Heme proteins–Diversity in structural characteristics, function, and folding

ARTICLE *in* PROTEINS STRUCTURE FUNCTION AND BIOINFORMATICS · AUGUST 2010

Impact Factor: 2.63 · DOI: 10.1002/prot.22747 · Source: PubMed

3 AUTHORS, INCLUDING:

Abdullah Kahraman
University of Zurich
19 PUBLICATIONS **811** CITATIONS

PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Heme proteins—Diversity in structural characteristics, function, and folding

Lorna J. Smith,[1]* Abdullah Kahraman,[2] and Janet M. Thornton[2]

[1] Department of Chemistry, University of Oxford, Inorganic Chemistry Laboratory, Oxford OX1 3QR, United Kingdom
[2] European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SD, United Kingdom

ABSTRACT

The characteristics of heme prosthetic groups and their binding sites have been analyzed in detail in a data set of nonhomologous heme proteins. Variations in the shape, volume, and chemical composition of the binding site, in the mode of heme binding and in the number and nature of heme–protein interactions are found to result in significantly different heme environments in proteins with different functions in biology. Differences are also seen in the properties of the apo states of the proteins. The apo states of proteins that bind heme permanently in their functional form show some disorder, ranging from local unfolding in the heme binding pocket to complete unfolding to give a random coil. In contrast, proteins that bind heme transiently are fully folded in their apo and holo states, presumably allowing both apo and holo forms to remain biologically active resisting aggregation or proteolysis. The principles identified here provide a framework for the design of *de novo* proteins that will exhibit tight heme ligand binding and for the identification of the function of structural genomic target proteins with heme ligands.

Proteins 2010; 78:2349–2368.

Key words: heme; cytochrome; binding pocket; ligand; apo state.

INTRODUCTION

Proteins containing heme groups have been studied intensely for many years. As early as 1850 hemoglobins from several different species of animal had been recognised and crystallised, with Teichmann in 1853 identifying that hemoglobin was associated with a substance, known as hemin, that contained iron.[1] The complete structure of the hemin (the chloride of heme) was determined by Fischer who reported the synthesis of this molecule in 1929.[2] Atomic resolution structural information about the characteristics of the heme binding site in myoglobin and hemoglobin came in the late 1950s, these proteins being among the first to have their structures determined by X-ray diffraction.[3,4] Since then extensive structural studies of proteins containing heme groups have continued and today the protein data bank contains over 2300 structures of proteins containing one or more heme groups.

Much of the current research involving heme proteins concentrates on understanding the details of their biological function and how this relates to the protein structure.[5–13] However, another area of active research is the design of new proteins binding heme groups and related ligands. Much exciting progress is being made in this *de novo* design of heme proteins at present.[14–20] Heme proteins are being developed as novel biomolecular with non-linear optical properties, as molecules that self-assemble at an interface providing an ordered material with electron transfer properties, and as "smart" materials that respond to stimuli such as changes in pH, ionic strength or redox environment.[21–29] The design and development of such heme proteins with specific structural and functional properties requires a firm understanding of the heme group and its binding sites in proteins. One strategy for obtaining such an understanding is through the systematic study of heme protein structures currently available.[9,30–34] This is the approach we have used here. We have concentrated particularly on identifying the characteristics of the binding pockets that recognize and bind heme molecules, as well as the features that enable heme groups to perform different functions in biology.

METHODS

Data set of heme proteins

To construct the data set of heme proteins, all the proteins with coordinated heme ligands in the Protein Data Bank (version at April 2009)[35] were identified.

Additional Supporting Information may be found in the online version of this article.

Abdullah Kahraman's current address is Institute of Molecular Systems Biology, ETH Zürich, Wolfgang-Pauli Str. 16, Zürich 8093, Switzerland.
*Correspondence to: Lorna J. Smith, Inorganic Chemistry Laboratory, University of Oxford, South Parks Road, Oxford OX1 3QR, United Kingdom. E-mail: lorna.smith@chem.ox.ac.uk.
Received 4 November 2009; Revised 31 March 2010; Accepted 2 April 2010
Published online 9 April 2010 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.22747

Short synthetic peptides associated with heme groups were excluded. The structure type of each protein was assigned using the CATH Protein Structure Classification database (version 3.2.0),[36,37] concentrating on the domains in a protein where a heme group is bound. One example of a protein from each CATH homologous superfamily that was identified was then taken to form the data set of heme proteins (Table I). Where possible we chose proteins whose structures had been determined at high resolution and proteins for which there is additional information (e.g. about the apo state) available in the literature. Some recently determined protein structures have not yet been classified in the CATH database. In these cases the structural classification from the SCOP database (version 1.73) was used, or if not available from the pre-SCOP database.[38,39] These proteins were only included in the heme data set if they belonged to a different SCOP family from all the other proteins in the data set. The proteins in the heme data set were then divided into six subsets on the basis of the role of the heme group in the protein identified from the literature (see Supporting Information Table I).

