×10^{-5} s^{-1} for horseradish peroxidase to 6.5 s^{-1} for cytochrome P450.[12] For the binary code proteins, the range is much narrower. As shown in Table 1, the dissociation rate constants for the binary code proteins range from 0.028 to 0.11 s^{-1}. These values are near the middle of the range for natural proteins, and are 1.5- to 6.5-fold faster than for myoglobin (0.017 s^{-1}). While it is tempting to suggest that the faster dissociation rates relative to myoglobin indicate a more exposed CO binding site, this would be an oversimplification as the rate of dissociation is also affected by the strength of the bond between the proximal ligand and the heme iron.[12] (Note that the dissociation rate constant (0.01 sec^{-1}) of the heme octapeptide with its fully exposed CO site is close to that of myoglobin (0.017 s^{-1})).

It is interesting to compare our collection of de novo heme proteins with a collection of single-site mutants of myoglobin. Although both the k_{on} and the k_{off} of wild-type myoglobin are slower than those of our binary code proteins, various single-site substitutions in myoglobin can speed up both rate constants to the same range as observed for our binary code proteins. For example, substitution of the distal histidine of myoglobin by any of nine different amino acids increases the k_{on} to the same range as that of our binary code proteins.[22] Likewise, eight different mutants of the distal histidine and several other mutations in the CO site increase the myoglobin k_{off} to values similar to those of our binary code proteins.[22] Myoglobin has undergone countless generations of evolutionary optimization leading to a finely tuned binding site for a small diatomic ligand. Yet, minor perturbations of this binding site by any of several different mutations yield binding kinetics similar to those observed for our unselected and nonoptimized binary code proteins.

The resonance Raman experiments indicate that none of the eight binary code proteins contribute a H-bond donor or a lone pair for interaction with the CO. Although such interactions are often observed in the CO adducts of natural heme proteins,[14−18] it is not surprising that they are absent in our de novo proteins. Indeed, we presume that the fortuitous occurrence of specific directional interactions with a bound ligand would be rare among proteins isolated from a library of de novo structures that were neither explicitly designed nor genetically selected for CO binding.

The equilibrium dissociation constants (K_D) for CO binding

(18) Ramsden, J.; Spiro, T. G. *Biochemistry* **1989**, *28*, 3125−3128.

(19) The HWHH's (half width half height) of the C−O stretching bands in the de novo proteins are 10−12 cm^{-1}. It might be postulated that the width of the C−O stretching band would indicate solvent exposure due to inhomogeneous broadening via interaction between CO and the aqueous solvent. However, the HWHH of the C−O stretching band in Microperoxidase-11, a water soluble heme undecapeptide in which the CO is completely exposed, has been reported as 8 cm^{-1} (ref 20), somewhat narrower than the HWHH of the de novo proteins (10−12 cm^{-1}). Thus, bandwidth cannot be used to indicate solvent exposure.

(20) Laberge, M.; Vreugdenhill, A. J.; Vanderkooi, J. M.; Butler, I. S. *J. Biomol. Struct. Dyn.* **1998**, *15*, 1039.

(21) Rohlfs, R. J.; Mathews, A. J.; Carver, T. E.; Olson, J. S.; Springer, B. A.; Egeberg, K. D.; Sligar, S. G. *J. Biol. Chem.* **1990**, *265*, 3168−3176.

(22) Springer, B. A.; Sligar, S. G.; Olson, J. S.; Phillips, G. N., Jr. *Chem. Rev.* **1994**, *94*, 699−714.

(23) Antonini, E.; Brunori, M. *Hemoglobin and Myoglobin in Their Reactions with Ligands*; Frontiers of Biology, Vol. 21; Elsevier: New York, 1971.

