

De novo heme proteins from designed combinatorial libraries

NINA R. L. ROJAS,^{1,4} SATWIK KAMTEKAR,^{1,5} CYRENA T. SIMONS,^{1,6} JEREMY E. MCLEAN,^{2,7}
KATHLEEN M. VOGEL,¹ THOMAS G. SPIRO,¹ RAMY S. FARID,³ AND MICHAEL H. HECHT^{1,2}

¹Department of Chemistry, Princeton University, Princeton, New Jersey 08544

²Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544

³Department of Chemistry, Rutgers University, Newark, New Jersey 07102

(RECEIVED May 19, 1997; ACCEPTED September 19, 1997)

Abstract

We previously reported the design of a library of de novo amino acid sequences targeted to fold into four-helix bundles. The design of these sequences was based on a "binary code" strategy, in which the patterning of polar and nonpolar amino acids is specified explicitly, but the exact identities of the side chains is varied extensively (Kamtekar S, Schiffer JM, Xiong H, Babik JM, Hecht MH, 1993, *Science* 262:1680–1685). Because of this variability, the resulting collection of amino acid sequences may include de novo proteins capable of binding biologically important cofactors. To probe for such binding, the de novo sequences were screened for their ability to bind the heme cofactor. Among an initial collection of 30 binary code sequences, 15 are shown to bind heme and form bright red complexes. Characterization of several of these de novo heme proteins demonstrated that their absorption spectra and resonance Raman spectra resemble those of natural cytochromes. Because the design of these sequences is based on global features of polar/nonpolar patterning, the finding that half of them bind heme highlights the power of the binary code strategy, and demonstrates that isolating de novo heme proteins does not require explicit design of the cofactor binding site. Because bound heme plays a key role in the functions of many natural proteins, these results suggest that binary code sequences may serve as initial prototypes for the development of large collections of functionally active de novo proteins.

Keywords: binary code; combinatorial libraries; de novo proteins; heme binding; protein design

The two central goals of protein design are: (1) to devise novel amino acid sequences that fold into particular target structures; and (2) to obtain from such sequences proteins that perform specific functions.

Our strategy for achieving the first goal employs binary patterning of polar and nonpolar amino acids as the cornerstone for designing target protein structures (Kamtekar et al., 1993). In this strategy, the sequence locations of polar and nonpolar residues are specified explicitly, but the precise identities of these residues are not constrained. Thus, a binary pattern designed to fold into a particular target structure is compatible with many different amino acid sequences. We tested the utility of the binary code strategy by designing a binary pattern for a four-helix bundle, constructing a

combinatorial library of genes to encode this pattern, and then expressing from these genes a collection of novel amino acid sequences that are compatible with this binary pattern (Kamtekar et al., 1993). Sequence diversity was generated in the library of synthetic genes by using the degenerate DNA codon NAN (where N represents a mixture of A, G, T, and C) at all positions destined to encode polar amino acids (Lys, His, Glu, Gln, Asp, and Asn) and NTN at all positions destined to encode nonpolar amino acids (Met, Leu, Ile, Val, and Phe).

Purification and characterization of an initial collection of these de novo proteins revealed that the majority of the sequences fold into compact α -helical structures (Kamtekar et al., 1993). More recent work has demonstrated that several of the proteins adopt structures with native-like properties (Roy et al., 1997a, 1997b).

Although the binary code strategy appears well-suited for exploring the sequence requirements for designing particular structures, the combinatorial underpinnings of this strategy preclude the explicit design of catalytic sites. Can the binary code strategy nonetheless yield prototypes toward the eventual goal of producing novel proteins that perform specific functions?

An appealing strategy for pursuing this second goal of protein design is to achieve function by co-opting a cofactor. One cofactor used frequently by nature is the heme group. Natural heme proteins

Reprint requests to: Michael H. Hecht, Department of Chemistry, Princeton University, Princeton, New Jersey 08544; e-mail: hecht@princeton.edu.

⁴Present address: Department of Chemistry, Ateneo de Manila University, Manila, Philippines.

⁵Present address: Department of Molecular Biophysics & Biochemistry, Yale University, New Haven, Connecticut.

⁶Present address: New York University School of Medicine, New York, New York.

⁷Present address: Graduate program in Biophysics & Structural Biology, Brandeis University, Waltham, Massachusetts.

perform a variety of biological functions ranging from oxygen binding to electron transport (Dolphin, 1979). In natural heme proteins, the heme moiety is sometimes attached to the protein covalently, but, in most cases, the cofactor is bound by noncovalent interactions. Thus, covalent attachment sites presumably are not required for the successful design of functionally active de novo heme proteins. Among the structures of natural heme proteins, several are four-helix bundles (e.g., cytochrome b_{562} and cytochrome c' ; Mathews et al., 1979; Finzel et al., 1985; Hamada et al., 1995). Thus, formation of a heme binding site is clearly compatible with the architecture of our target structure.

De novo four-helix bundles capable of binding heme have been constructed by DeGrado, Dutton, and coworkers. For example, Choma et al. (1994) used computer modeling to guide the rational design of histidine side chains at key positions in the sequence of their de novo four-helix bundle α_2 . Heme binding sites in de novo proteins need not be limited to one heme per protein, and multi-heme proteins have been designed by Dutton and coworkers as "molecular maquettes" of the natural proteins involved in respiration and photosynthesis (Robertson et al., 1994; Rabanal et al., 1996; Gibney et al., 1997). The success of these initial forays into the design of novel heme proteins prompted us to question whether our binary code α -helical proteins might bind heme.

To achieve heme binding, a de novo sequence must include residues, such as histidine and methionine, that can coordinate the central iron. However, the mere occurrence of these residues is not sufficient. Many natural sequences are rich in histidine and methionine, but nevertheless fail to bind heme. Thus, the sequence locations and spatial orientations of these residues are clearly important. Histidine and methionine are well-represented in our library of binary code sequences (Kamtekar et al., 1993), but neither the number nor the locations of these residues were specified a priori. Because the binary code relies on combinatorial methods, a heme binding site could not be designed explicitly. We now ask whether such explicit design is essential. Can the binary code yield proteins that bind heme?

Here we describe a screen of the binary code proteins and demonstrate, surprisingly, that approximately half of the sequences bind heme. Upon addition of hemin chloride, these proteins yield bright red solutions and the absorption spectra of these heme/protein complexes resemble those of natural heme proteins. Further biophysical characterization demonstrates that several of the de novo heme proteins display resonance Raman spectra and binding affinities resembling those of natural cytochromes.

Results

Screening for heme binding

A convenient screen for heme proteins relies on their bright red color. This color is associated with a characteristic absorption spectrum in the visible region, which includes a large Soret peak between 410 nm and 430 nm, as well as weaker absorbances at approximately 550 nm (Antonini & Brunori, 1971; Moore & Pettigrew, 1990). In contrast, hemin chloride in neutral aqueous solution is greenish-brown, and displays a broad absorption with a maximum between 360 nm and 365 nm. The visually striking and spectroscopically quantifiable red color of heme proteins was used to screen our library of de novo sequences for their ability to bind heme.

To rapidly screen a collection of proteins, we developed an assay that does not require chromatographic purification of each

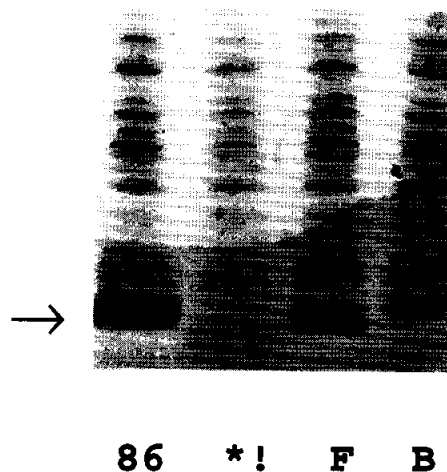


Fig. 1. Coomassie-stained SDS-polyacrylamide gel of semi-pure samples used for the screen. Arrow points to the protein of interest. Sequences of proteins 86, F, and B are shown in Figure 3. *! indicates a control sample in which the third codon of the gene is a nonsense codon.

protein in the collection. We had demonstrated previously that overexpressed proteins can be liberated from *Escherichia coli* cells by a freeze-thaw procedure (Johnson & Hecht, 1994; Roy et al., 1997a). The resulting "semi-pure" samples contain relatively limited amounts of bacterial contaminants. Acidifying the samples to pH 4 precipitates most of the remaining contaminants. These rapidly prepared samples were dialyzed into an appropriate buffer, and shown to be sufficiently pure for our screen. The SDS-PAGE profile of a typical set of samples is shown in Figure 1. The main component of each sample is the de novo protein.

Screening these samples demonstrated that many of the proteins in our collection bind heme and yield the familiar red color (Figs. 2, 3). Moreover, the absorption spectra of these de novo heme-binding proteins are similar to those of natural heme proteins, and include both the Soret peak at 412–414 nm and the

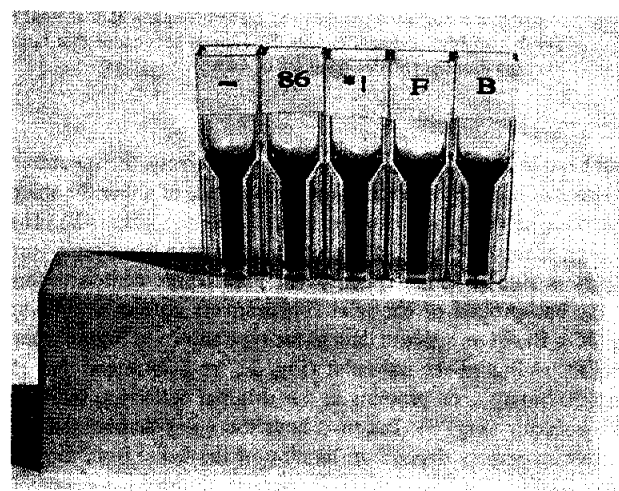


Fig. 2. Red color associated with formation of a heme-protein complex. All samples contain $\sim 55 \mu\text{M}$ hemin chloride. Left to right: No protein (-); protein 86; negative control from cells expressing a nonsense mutant (*!); protein F; protein B.