Analysis of the heme binding sites

The heme group solvent accessible surface area was calculated in each heme protein using the program NACCESS.[40] This uses the Lee and Richards method[41] in which a probe of radius 1.4 Å is rolled around the surface of the molecule. The program HBPLUS was used to determine hydrogen atom positions, when they were not included in the protein coordinate set, and to identify hydrogen bonds between the heme group and the protein.[42] The positions of hydrogen atoms were taken from references.[43,44] Hydrogen bonds were identified between hydrogen atoms (H) attached to donor atoms (D) and acceptor atoms (A) if the H-A distance is <2.7 Å, the D-A distance is <3.3 Å, the D-H-A angle is $>90^\circ$, and the H-AAA angle is $>90^\circ$, where the AA atom is the one attached to the acceptor, usually preceding it along the amino acid chain. The program LIGPLOT[45] was used to identify nonbonded contacts between the protein and heme group; a nonbonded contact is identified if the atoms concerned are less than 3.9 Å apart. We describe the residues that make nonbonded contacts with the heme as "heme binding site residues."

The characteristics of the heme binding sites in homologous proteins were analyzed using data from the ConSurf-HSSP database.[46] For this section of the analysis only proteins in the data set where 50 or more aligned sequences are available have been considered (10 proteins in total). This database gives a conservation score for each amino acid in a protein based on multiple sequence alignments taken from the homology-derived secondary structure of proteins (HSSP) database.[47] The conservation scores are calculated using the Rate4Site algorithm[48] which takes into account the phylogenetic relationships between the homologs and the stochastic nature of the evolutionary process. The scores are normalized, so that the average score for all residues is zero, and the standard deviation is one. The lowest score represents the most conserved position in a protein. Considering the residues in the heme binding site as identified by LIGPLOT,[45] the mean normalized conservation scores for the heme binding site residues in the proteins analyzed range from -0.32 to -0.99 making it clear that residues within the heme binding site are conserved more than average.

The shape and volume of the heme binding clefts were defined and analyzed using the approach and methods described in Kahraman *et al.*[49] In this approach, the binding site cleft was initially identified using SURFNET,[50] which detects protein cavities by inserting spheres between protein atoms. The binding cleft was then reduced, so it more accurately reflected the shape of the cleft where the heme binds. To do this only the SURFNET spheres which were within 0.3 Å of the protein atoms that interact with the heme group were retained. Next, the cleft was translated to the coordinate origin and rotated according to its principle components. After the transformation, the surface of the cleft was approximated by a discrete radial function. The radial function was described by a spherical harmonics expansion of the order $l_{max} = 14$. The expansion returned a set of 225 coefficients that uniquely described the shape of a cleft. To compare two cleft shapes with each other their expansion coefficients were compared using a Euclidean metric in a 225 dimensional space producing smaller coefficient distances for more similar shapes. For more detailed information about the methodology see.[49,51] All pairwise coefficient distances between all clefts in the data set were collected in an all against all coefficient distance matrix. The matrix was used to cluster the heme binding sites based on their shape similarity a hierarchical agglomerative clustering method. This method started off with all the clefts in their own cluster and then iteratively merged the two clusters closest to each other following the Ward method.[52] Ward's clustering minimises the increase of the sum of squared errors in every iteration. We used the R-implementation of the agnes package.[53] To determine the optimal number of clusters, we used the Kelley plot, whose algorithm is described in Ref. 54 in detail. In this representation, the number of clusters on the x-axis is plotted against a penalty function on the y-axis, whose value depends on the average spread of the values in the cluster and the total number of clusters at iteration step i of the clustering. The optimal number of clusters is the one that results in the minimal penalty value.