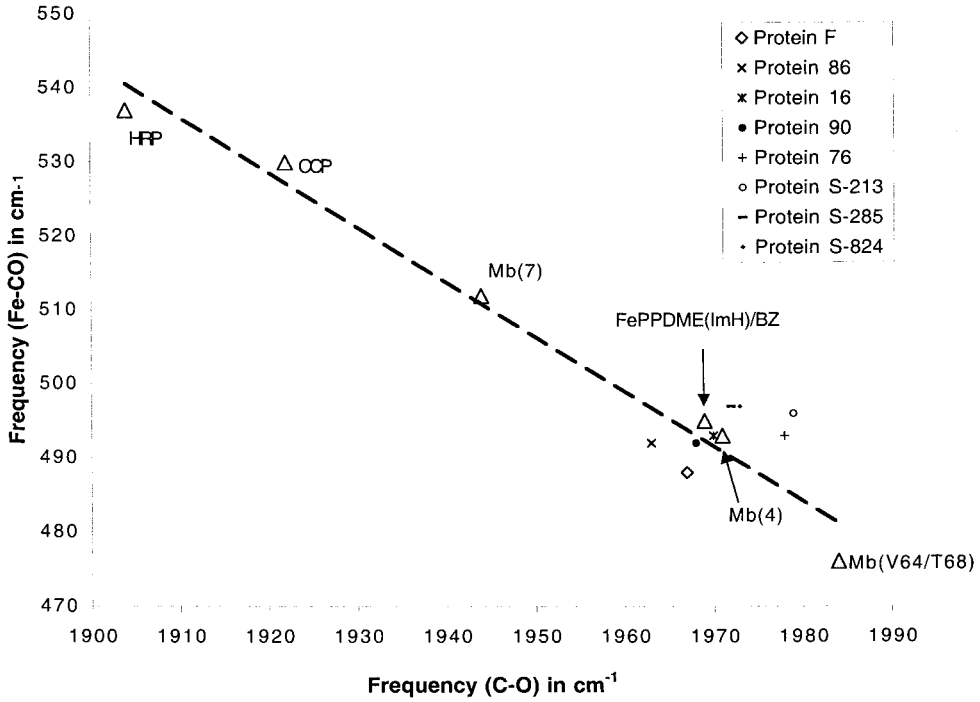

Figure 5. Plot of C−O and Fe−CO stretching frequencies determined by resonance Raman spectroscopy. Model compounds and natural heme proteins are designated by triangles. Symbols for the de novo heme proteins are indicated in the figure. HRP: horseradish peroxidase forms a hydrogen bond with the oxygen atom of the CO. CCP: cytochrome *c* peroxidase forms a weak hydrogen bond with the bound CO. Mb(pH7): sperm whale myoglobin forms an indirect hydrogen bond with the bound CO. Mb(pH4): myoglobin at pH 4.0 has no hydrogen bonding nor lone pair interactions with the CO. FePPDME(ImH/BZ): protoporphyrin IX dimethylester is a model compound with imidazole as one ligand in a benzene solvent. It has no hydrogen bonding nor electron lone pair interactions with the CO. Mb(V64/T68): Double mutant of myoglobin which has a lone pair of electrons pointing directly toward the bound CO.

to the binary code heme proteins range from 5.5 to 25 nM. For comparison, the K_D for CO binding to equine myoglobin is 22 nM. Equine myoglobin has evolved to selectively bind diatomic oxygen rather than CO, yet myoglobin's K_D for carbon monoxide is similar to that of the unselected de novo heme proteins.

It is interesting to compare the CO binding properties of the second generation proteins (the "S" series) with protein 86 from the first generation. As noted in the Introduction, the second generation proteins were derived from protein 86 by a combinatorial binary code lengthening of the α-helices. We show elsewhere[8] that lengthening these helices is sufficient to transform a molten globule structure (86) into well-folded nativelike structures. Thus, in terms of packing and rigidity, the second generation proteins are very different from the parental sequence 86.[8] Nonetheless, the CO binding properties of the second generation proteins are fairly similar to those measured for protein 86. (k_{on} for 86 = 5.1 μM^{-1} sec^{-1}, and the second generation proteins range from 2.9 to 5.3 μM^{-1} sec^{-1}; k_{off} for 86 = 0.028 s^{-1}, and the second generation proteins range from 0.031 to 0.046 s^{-1}; K_D for 86 = 5.5 nM, and the second generation proteins range from 5.9 to 16 nM.) The similarity of the CO binding properties of the second generation proteins to those of protein 86 may seem surprising. However, there are two reasons why this similarity is quite reasonable. First, the local environment of the heme and CO-binding site in the second generation proteins are probably quite similar to those in protein 86. Since protein 86 already bound heme (and CO) elongating the α-helixes in the construction of the second generation library presumably did not create a new binding site, but rather extended the structure of the proteins in a region distant from the heme site. Second, the nativelike properties of the second generation proteins were demonstrated for the apo forms of these sequences,[8] and it is unlikely that these well-packed structures are maintained upon insertion of the heme macrocycle into the hydrophobic core. Thus, even the second generation proteins are likely to be somewhat flexible in their heme bound forms. A similar result was reported by Dutton and co-workers, who designed a 4-helix bundle that was relatively nativelike in its apo form, but became less ordered upon binding heme.[24]