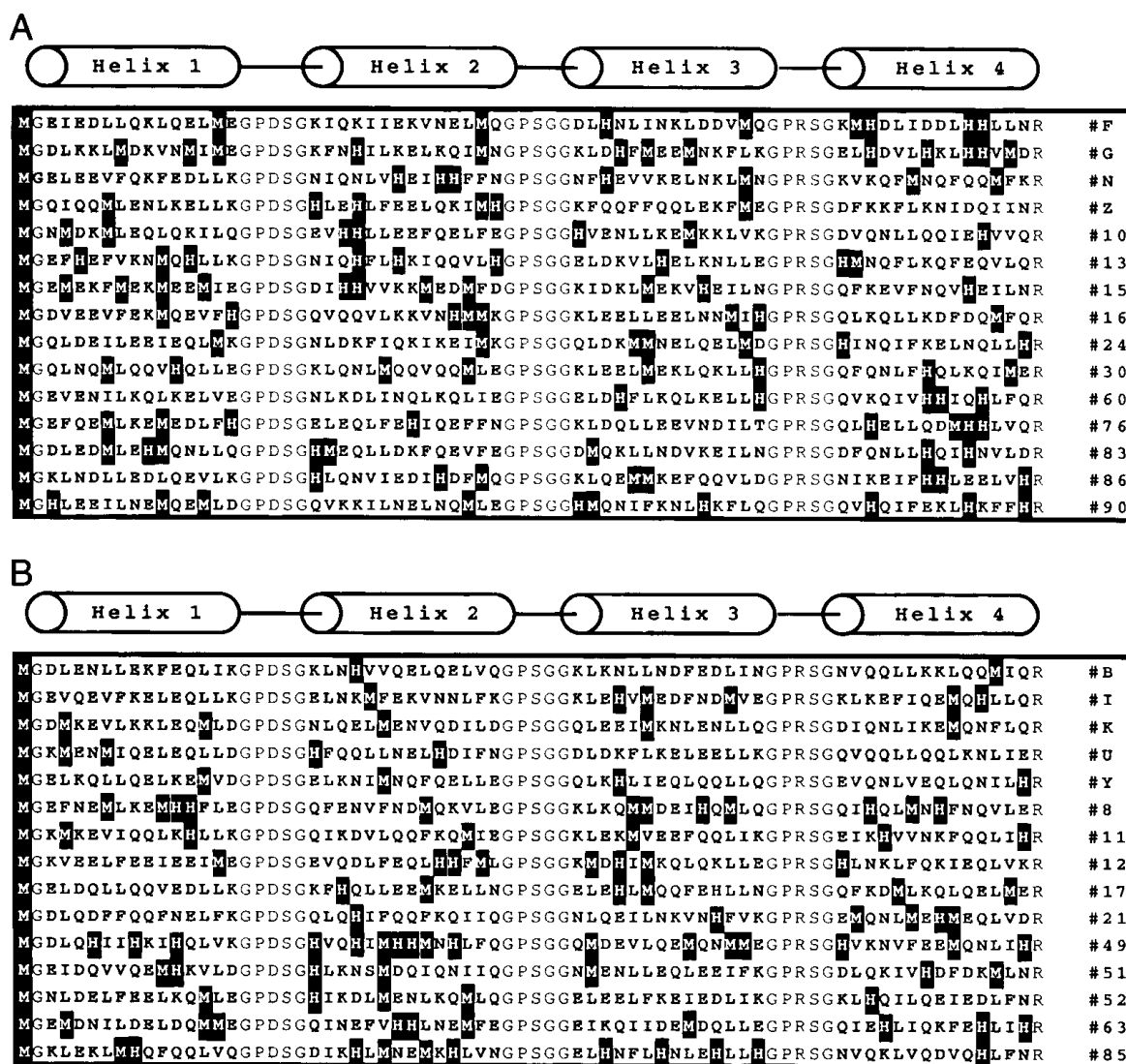


Fig. 3. Amino acid sequences of the 30 proteins screened for heme binding. Histidine and methionine residues are highlighted. [Note: The original collection published by Kamtekar et al. (1993) contained 29 sequences. The sequence of protein #21 has since been verified and is now included]. **A:** Fifteen sequences that bind heme (average number of His residues per protein is 4.3, and average number of Met residues is 4.9). **B:** Fifteen sequences that fail to bind heme (average number of His residues per protein is 3.5, and average number of Met residues is 4.8).

broad band at around 550 nm (Fig. 4A). Conversely, the spectra of proteins that fail to bind heme show no evidence of either the sharp Soret peak or the broad band at around 550 nm (Fig. 4B). The spectrum of the residual *E. coli* proteins present in these semi-pure samples is similar to the spectrum of pure heme, demonstrating that the background of bacterial contaminants do not bind heme (Fig. 4C). Based on spectral data of the type shown in Figure 4 and visual inspection of the red color (Fig. 2), we determined that 15 of the 30 binary code proteins in our original collection bind the heme cofactor (Fig. 3). Because both the color screen and the spectroscopic assay depend on ligation of the heme iron, the observed binding must include such ligation. Spurious nonspecific "stickiness" between the protein and the nonpolar heme moiety would not suffice to yield the observed results.

The frequent occurrence of heme ligation in our collection led us to question whether the mere presence of histidines and methi-

onines in a de novo sequence might be sufficient for heme binding. This is clearly *not* the case. Many sequences in the collection contain several His and Met residues, but fail to bind heme (Fig. 3). For example, sequence 49 has 10 His and 8 Met residues, but the spectrum of this protein shows no detectable Soret peak. Overall, the number of His and Met residues, and the ability of a sequence to bind heme are not correlated. Although residues capable of coordinating the heme iron are clearly important for heme binding, their mere inclusion into a de novo sequence is not sufficient. The locations of the His and/or Met side chains and the overall structure of the protein must play key roles in forming a binding site.

Spectroscopic characterization of de novo heme proteins

Several of the proteins that showed strong Soret absorbance in the screen were purified for further characterization. The absorption

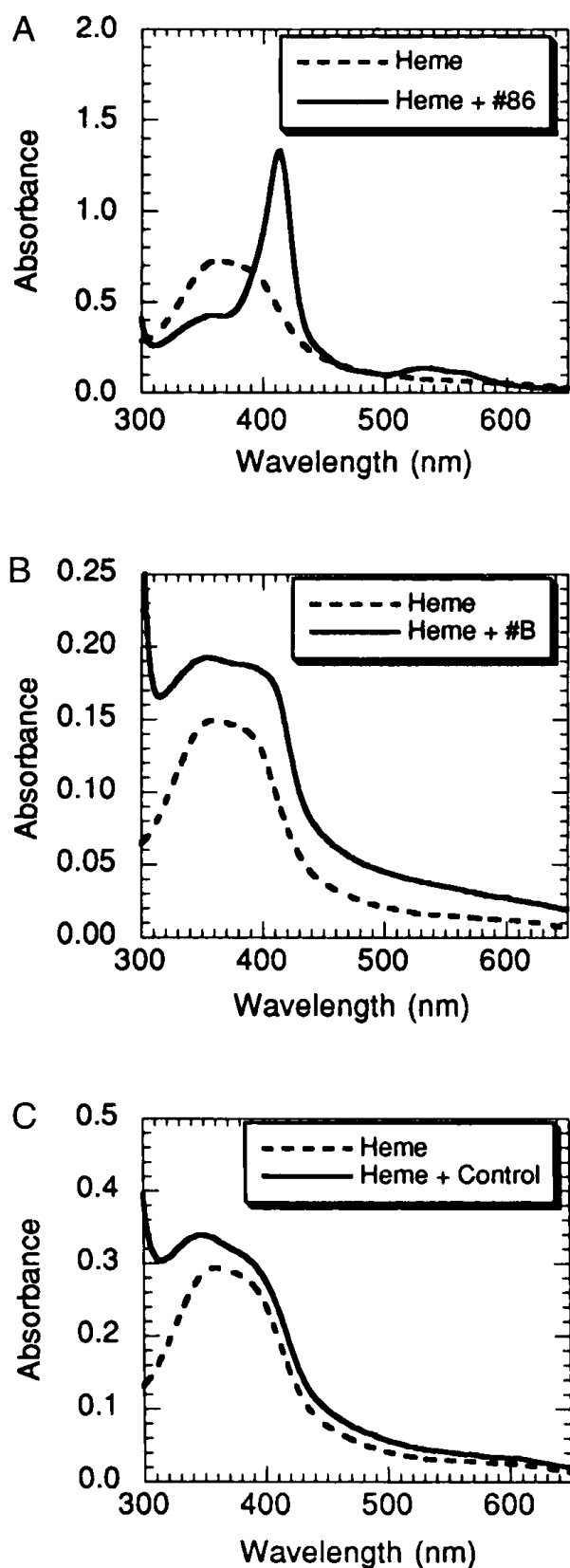


Fig. 4. Examples of absorption spectra from the heme binding screen. A: Protein 86. B: Protein B. C: Negative control (from cells expressing a nonsense mutant).

spectra of three purified proteins (86, F, and G) in both the oxidized and reduced forms are shown in Figure 5. These spectra are very similar to those of natural heme proteins. Comparison of the peaks in the spectra of both the reduced and the oxidized forms of the de novo proteins indicates that they are similar to *b*-type cytochromes (Table 1). For example, in the spectra of both the de novo proteins and the *b*-type cytochromes, the Soret band is shifted to higher wavelength following reduction of the heme (Table 1). Reduction also causes the broad absorbance near 550 nm to resolve into two distinct bands at 530 nm and 560 nm. This band splitting is typical of cytochromes and suggests that the reduced heme complexes of the de novo proteins contain low spin Fe(II) and are six-coordinate (Moore & Pettigrew, 1990).