The binary code proteins were not explicitly designed to bind heme or CO. Nor were they subjected to genetic selections (either in vivo or in vitro) for heme or CO binding. Therefore these libraries of de novo proteins provide a unique opportunity for an unbiased assessment of the CO binding potential of heme proteins that have neither been prejudiced by rational design nor selected by evolution. Our characterization of these proteins shows that both the CO binding kinetics and the spectroscopic properties of the de novo heme proteins fall within the range of values observed for natural heme proteins. However, compared to natural proteins, the de novo proteins span a narrower range. We propose that this range be considered as the "default" range for CO binding in α-helical proteins. Natural proteins with binding kinetics or spectroscopic properties that deviate significantly from these default values were selected by evolution for specific biological properties. Likewise, future designs of de novo heme proteins with desired properties that differ from these default values presumably will require optimization by explicit design, evolution in vitro, or both.

Experimental Section

Protein Purification and Sample Preparation. Proteins were overexpressed in *Escherichia coli* strain X90(DE3) grown in 2xYT as described previously.[1] The proteins were extracted from cells using a freeze−thaw protocol[25] and then solubilized in 100 mM $MgCl_2$.[3]

(24) Gibney, B. R.; Rabanal, F.; Reddy, K. S.; Dutton, P. L. *Biochemistry* **1998**, *37*, 4635−4643.

(25) Johnson, B. H.; Hecht, M. H. *Bio/Technology* **1994**, *12*, 1357−1360.

Cellular contaminants were removed by acid precipitation in 50 mM sodium acetate buffer (pH 4.0). The resulting supernatant was loaded onto a POROS HS cation-exchange column (PerSeptive Biosystems), and eluted using a gradient of NaCl from 0 to 1.5 M. Purified proteins were concentrated, and buffer was exchanged to 50 mM Tris, 50mM NaCl, pH 8.0 using Centricon Plus-20 filters (Millipore).

Reconstitution with heme was accomplished using a freshly prepared stock solution of 1−2 M hemin chloride (Sigma) in 0.1 M NaOH. Aliquots of this stock solution were added to the protein samples until a light pink solution was observed. Typical heme protein solutions were 10−15 μM, with Soret peaks between 0.8 and 1.6 absorbance units. Protein concentrations were estimated to be at least 10-fold greater than total heme concentration, thereby minimizing the amount of unbound heme.

CO Association Kinetics. CO Association kinetics were measured using time-resolved UV/visible absorbance spectroscopy to monitor the rebinding of CO to the heme after flash photolysis. Heme protein samples were saturated with CO and reduced with sodium dithionite. Samples were then transferred to an airtight 2 mm path length quartz cuvette that had been flushed with CO. Flash photolysis was accomplished using a Quanta Ray GCR-170 Nd:YAG laser/MOPO-710 (Spectra Physics) with a 7 ns pulse width. Excitation wavelengths were tuned between 538 and 542.5 nm to maximize overlap with the beta Q_o band of the heme. CO reassociation kinetics were monitored by recording absorbance difference spectra as a function of time. Difference spectra were collected using a 1103 EG&G Xe flash lamp probe source and a Princeton Instruments IRY-700/15 image intensifier diode array detector. The image intensifier was gated for 20 ns using an Avtech AVRL-1-PS pulse generator. Timing synchronization of the intensifier, flash lamp probe, and diode array detector readout was controlled using a Stanford Research Systems DG535 delay gate generator. The methodology is that of Goldbeck and Kliger[26] except that timing synchronization was taken from the laser Q switch. Difference spectra were recorded from 40 ns to 2 ms after photolysis. Approximately 50 time points were recorded for each sample; each averaged over 4 shots. Data were analyzed using Matlab (Mathworks Inc.). Kinetic rates were calculated by singular value decomposition of time/wavelength/intensity matrices as described by Hofrichter et al. [27]

CO Dissociation Kinetics. Reconstituted heme protein was placed in an airtight vial and saturated with carbon monoxide (Matheson). Aqueous sodium dithionite was added dropwise to reduce the heme and scavenge any residual oxygen. Aqueous potassium ferricyanide, saturated with CO, was used as the oxidative trapping agent.[11] Following dissociation of CO from the 6-coordinate ferrous heme protein, ferricyanide oxidizes the heme to the 5-coordinate ferric state, thereby preventing rebinding of CO. This reaction can be monitored by following the loss of the Fe^{2+} Soret peak at 421 nm. The general scheme is as follows:

$$\mathrm{PFe(II)-CO} \xrightarrow{k_{off}} \mathrm{PFe(II)} + \mathrm{CO} \quad (1)$$

$$\mathrm{PFe(II)} + \mathrm{CO} \xrightarrow{k_{on}} \mathrm{PFe(II)-CO} \quad (2)$$