In *b*-type cytochromes, the heme iron is typically ligated either by two His residues (e.g., cytochrome *b₅*) or by a His/Met combination (e.g., cytochrome *b₅₆₂*). In cases where a Met is involved, the spectrum of the oxidized form of the protein typically includes a very weak band at 695 nm, which is due to sulfur-to-iron charge transfer (Dickerson & Timkovitch, 1975; Moore & Pettigrew, 1990). For our de novo proteins, the spectra of the oxidized forms show no evidence of a peak at 695 nm. However, because the absorbance due to sulfur-to-iron charge transfer is inherently weak (ϵ_{695} can be >100-fold lower than ϵ_{Soret} ; Moore & Pettigrew, 1990), these experiments do not rule out coordination by methionine.

Resonance Raman spectroscopy

The coordination and spin state of the heme iron was also analyzed by resonance Raman spectroscopy. Spectra were measured for proteins F and 86 in both the oxidized form (Fig. 6A) and the reduced form (Fig. 6B). In both oxidation states, the spectra of the de novo proteins resemble those of natural heme proteins (Table 2).

The most prominent feature in the spectra of the oxidized proteins is the ν_4 band at $1,376\text{ cm}^{-1}$. This band is characteristic of a ferric heme species and is readily observed in the spectra of natural heme proteins (Table 2A). In addition to this prominent band, three smaller porphyrin marker bands are also observed. These bands, ν_3 , ν_2 , and ν_{10} , are sensitive to the spin state of the heme iron. For example, low spin adducts yield ν_3 frequencies above $1,500\text{ cm}^{-1}$, whereas high spin adducts yield ν_3 frequencies at or below $1,500\text{ cm}^{-1}$ (Table 2A). For proteins F and 86, ν_3 occurs at $1,507\text{ cm}^{-1}$, suggesting that the iron in these de novo proteins is 6-coordinate and low spin. Likewise, the frequencies of the ν_2 and ν_{10} bands in the spectra of F and 86 demonstrate that these proteins form heme complexes that are 6-coordinate and low spin (Table 2A).

The resonance Raman spectra of proteins F and 86 in the reduced form are also informative (Fig. 6B). Reduction of the complex shifts the prominent ν_4 band to lower frequencies, exactly as seen for the natural heme proteins (Fig. 6; Table 2). The ν_3 band is split into two bands in the spectra of the reduced proteins (Fig. 6B). Although these bands are weak and somewhat difficult to characterize, the splitting suggests the presence of both 6-coordinate, low spin, and 5-coordinate, high spin complexes (Table 2B). A likely explanation for this result would be laser-induced photolability of one of the axial ligands. Although such photolability is unusual in resonance Raman studies of heme proteins, it is not unprecedented and has been observed for cytochrome *c* peroxidase (Smulevich et al., 1989) and for guanylyl cyclase (Yu et al., 1994).

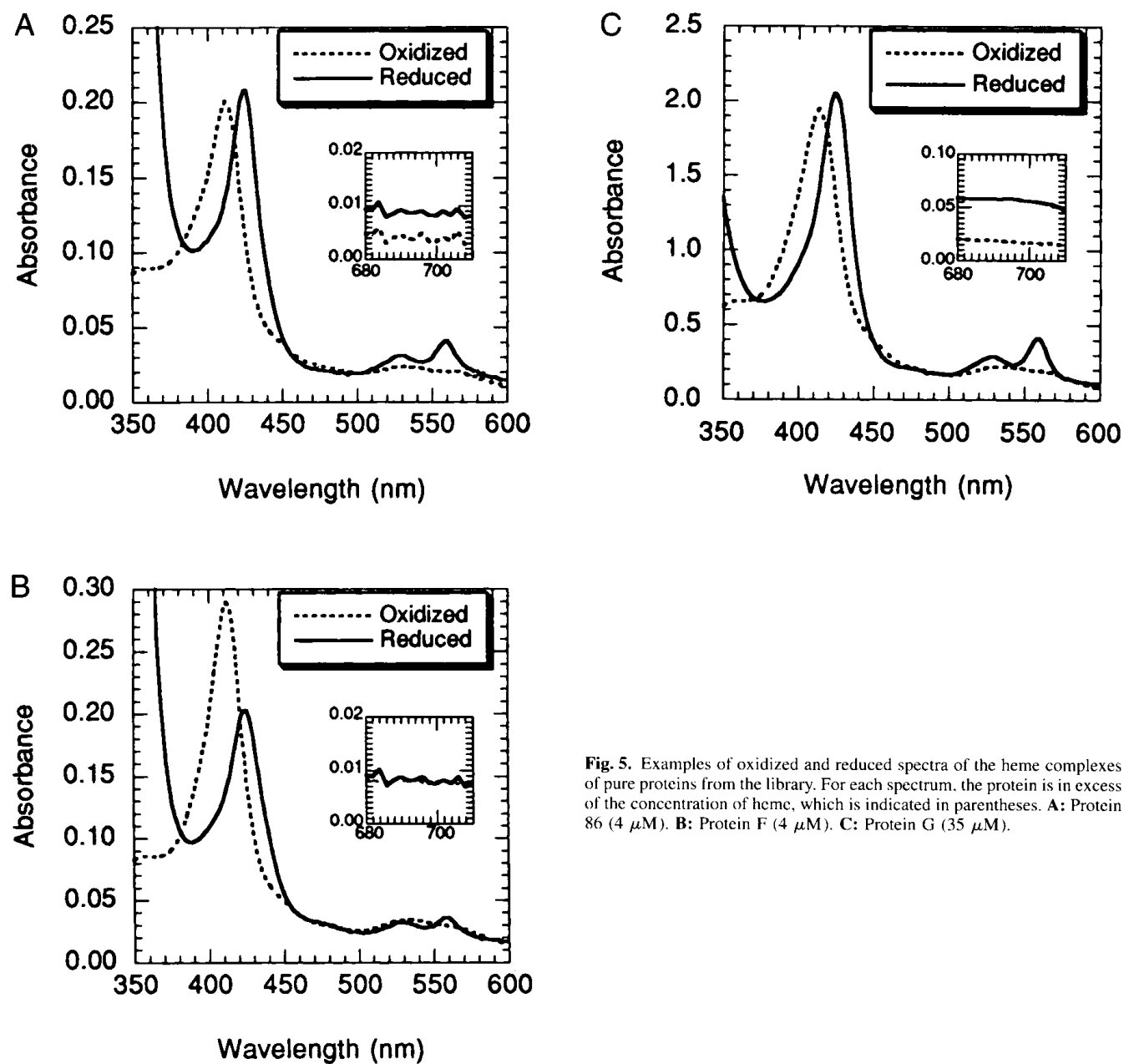


Fig. 5. Examples of oxidized and reduced spectra of the heme complexes of pure proteins from the library. For each spectrum, the protein is in excess of the concentration of heme, which is indicated in parentheses. A: Protein 86 ($4 \mu\text{M}$). B: Protein F ($4 \mu\text{M}$). C: Protein G ($35 \mu\text{M}$).

The ν_{11} bands for proteins F and 86 occur at $1,558$ and $1,554 \text{ cm}^{-1}$, respectively. These frequencies are similar to the value of $1,551 \text{ cm}^{-1}$ seen for Met/His ligation in cytochrome *c*, and different from the values below $1,540 \text{ cm}^{-1}$ typically seen for bis-His ligation (Table 2B). The locations of the ν_2 and ν_{10} bands in the spectra of the reduced de novo proteins are also consistent with the Met/His ligation seen in cytochrome *c* (Table 2B).

In summary, the resonance Raman spectroscopy, like the absorption spectroscopy, demonstrates that the de novo proteins form complexes with the heme cofactor in which the central iron is low spin and 6-coordinate. In addition, the resonance Raman studies suggest that the two axial ligands are methionine and histidine.

Heme affinity

Heme-binding titration curves were measured for several of the de novo proteins. To obtain apparent dissociation constants (K_{app}) from these titration curves, two assumptions were made. First, we neglected the tendency of heme to aggregate (White, 1978), and assumed that all heme exists in one of two states: either bound by protein or free in solution. A consequence of this assumption is that the true affinity of the proteins for heme should be greater than the measured affinity. The second assumption is that heme and protein form 1:1 complexes.

Of the proteins studied thus far, protein 86 has the greatest affinity for heme, with a K_{app} of $7 \times 10^{-7} \text{ M}$ (Fig. 7A). The

Table 1. Comparison of the visible absorption bands (nm) of the heme complexes of the de novo proteins with several natural heme proteins

Protein	Ligation	Reduced			Oxidized			695
		Soret	β	α	Soret	β	α	
Protein 86	?	426	530	560	412	~530		
Protein F	?	424	530	560	412	~530		
Protein G	?	424	530	560	414	~530		
<i>a</i> -Type cytochromes ^a		439–443	Absent	592–604	414–431		~600	
<i>b</i> -Type cytochromes ^a		408–449	526–546	555–567	400–419	520–530		
cyt <i>b</i> ₅₆₂ ^b	His/Met	427	531.5	562	418	~530	695	
<i>b</i> ₅ microsome ^c	His/His	423	527	555	410	~530		
<i>b</i> ₅ mitochond ^d	His/His	423	525	556	412	~530		
<i>c</i> -Type cytochromes ^a		415–423	521–527	550–558	390–415	520–530		
Horse heart cyt <i>c</i> ^{h,e}	His/Met	415	521	550	411	~530	695	
Myoglobin ^e	His	434		556				
Mb (O ₂)	His/O ₂	418	543	581				
Mb (CO)	His/CO	423	542	579				
Mb ⁻ (H ₂ O)	His/H ₂ O				409.5	505	635	
Hemoglobin ^e	His	430		555				
Hb (O ₂)	His/O ₂	415	541	577				
Hb (CO)	His/CO	419	540	569				
Hb ⁻ (H ₂ O)	His/H ₂ O				405	500	631	
Hb ⁺ (imidazole)	His/imidazole				411	534	560	

^aData for the classes of cytochromes are taken from Lemberg and Barrett (1973); Kirchenbaum (1974); Dickerson and Timkovitch (1975); Moore and Pettigrew (1990); and Yamanaka (1992).

^bLemberg and Barrett (1973).

^cBeck von Bodman et al. (1986).

^dRivera et al. (1992).

^ePettigrew and Moore (1987).

^fAntonini and Brunori (1971). Data are for sperm whale myoglobin, reduced human hemoglobin, and oxidized horse hemoglobin.

binding curve obtained at higher concentrations supports a 1:1 binding stoichiometry (Fig. 7B). For protein G, the K_{app} is 2×10^{-6} M (Fig. 7C). For protein F, the titration curve can be fit to a K_{app} of 1×10^{-5} M (Fig. 7D), but titrations done at higher concentrations of heme yield different values and the curves do not reach saturation, indicating the presence of additional weak heme binding sites.

Effect of heme binding on protein structure and stability

Size-exclusion chromatography using Superdex 75 (Pharmacia) demonstrates that the apo forms of proteins F, G, and 86 are monomeric (S. Roy & M.H. Hecht, unpubl.). The apo form of protein F has also been shown to be monomeric by equilibrium ultracentrifugation (Kamtekar, 1995). Sizing studies of the heme-bound proteins have proved difficult because (1) high-resolution gel filtration resins (e.g., Superdex 75) strip heme from noncovalent protein complexes; and (2) ultracentrifugation studies are complicated by background absorbance due to free or aggregated heme (Kamtekar, 1995).

The heme-bound forms of the proteins were compared to the apo forms by size-exclusion chromatography using Sephadex G-50. This resin does not provide as high a level of resolution as Superdex 75, but was used nonetheless because it does not strip heme as extensively. However, because some stripping does occur, the data

are difficult to interpret. For protein 86, which has the highest heme affinity of the proteins studied thus far, the heme-bound form elutes slightly earlier than the apo form. Thus, heme binding clearly does not promote large-scale aggregation. The slightly earlier elution time of the heme-bound form suggests either that the protein expands upon binding, or that binding induces dimerization. The latter interpretation is less likely because (1) the molecular weight estimated for the heme-bound form is less than double that of the apo form (Simons, 1996); and (2) the stoichiometry of binding for protein 86 is 1:1 (see above and Fig. 7B). In the overall collection of binary code heme proteins, we expect to observe a range of behaviors. Some sequences will form dimers upon heme binding (via His/His or His/Met ligation from two chains), whereas others will bind heme as monomeric proteins.

CD spectroscopy was used to determine the effect of heme binding on the secondary structure of the de novo proteins. The CD spectra of proteins 86 and G are shown in Figure 8. Both in the absence and in the presence of heme, these proteins are predominantly α -helical, displaying characteristic minima at 208 nm and 222 nm. Formation of the heme complex does not affect the α -helical content of either protein appreciably.

The thermal stabilities of proteins 86 and G were measured in the presence and absence of heme (Fig. 9). The thermal denaturation of protein 86 is cooperative and sigmoidal (Fig. 9A). This suggests that the folded structure of this protein may possess native-

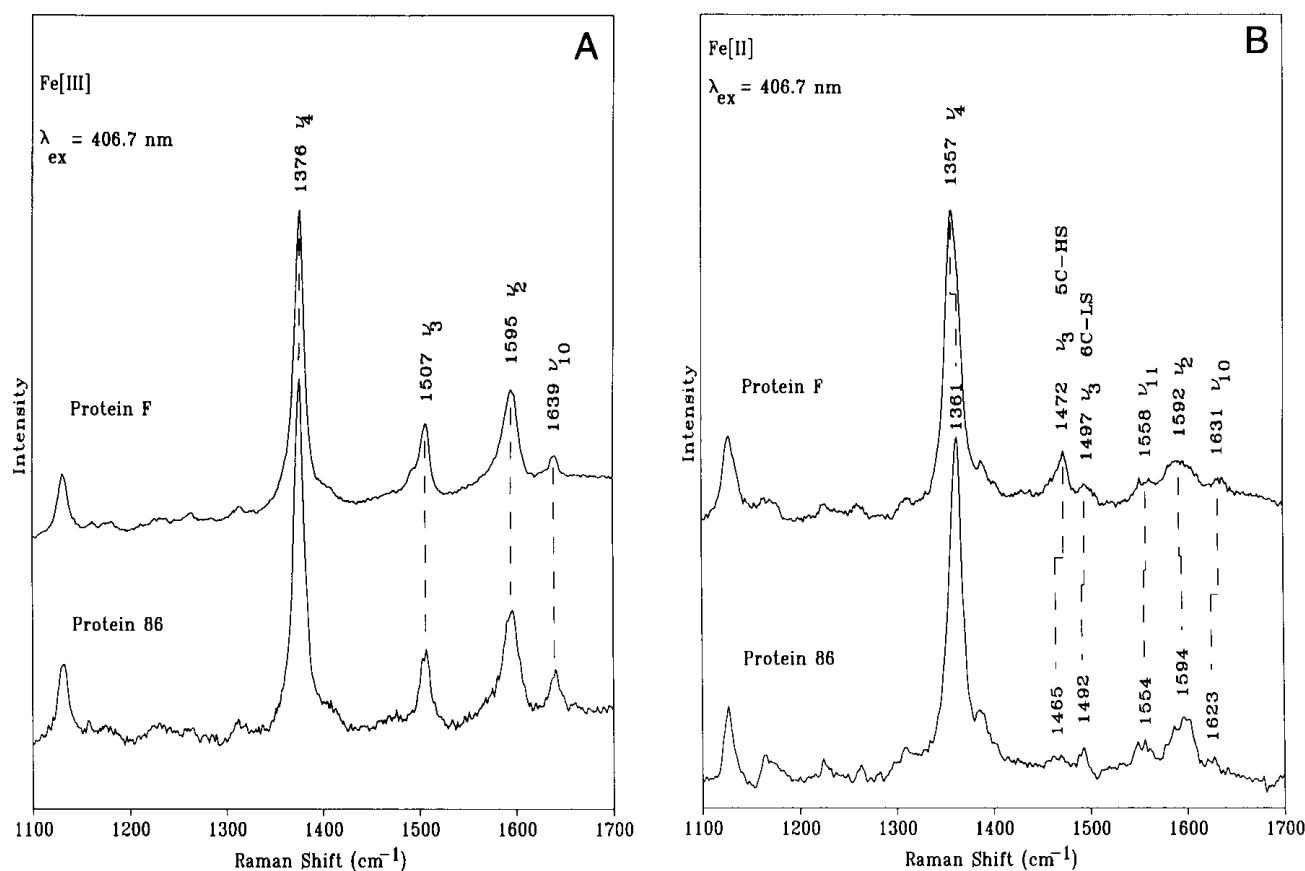


Fig. 6. Mid-frequency resonance Raman spectra of the complexes of proteins 86 and F with mesoporphyrin IX. A: Oxidized. B: Reduced.

like properties both in the absence (S. Roy & M.H. Hecht, unpubl.) and in the presence of heme. Conversely, the thermal denaturation of protein G is not cooperative, suggesting that this protein is less native-like. These results suggest that in some of the de novo proteins, the heme binding site may be pre-organized, whereas in others it probably is not.

Binding of the heme cofactor enhances the thermal stabilities of both proteins (Fig. 9). For the cooperative sigmoidal denaturation of protein 86, the addition of heme increases the melting temperature by 8 °C.

The α -helical CD spectra of the heme complexes, coupled with the observation that heme enhances the stability of the structure responsible for these CD spectra, indicates that heme binds to the folded α -helical conformations of these de novo proteins.

Discussion

Binding studies previously performed on libraries of proteins have been limited to either short peptides (Scott & Smith, 1990) or naturally occurring proteins with sections of randomized sequence (Roberts et al., 1992). Thus, although cofactor binding studies on libraries of RNA molecules have begun to reveal possible mechanisms of evolution in the RNA world (Sassanfar & Szostak, 1993; Lorsch & Szostak, 1994), our picture of the early evolution of proteins is much less detailed. For example, it would be interesting to determine what fraction of a pool of "foldable" sequences can

also bind cofactors such as heme. The results reported here represent an initial attempt to answer such questions.

In our screen, we found that 15 of the 30 sequences in our binary code collection exhibit some degree of heme binding. This high frequency is rather surprising. Earlier attempts to devise de novo heme proteins have been based on the explicit design of the binding site (Choma et al., 1994; Robertson et al., 1994; Rabanal et al., 1996; Gibney et al., 1997). In contrast, the binary code sequences described here are based on global design strategies, and were not designed explicitly to bind heme. Nonetheless, half the binary code sequences bind heme. Moreover, protein 86 binds with the same dissociation constant ($K_{app} = 7 \times 10^{-7}$ M) as that reported by Choma et al. (1994) for their rationally designed heme protein, VAVH₂₅(S-S). These findings highlight the power of the combinatorial binary code strategy, and demonstrate that isolating de novo heme proteins need not require explicit design of the cofactor binding site.