$$\mathrm{PFe(II)} + \mathrm{Fe(CN)_6}^{3-} \xrightarrow{k_{ox}} \mathrm{PFe(III)} + \mathrm{Fe(CN)_6}^{4-} \quad (3)$$

where PFe(II) denotes the reduced form of the heme protein and PFe(III) denotes the oxidized form. Assuming steady state in [PFe(II)],

$$k_{app} = (k_{off})(k_{ox})[\mathrm{Fe(CN)_6}^{3-}]/[(k_{on})[\mathrm{CO}] + (k_{ox})[\mathrm{Fe(CN)_6}^{3-}]] \quad (4)$$

In reciprocal form this becomes

$$1/k_{app} = (k_{on}[\mathrm{CO}]/(k_{ox})(k_{off})[\mathrm{Fe(CN)_6}^{3-}]) + (1/k_{off}) \quad (5)$$

A plot of $1/k_{app}$ versus $1/[\mathrm{Fe(CN)_6}^{3-}]$ yields a straight line with a *Y*-intercept of $1/k_{off}$.

Kinetic measurements were accomplished using a HI-TECH Scientific SF-61 DX2 Double Mixing Stopped Flow System, equipped with a Xe^+ Hg^+ arc lamp and Brandenburg 4479 photomultiplier module detector. k_{app} was determined from the first-order decay monitored at 421 nm upon fast-mixing of the CO-saturated ferrous heme protein with CO-saturated potassium ferricyanide. Kinetics were measured at several different concentrations of potassium ferricyanide (see Figures 2 and 3). All solutions were kept in airtight syringes prior to injection into the stopped flow apparatus.

Resonance Raman Spectroscopy. Heme protein samples (40 μM) were prepared in 50 mM sodium phosphate buffer, 200 mM NaCl, pH 6.8, and reduced with a minimal amount of dithionite. Carbon monoxide adducts were prepared by flowing CO (Matheson) over the protein for >20 min. ^{13}CO adducts were prepared similarly using ^{13}CO (99%, Cambridge Isotope Laboratories) from a 0.25 L glass break-neck flask.

Resonance Raman spectra were measured using excitation from the 413.1 nm line of a Kr^+ laser (Coherent). Backscattered light was focused onto the sample in a spinning NMR tube using a cylindrical lens to minimize photodecomposition. The light was dispersed with a triplemate spectrograph (Spex triplemate) and detected with an intensified photodiode array (Princeton Instruments). All spectra were calibrated with indene as a standard. Frequencies are accurate to ± 1 cm^{-1} for isolated bands. Spectra were processed with Grams/32 software (Galactic Industries Corp.).

Appendix

Derivation of the CO Dissociation Rate Constant. The apparent rate of the Fe(II)−CO dissociation reaction is

$$\mathrm{rate} = k_{app}[\mathrm{PFe(II)-CO}] \quad (6)$$

The rate of formation of heme Fe(III) is:

$$\mathrm{rate} = \frac{\mathrm{d[PFe(III)]}}{\mathrm{d}t} = k_{ox}[\mathrm{PFe(II)}][\mathrm{Fe(CN)_6}^{3-}] \quad (7)$$

Combining eqs 6 and 7 gives the rate equation:

$$k_{app}\,[\mathrm{Fe(II)-CO}] = k_{ox}[\mathrm{PFe(II)}][\mathrm{Fe(CN)_6}]^{3-} \quad (8)$$

Assuming steady state in PFe(II) and using eqs 1−3 we find that:

$$\frac{\mathrm{d[PFe(II)]}}{\mathrm{d}t} = -k_{on}[\mathrm{PFe(II)}][\mathrm{CO}] + k_{off}[\mathrm{PFe-CO}] - k_{ox}[\mathrm{PFe(II)}][\mathrm{Fe(CN)6}]^{3-} = 0 \quad (9)$$

This yields

$$[\mathrm{PFe(II)}] = \frac{k_{off}[\mathrm{PFe-CO}]}{k_{on}[\mathrm{CO}] + k_{ox}[\mathrm{Fe(CN)_6}]^{3-}} \quad (10)$$

Substituting eq 10 into eq 8 yields eq 4.

JA0036007

(26) Goldbeck, R. A.; Kliger, D. S. *Methods Enzymol.* **1993**, *226*, 147−177.

(27) Hofrichter, J.; Ansari, A.; Jones, C. M.; Deutsch, R. M.; Sommer, J. H.; Henry, E. R. *Methods Enzymol.* **1994**, *232*, 387−415.