Our finding that sequences not designed explicitly to bind heme may nonetheless do so is not unprecedented—Choma et al. (1994) found that a control peptide displayed heme binding properties that were superior to those of a peptide explicitly designed to bind heme. The control peptide (called "retro" because it had been synthesized with the reverse sequence) bound heme more tightly and yielded a pinker color (compared to yellow-pink) than the designed peptide. Furthermore, like natural cytochromes, although in contrast to the designed peptide, the retro peptide gave rise to

Table 2. Comparison of the resonance Raman frequencies for porphyrin skeletal modes for the de novo proteins F and 86 and several heme models^a

A. Oxidized form							
Sample	Ligation	ν_4	ν_3	ν_2	ν_{10}	Reference	
Protein F (MPIX)	?	1,376	1,507	1,595	1,639	This work	
Protein 86 (MPIX)	?	1,376	1,507	1,595	1,639	This work	
6-Coordinate, low spin							
MPIX(Im) ₂	His/His	1,375	1,505	1,599	1,640	Spiro and Burke (1976)	
PPIX(Im) ₂	His/His	1,373	1,502	1,579	1,640	Choi et al. (1982)	
Cyt <i>b</i> ₂	His/His	1,374	1,505	1,579	1,640	Desbois et al. (1989)	
Cyt <i>b</i> ₅	His/His	1,372	n.a.	1,585	1,639	Kitagawa et al. (1976)	
Cyt <i>c</i>	Met/His	1,371	1,501	1,585	1,635	Hu et al. (1993)	
Met ₈₀ Cys Cyt <i>c</i>	Cys/His	1,376	1,503	1,589	1,636	Smulevich et al. (1994)	
6-Coordinate, high spin							
Hb(F)	His/F	1,373	1,482	1,555	1,608	Spiro and Streckas (1974)	
5-Coordinate, high spin							
MPIX(F)	F	1,374	1,495	1,572	1,632	Spiro and Burke (1976)	
HRP	His	1,375	1,500	1,575	1,630	Rakshit and Spiro (1974)	
B. Reduced form							
Sample	Ligation	ν_4	ν_3	ν_{11}	ν_2	ν_{10}	Reference
Protein F (MPIX)	?	1,357	1,472/1,497	1,558	1,592	1,631	This work
Protein 86 (MPIX)	?	1,361	1,465/1,492	1,554	1,594	1,623	This work
6-Coordinate, low spin							
MPIX(Im) ₂	His/His	1,358	1,490	1,537	1,600	1,620	Spiro and Burke (1976)
PPIX(Im) ₂	His/His	1,359	1,493	1,539	1,584	1,617	Choi et al. (1982)
Cyt <i>b</i> ₂	His/His	1,361	1,493	1,537	1,580	1,620	Desbois et al. (1989)
Cyt <i>b</i> ₅	His/His	1,362	1,493	n.a.	1,585	1,617	Kitagawa et al. (1976)
Cyt <i>c</i>	Met/His	1,364	1,496	1,551	1,596	1,626	Hu et al. (1993)
Met ₈₀ Cys Cyt <i>c</i>	Cys/His	1,360	1,494	1,537	1,592	1,625	Smulevich et al. (1994)
5-Coordinate, high spin							
Deoxy Hb	His	1,358	1,473	1,546	1,557	1,607	Spiro and Streckas (1974)
HRP	His	1,358	1,472	n.a.	1,553	1,605	Spiro and Burke (1976)
PPIX(2-MeIm)	2-MeIm	1,357	1,471	1,547	1,562	1,604	Choi et al. (1982)

^aFrequencies are in cm⁻¹. MPiX, mesoporphyrin IX; PPIX, protoporphyrin IX; Im, imidazole; Cyt *b*₂, flavocytochrome *b*₂; Hb, hemoglobin; HRP, horseradish peroxidase; 2-MeIm, 2-methylimidazole.

pronounced splitting of the α/β peaks in the visible spectrum (Choma et al., 1994).

Among the proteins we have purified and examined thus far, we observe a range of binding profiles. Diversity in the properties of these proteins is to be expected; it reflects the enormous diversity in the sequences generated by the combinatorial methods of the binary code strategy. Some of the de novo sequences exhibit complex binding curves (e.g., protein 13; data not shown). However, several sequences, such as 86 and G, appear to bind heme stoichiometrically. In these cases, the apparent dissociation constants are on the order of 10⁻⁶–10⁻⁷ M, close to the $K_d = 0.9 \times 10^{-7}$ M estimated for the noncovalent complex of Fe(II) heme with apocytocrome *c* (Dumont et al., 1994).

The three-dimensional structures of our proteins have not been determined. Therefore, we cannot assess whether our proteins indeed contain the appropriate number and orientation of helices that define a four-helix bundle. Nor, in the absence of such structures, can we determine the precise extent to which heme binding may

change the fold of our proteins. Nevertheless, a model of protein 86 (which has the highest heme affinity of the sequences examined thus far) shows that heme binding in our library can occur within a four-helix bundle in a manner consistent with the experimental data. When protein 86 is modeled as a four-helix bundle, one His/His pair and six possible His/Met pairs can occur in the same layer of the bundle (Fig. 10A). However, a head-on view of the four-helix bundle (Fig. 10B,C) shows that, for all but one of these pairs, the ligating atoms (S of Met and N ϵ of His) are too far apart to ligate the iron. The only His/His or His/Met pair close enough to ligate the iron is His₃₁/Met₄₆ (Fig. 10C). (Although His₃₁/Met₁ makes a contact in Fig. 10B, this pair was excluded by the analysis in Fig. 10A.) Binding to iron by His₃₁ and Met₄₆ is consistent with the resonance Raman data, which suggests His/Met ligation.

By using His₃₁ and Met₄₆ to ligate the heme iron, a binding site was modeled between the second and third helices of the bundle. These two helices were superimposed onto the first and fourth helices of cytochrome *b*₅₆₂ (which ligates its heme iron via Met₇

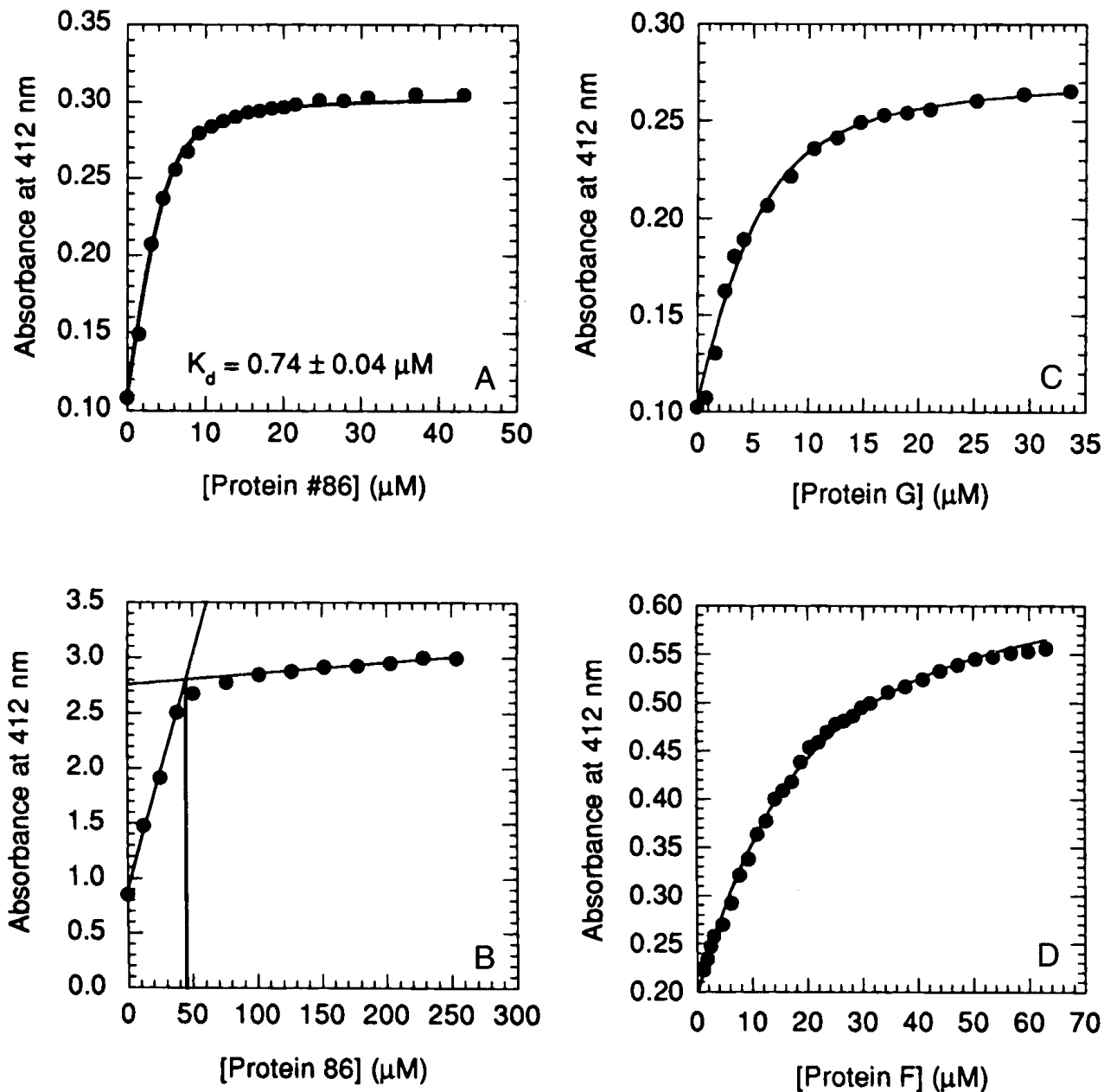


Fig. 7. Titration curves. **A:** Protein 86 into 5.1 μM heme. Solid line shows the fit at $K_{app} = 7.4 \pm 0.4 \times 10^{-7}$ M assuming 1:1 binding. **B:** Protein 86 into 40 μM heme; 1:1 proteins/heme from the breakpoint of the curve. **C:** Protein G into 5.1 μM heme. Solid line shows the fit at $K_{app} = 2.0 \pm 0.1 \times 10^{-6}$ M assuming 1:1 binding. **D:** Protein F into 9.7 μM heme. Solid line shows the fit for $K_{app} = 1.2 \pm 0.02 \times 10^{-5}$ M assuming 1:1 binding. A titration using a higher concentration of heme (40 μM) yields $K_{app} = 4.5 \times 10^{-5}$ M, but also does not reach saturation (even at a 5.5 molar excess of protein over heme). Therefore, the extinction coefficient of the heme protein complex was estimated from curve fitting.

and His₁₀₂). The remaining helices were then placed in positions mimicking the four-helix bundle structure of cytochrome *b*₅₆₂ (Hamada et al., 1995). The resulting model (Fig. 11) demonstrates that, although the binary code places methionine residues on the nonpolar face and histidine residues on the polar face of the helices, a His/Met pair at the helix-helix interface can participate in heme binding.

The 30 proteins screened thus far represent an extremely small fraction of the total number of sequences that can be encoded by

the binary patterning used for this structural motif. (24 buried and 32 surface positions, chosen from five nonpolar and six polar residues, respectively, can yield $5^{24} \times 6^{32} = 5 \times 10^{41}$ possible sequences; Kamtekar et al., 1993). Because, among these few proteins, we have observed several that bind with moderate affinities, it seems likely that the far larger set of binary code sequences that adopt this fold probably contains many proteins with high affinities for heme.

Natural heme proteins presumably evolved from simpler precursors, which at some point in evolution developed the ability to

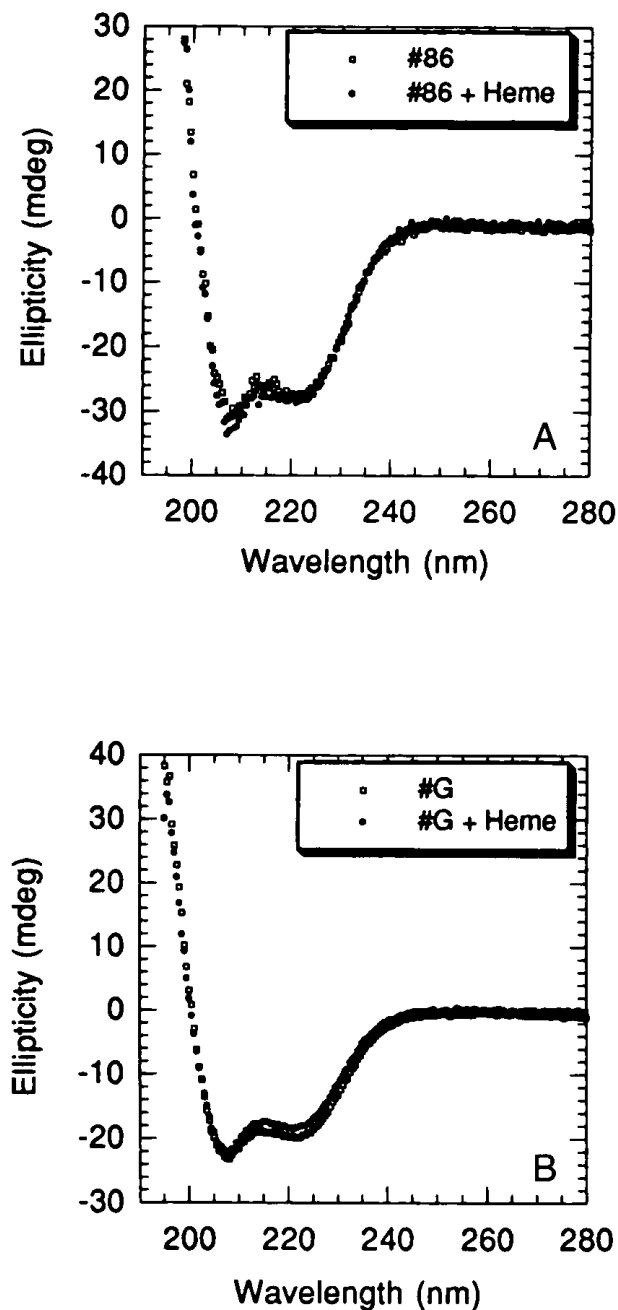


Fig. 8. CD spectra of the proteins in the presence and absence of heme. A: Protein 86. B: Protein G.

bind heme and use it as a functional cofactor. The binary code approach, in some measure, mimics this approach on the laboratory time scale. Indeed, the existence of the binary code distribution within the genetic code suggests a possible mechanism for the early evolution of proteins. Thus, our results indicate that obtaining sequences capable of folding into globular structures and binding specific cofactors may be considerably easier than expected.

Such a conclusion also has implications for de novo protein design. For the eventual design of functionally active de novo proteins, the binding of catalytically relevant cofactors represents

an appealing first step. Our finding that half the sequences in our initial collection are capable of binding the heme cofactor suggests that this first step is not a difficult one. It remains to be seen whether any of the proteins in this collection are catalytically active. However, we anticipate that, among the enormous number of heme binding proteins that are likely to occur in larger collections of binary code sequences, it is not unlikely that some will possess catalytic activity.

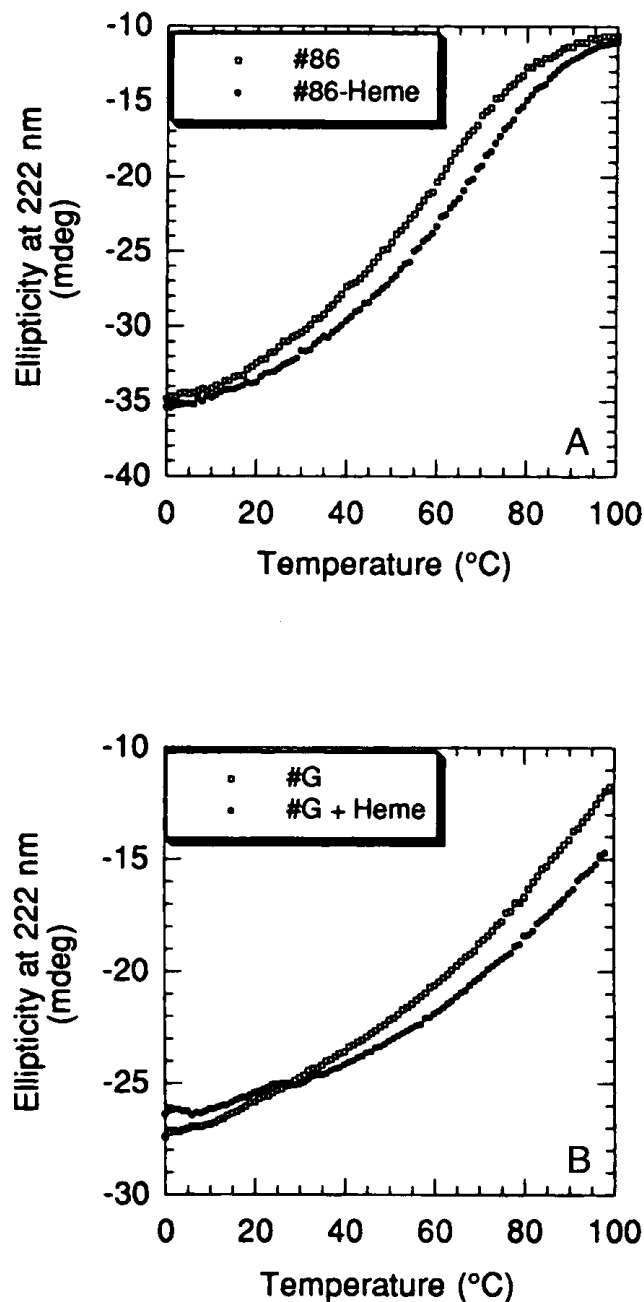


Fig. 9. Thermal denaturation profiles of the proteins in the presence and absence of heme, as monitored by the CD signal at 222 nm. A: Protein 86. B: Protein G.

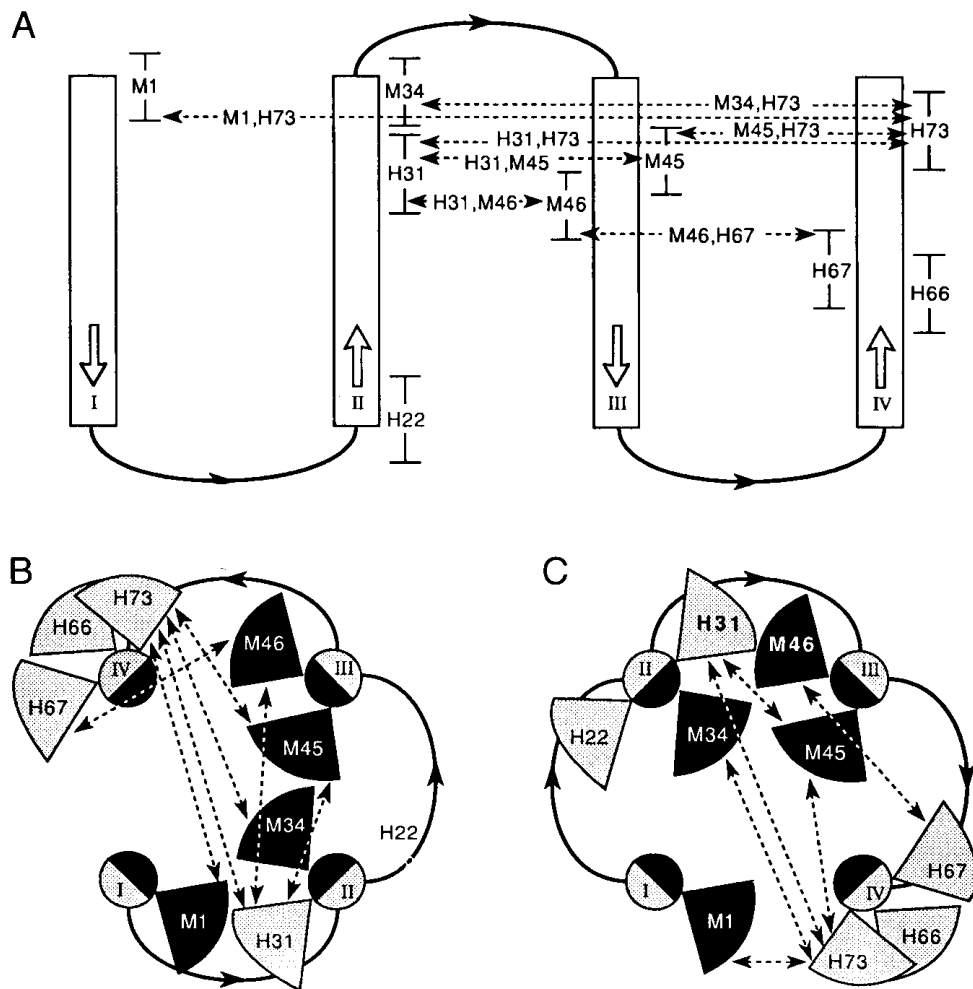


Fig. 10. **A:** Possible positions of His and Met residues along the α -helices. Uncertainty in positions arises from consideration of all allowable side-chain rotamers. Dashed horizontal arrows connect His/His and His/Met pairs that occur in the same layer of the bundle and could potentially bind heme. **B,C:** Head-on representations of the four-helix bundle in either left-turning (**B**) or right-turning (**C**) topologies (Efimov, 1993). α -Helices are represented by circles with the light and dark sides indicating polar and nonpolar faces, respectively. Pie-shaped sectors represent possible locations of an Fe atom bound either to the N ϵ of His or the S of Met. Uncertainty in positions arises from consideration of all allowable side-chain rotamers. Contact between two pie shapes indicates that the two residues can bind a single Fe atom between them. Of the six His/His and His/Met pairs shown in **A**, only His₃₁/Met₄₆ forms such a contact.

Materials and methods

Screening for heme binding

Proteins were expressed in *E. coli* strain X90(DE3) in 250 mL 2 \times YT medium and harvested as described previously (Kamtekar et al., 1993). Proteins were extracted from the cells using a freeze-thaw protocol (Johnson & Hecht, 1994) and resuspended in 5 mL of 100 mM MgCl₂ (Roy et al., 1997a). The freeze-thaw extracts were partially purified by acid precipitation of contaminants at pH 4.0 in 50 mM sodium acetate buffer. (Proteins 17, 21, and 49 were not acidified because of their low solubility at low pH.) The supernatant was dialyzed into 50 mM sodium phosphate, 200 mM NaCl at pH 7.0 and concentrated (Centricon, Pharmacia) to a final volume of 100 μ L. Total protein concentration was estimated using the Bio-Rad Protein Assay with BSA as a standard. The fraction of designed protein was approximated by Coomassie-stained SDS-PAGE gels analyzed using NIH Image 1.60.

These semi-pure samples were screened for heme binding by titrating concentrated protein solution into solutions of 5–30 μ M bovine hemin chloride (Sigma) in 50 mM sodium phosphate, 200 mM NaCl, pH 7.0. Titrations were carried to an endpoint at which protein was estimated to be in excess of heme by at least threefold. Solutions were equilibrated for 5 min after each addition of protein, and the spectrum from 230 to 800 nm was then measured on an HP 8452A diode array spectrophotometer using a cuvette with a 1-cm pathlength.

Protein purification

Chromatographic purification was accomplished by methods similar to those described previously (Kamtekar et al., 1993; Johnson & Hecht, 1994; Roy et al., 1997a). Supernatants from the acid precipitation (see above) were loaded onto an S-sepharose column (Sigma) in 50 mM sodium acetate, pH 4.0, and protein

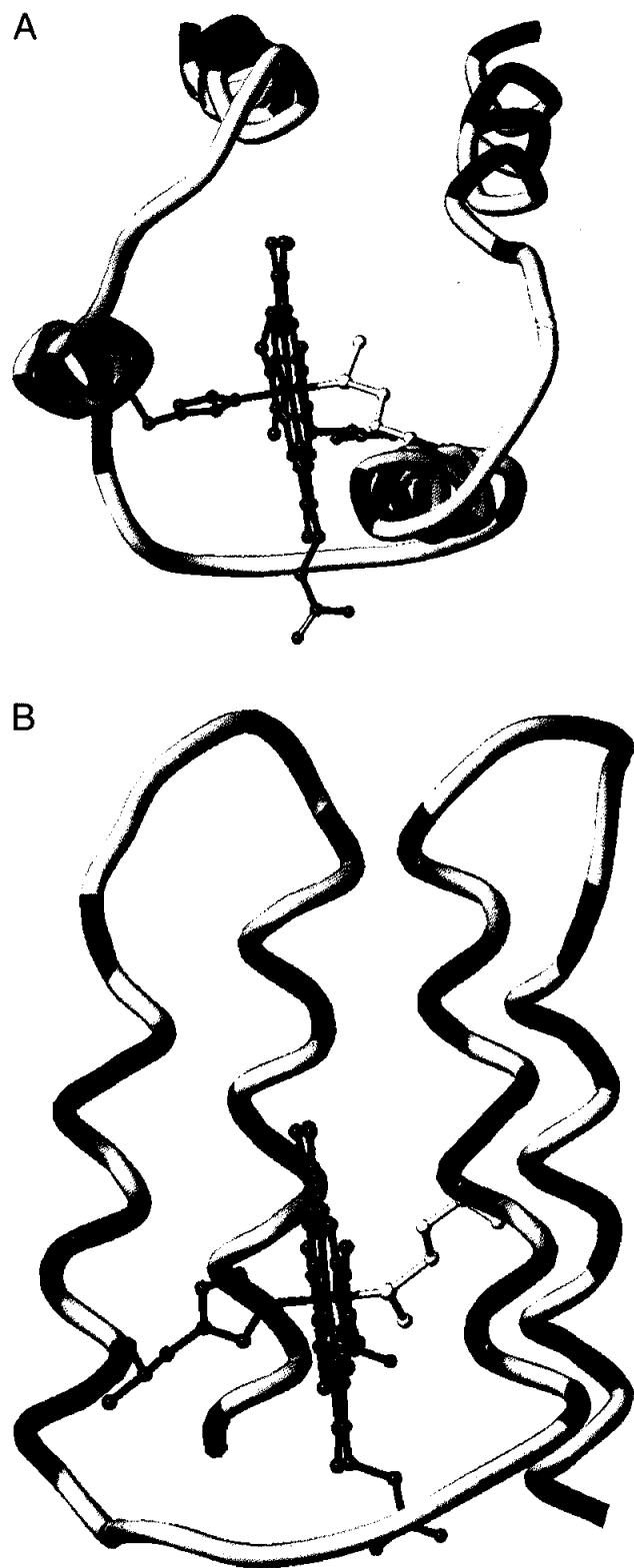


Fig. 11. Top and side view of a molecular model of the heme complex of protein 86. The protein backbone is shown with polar residues in blue, nonpolar residues in yellow, loop regions in gray, and the heme group in red. His₃₁, Met₄₆, and heme are shown in ball and stick representation. The sequence of protein 86 is MGKLNLDLLEDLQEVKGPDSGHLQN VIEDIHDFMQGPGSGGKIQEMMKEFQQVLDGPRSGNIKEIF HHLEELVHR.

was eluted by a linear gradient of NaCl (0–2 M). Final purification was achieved by size-exclusion chromatography (Superdex 75, Pharmacia). Protein concentrations were determined by amino acid analysis.

Optical spectra, binding stoichiometry, and determination of dissociation constant

Dissociation constants were determined by titrating protein into heme solutions containing 50 mM Tris, 50 mM NaCl, pH 8.0. Under these conditions, aggregation of heme is decreased (Hargrove et al., 1996). Absorbance was monitored at the Soret maximum (412 or 414 nm) and fitted to the equation for the dissociation constant assuming 1:1 binding using Kaleidagraph 3.0 for the Macintosh. Binding stoichiometry was derived from similar experiments at high heme concentrations, such that the breakpoint of the curve determines the ratio of heme to protein in the complex.

Visible spectra were measured at the end of similar titrations. Spectra of reduced samples were measured following addition of a minimum volume of an aqueous solution of sodium dithionite into deoxygenated solutions of the heme protein complex. All spectra were taken on an HP8452A diode array spectrophotometer using a cuvette with a 1-cm pathlength.

Heme concentration was determined spectrophotometrically by forming the reduced CO-complex of an aliquot of the hemin chloride stock solution and using an extinction coefficient of $147 \text{ mM}^{-1} \text{ cm}^{-1}$ at 407 nm (Hargrove et al., 1996).

The extinction coefficient for the Soret band of protein 86 in the oxidized form was estimated from the data in Figure 7B as approximately $75 \text{ mM}^{-1} \text{ cm}^{-1}$. For comparison, the extinction coefficient for cytochrome *b*₅₆₂ is $117 \text{ mM}^{-1} \text{ cm}^{-1}$ (Itagaki & Hager, 1966).

Resonance Raman spectroscopy

Resonance Raman spectra were measured using both hemin and iron mesoporphyrin IX (a derivative in which the heme vinyl groups are replaced by ethyl groups). The two porphyrin derivatives yielded similar results. Spectra from the mesoporphyrin experiments are shown because peak assignments are easier to interpret in the absence of the $\nu(\text{C}=\text{C})$ due to the hemin vinyl groups. Solutions contained $25 \mu\text{M}$ protein in 50 mM sodium phosphate, 200 mM NaCl, pH 6.8. The reduced form of the heme protein was prepared by adding a minimum volume of an aqueous solution of sodium dithionite into deoxygenated solutions of protein. Samples were excited using the 406.7 nm line of a Kr⁺ laser (Coherent Radiation). Back scattered light was focused onto the sample in a spinning NMR tube using a cylindrical lens to minimize photo-decomposition. The light was dispersed with a Spex triplemate monochromator grating (2,400 grooves/mm) and collected with an optical multichannel analyzer (Princeton Instruments). All spectra were calibrated with indene as a standard. The frequencies are accurate to $\pm 1 \text{ cm}^{-1}$ for isolated bands. Spectra were imported into and processed with Labcalc software (Galactic Industries Corp).

Protein secondary structure and stability

CD spectra were measured on an Aviv 62DS CD spectropolarimeter. Proteins were dissolved in 50 mM sodium phosphate, 200 mM NaCl, pH 7.0. Samples were in a 0.1-cm cuvette. Spectra

were recorded at 20 °C from 280 nm to 195 nm, and the signal was averaged over 5 s for steps of 1 nm.

Thermal denaturation curves were obtained by monitoring the ellipticity at 222 nm at 1 °C increments from 0 to 95 °C. Temperature equilibration time was 2 min and the signal was averaged over an additional minute. Reversibility was demonstrated by cooling the samples to 20 °C and recording their CD spectra.

Molecular modeling

Molecular models were generated using Sybyl (Tripos, Inc.). To construct the model in Figure 11, helical regions ($\phi = -58^\circ$, $\psi = -47^\circ$, $\omega = 180^\circ$) were first generated for the binary code sequences. Helices II and III were then docked to accommodate heme binding by setting the His₅₁ N ϵ -Fe bond distance as 2.0 ± 0.1 Å, the Met₄₆ S-Fe bond distance as 2.4 ± 0.1 Å, and the N-Fe-S bond angle as $180 \pm 5^\circ$. These helices were superimposed onto helices 1 and 4 of cytochrome *b*₅₆₂ (which provide the heme-binding axial ligands, Met₇ and His₁₀₂) and the remaining helices were placed in positions mimicking the four-helix bundle structure of cyt *b*₅₆₂ (Hamada et al., 1995). The resulting model was energy minimized using the Tripos Force Field with Pullman charges.

Acknowledgments

Supported by the Whitaker Foundation and by a Beckman Foundation Young Investigator Award to M.H.H.

References

- Antonini E, Brunori M. 1971. *Hemoglobin and myoglobin in their reactions with ligands*. Amsterdam: North-Holland Publishing Company.
- Beck von Bodman S, Schuler MA, Jollie DR, Sligar SG. 1986. Synthesis, bacterial expression, and mutagenesis of the gene coding for mammalian cytochrome *b*₅. *Proc Natl Acad Sci USA* 83:9443–9447.
- Choi S, Spiro TG, Langry KC, Smith KM, Budd DL, La Mar GN. 1982. Structural correlations and vinyl influences in resonance Raman spectra of protoheme complexes and proteins. *J Am Chem Soc* 104:4345–4351.
- Choma CT, Lear JD, Nelson MJ, Dutton PL, Robertson DE, DeGrado WF. 1994. Design of a heme-binding four-helix bundle. *J Am Chem Soc* 116:856–865.
- Desbois A, Tegoni M, Gervais M, Lutz M. 1989. Flavin and heme structures in lactate:cytochrome *c* oxidoreductase: A resonance Raman study. *Biochemistry* 1989:8011–8022.
- Dickerson RE, Timkovitch R. 1975. Cytochromes *c*. In: Boyer PD, ed. *The enzymes: XI. Oxidation–reduction, part A*. New York: Academic Press. pp 397–547.
- Dolphin D, ed. 1979. *The porphyrins. VII: Biochemistry, part B*. New York: Academic Press.
- Dumont ME, Corin AF, Campbell GA. 1994. Noncovalent binding of heme induces a compact apocytochrome *c* structure. *Biochemistry* 33:7368–7378.
- Efimov AV. 1993. Patterns of loop regions in proteins. *Curr Opin Struct Biol* 3:379–384.
- Finzel BC, Weber PC, Hardman KD, Salemme FR. 1985. Structure of ferri-cytochrome *c'* from *Rhodospirillum rubrum* at 1.67 Å resolution. *J Mol Biol* 186:627–643.
- Gibney BR, Rabanal F, Skalicky JJ, Wand AJ, Dutton PL. 1997. Design of a unique protein scaffold. *J Am Chem Soc* 119:2323–2324.
- Hamada K, Bethge PH, Mathews FS. 1995. Refined structure of cytochrome *b*₅₆₂ from *Escherichia coli* at 1.4 Å resolution. *J Mol Biol* 247:947–962.
- Hargrove MS, Barrick D, Olson JS. 1996. The association rate constant for heme binding to globin is independent of protein structure. *Biochemistry* 35:11293–11299.
- Hu S, Morris IK, Singh JP, Smith KM, Spiro TG. 1993. Complete assignment of cytochrome *c* resonance Raman spectra via enzymatic reconstitution with isotopically labeled hemes. *J Am Chem Soc* 115:12446–12458.
- Itagaki E, Hager LP. 1966. Studies on cytochrome *b*₅₆₂ of *Escherichia coli*. *J Biol Chem* 241:3687–3695.
- Johnson BH, Hecht MH. 1994. Recombinant proteins can be released from *E. coli* cells by repeated cycles of freezing and thawing. *Biotechnology* 12:1357–1360.
- Kamtekar S. 1995. A binary code for protein design [thesis]. Princeton, New Jersey: Princeton University.
- Kamtekar S, Schiffer JM, Xiong H, Babik JM, Hecht MH. 1993. Protein design by binary patterning of polar and nonpolar amino acids. *Science* 262:1680–1685.
- Kirichenbaum DM. 1974. *Atlas of protein spectra in the ultraviolet and visible regions, vol. 2*. New York:IFI/Plenum.
- Kitagawa T, Kyogoku Y, Iizuka T, Saito MI. 1976. Nature of the iron–ligand bond in ferrous low spin hemoproteins studied by resonance Raman scattering. *J Am Chem Soc* 98:5169–5173.
- Lemberg R, Barrett J. 1973. *Cytochromes*. London: Academic Press.
- Lorsch JR, Szostak JW. 1994. In vitro selection of RNA aptamers specific for cyanocobalamin. *Biochemistry* 33:973–982.
- Mathews FS, Bethge PH, Czerwinski EW. 1979. The structure of cytochrome *b*₅₆₂ from *Escherichia coli* at 2.5 angstroms resolution. *J Biol Chem* 254:1699–1706.
- Moore GR, Pettigrew GW. 1990. *Cytochromes c: Evolutionary, structural and physicochemical aspects*. Berlin/Heidelberg: Springer-Verlag.
- Pettigrew GW, Moore GR. 1987. *Cytochromes c: Biological aspects*. Berlin Heidelberg: Springer-Verlag.
- Rabanal F, DeGrado WF, Dutton PL. 1996. Toward the synthesis of a photosynthetic reaction center maquette: A cofacial porphyrin pair assembled between 2 subunits of a synthetic four-helix bundle multiheme protein. *J Am Chem Soc* 118:473–474.
- Rakshit G, Spiro TG. 1974. Resonance Raman spectra of horseradish peroxidase: Evidence for anomalous heme structure. *Biochemistry* 13:5317–5323.
- Rivera M, Barillas-Mury C, Christensen KA, Little JW, Wells MA, Walker FA. 1992. Gene synthesis, bacterial expression, and ¹H NMR spectroscopic studies of the rat outer mitochondrial membrane cytochrome *b*₅. *Biochemistry* 31:12233–12240.
- Roberts BL, Markland W, Ley AC, Kent RB, White DW, Guterman SK, Ladner RC. 1992. Directed evolution of a protein: Selection of potent neutrophil elastase inhibitors displayed on M13 fusion phage. *Proc Natl Acad Sci USA* 89:2429–2433.
- Robertson DE, Farid RS, Moser CC, Urbauer JL, Mulholland SE, Pidikiti R, Lear JD, Wand AJ, DeGrado WF, Dutton PL. 1994. Design and synthesis of multi-haem proteins. *Nature* 368:425–432.
- Roy S, Helmer KJ, Hecht MH. 1997a. Detecting native-like properties in combinatorial libraries of de novo proteins. *Folding & Design* 2:89–92.
- Roy S, Ratnaswamy G, Boice JA, Fairman R, McLendon G, Hecht MH. 1997b. A protein designed by binary patterning of polar and nonpolar amino acids displays native-like properties. *J Am Chem Soc* 119:5302–5306.
- Sassanfar M, Szostak JW. 1993. An RNA motif that binds ATP. *Nature* 364:550–553.
- Scott JK, Smith GP. 1990. Searching for peptide ligands with an epitope library. *Science* 249:386–390.
- Simons CT. 1996. Heme affinity in a library of de novo four-helix bundles [thesis]. Princeton, New Jersey: Princeton University.
- Smulevich G, Bjerrum MJ, Gary HB, Spiro TG. 1994. Resonance Raman spectra and the active site structure of semisynthetic Met80Cys horse heart cytochrome *c*. *Inorganic Chemistry* 33:4629–4634.
- Smulevich G, Miller MA, Gosztola D, Spiro TG. 1989. Photodissociable endogenous ligand in alkaline reduced cytochrome *c* peroxidase implicates distal protein tension. *Biochemistry* 28:9905–9908.
- Spiro TG, Burke JM. 1976. Protein control of porphyrin conformation: Comparison of resonance Raman spectra of heme proteins with mesoporphyrin IX analogues. *J Am Chem Soc* 98:5482–5489.
- Spiro TG, Streckas TC. 1974. Resonance Raman spectra of heme proteins: Effects of oxidation and spin state. *J Am Chem Soc* 96:338–345.
- White WJ. 1978. Aggregation of porphyrins and metalloporphyrins. In: Dolphin D, ed. *The porphyrins. V. Physical chemistry, part C*. New York: Academic Press.
- Yamanaka T. 1992. *The biochemistry of bacterial cytochromes*. Tokyo: Japan: Scientific Societies Press.
- Yu AE, Hu S, Spiro TG, Burstyn JN. 1994. Resonance Raman spectroscopy of soluble guanylyl cyclase reveals displacement of distal and proximal heme ligands by NO. *J Am Chem Soc* 116:4117–4118.