Table I
Data Set of Heme Binding Proteins

Protein, PDB code, and literature citation	CATH class, architecture, and homologous superfamily code[a]	Number and type of heme	Axial ligands	Heme binding mode[b]	Heme solvent accessible surface area (Å^2)	Heme E° (mV)[c]
Function—electron transfer						
Mitochondrial cytochrome *c* from yeast, 1ycc[76]	Mainly alpha, orthogonal bundle, 1.10.760.10	1 *c*-type	His, Met	2	30	290.0
Cytochrome b_{562} from *E. coli*, 256b[132]	Mainly alpha, up-down bundle, 1.20.120.10	1 *b*-type	His, Met	1	120	160.0
Bacterioferritin cytochrome b_1 from *E. coli*, 2vxi[133]	Mainly alpha, up-down bundle, 1.20.1260.10	1 *b*-type	Met, Met	1	81	
Cytochrome *f* from turnip, 1hcz[134]	Mainly beta, sandwich, 2.60.40.830	1 *c*-type	His, amino Tyr 1	1	38	
Cytochrome domain of cellobiose dehydrogenase from *Phanerochaete chryosporium*, 1d7b[73]	Mainly beta, sandwich, 2.60.40.1210	1 *b*-type	His, Met	1+ ring B	144	164.0
Bovine cytochrome b_5, 1cyo[89]	Alpha beta, roll, 3.10.120.10	1 b-type	His, His	1	213	5.1
Function—binds small molecules						
Sperm whale myoglobin, 1a6n[135]	Mainly alpha, orthogonal bundle, 1.10.490.10	1 *b*-type	His, vacant	1	149	120.0
Nitrophorin 4 from *Rhodnius prolixus*, 1d2u[57]	Mainly beta, barrel, 2.40.128.20	1 *b*-type	His, vacant	1	86	
CO-sensing protein, N-terminal domain from *Rhodospirillum rubrum*, 1ft9[70]	Mainly beta, sandwich, 2.60.120.10	1 *b*-type	His, Pro reduced; Cys, Pro oxidized	1	183	−260.0
Histidine kinase FixL heme domain from *Bradyrhizobium japonicum*, 2vv6[136]	Alpha beta, 2-layer sandwich, 3.30.450.20	1 *b*-type	His, vacant	1	49	68.0
H-NOX domain from *Thermoanaerobacter tengcongensis*, 1u56[137]	Alpha beta, complex, 3.90.1520.10	1 *b*-type	His, ligand	Buried	5	
Nitrophorin from *Cimex lectularius*, 1ntf[138]	Alpha beta, 4-layer sandwich 3.60.10.10	1 *b*-type	Cys, vacant	2	104	
Function—enzyme						
Cytochrome *c* peroxidase from yeast, 2cyp[77]	Mainly alpha, orthogonal bundle, 1.10.420.10	1 *b*-type	His, water	1	25	−194.0
Cytochrome P-450$_{cam}$ from *Pseudomonas putida*, 1phc[139]	Mainly alpha, orthogonal bundle, 1.10.630.10	1 *b*-type	Cys, water or vacant	Buried	21	−303.0
Chloroperoxidase from Fungus (*Caldariomyces fumago*), 1cpo[75]	Mainly alpha, orthogonal bundle, 1.10.489.10	1 *b*-type	Cys, water or vacant	Buried	3	−140.0

(Continued)

Table I
(Continued)

Protein, PDB code, and literature citation	CATH class, architecture, and homologous superfamily code[a]	Number and type of heme	Axial ligands	Heme binding mode[b]	Heme solvent accessible surface area (Å^2)	Heme E° (mV)[c]
Human myeloperoxidase, 1cxp[66]	Mainly alpha, orthogonal bundle, 1.10.640.10	1 Unusual ester and sulphonium ion linkages	His, His	1	62	
C-terminal heme d_1 domain of cytochrome cd_1 nitrite reductase from *T. pantotropha*, 1qks[58]	Mainly beta, 8 propeller, 2.140.10.20	1 d_1 heme	Tyr, His oxidized, His, vacant reduced	Buried	17	
Catalase I from human, 1dgf[83]	*Multidomain proteins alpha and beta; heme dependent catalase-like*	1 *b*-type	Tyr, vacant	Buried	2	
Nitric oxide synthase chain A domain 1 from cow, 1nse[90]	Alpha beta, complex, 3.90.340.10	1 *b*-type	Cys, vacant	1	53	
Human indoleamine 2,3-dioxygenase, 2d0t[140]	*All alpha proteins, indoleamine 2,3-dioxygenase-like*	1 *b*-type	His, ligand	1	83	
Sulfite reductase from *E. coil*, 1aop[65]	Alpha beta, 2-layer sandwich, 3.30.413.10	1 siro heme	Cys, vacant/phosphate	1	107	−340.0
Function—transient heme binding						
Heme complex of heme oxygenase from *Corynebacterium diphtheriae*, 1iw0[56]	Mainly alpha, Up-down bundle, 1.20.910.10	1 *b*-type	His, water	Mode 2	158	
Heme binding protein A from *Serratia marcescens*, 1b2v[141]	Alpha beta, 2-layer sandwich, 3.30.1500.10	1 *b*-type	His, Tyr	1+ring C	169	
Hemopexin from rabbit in complex with heme, 1qjs[55]	Mainly beta, 4 propeller, 2.110.10.10	1 *b*-type	His, His	2+ring B	176	
Heme transport protein HemS from *Yersinia enterocolitica*, 2j0p[117]	*Multi-domain proteins (alpha and beta), heme iron utilization protein-like*	1 *b*-type	His, water	3+ring C	167	
Iron-regulated surface determinant protein C IsdC from *Staphylococcus aureus*, 2o6p[59]	*All beta proteins immunoglobulin-like β-sandwich*	1 *b*-type	Tyr, vacant	1	258	
Function—multiheme						
Cytochrome subunit of the photosynthetic reaction centre from *Rhodopseudomonas viridis*, 2i5n[109]	Mainly alpha, orthogonal bundle, 1.10.468.10	4 *c*-type	Heme 1 His, Met	3	59	−60.0
			Heme 2 His, Met	Buried	0.4	300.0
			Heme 3 His, Met	Buried	2	370.0
			Heme 4 His, His	1	42	10.0

(Continued)

Table I
(Continued)

Protein, PDB code, and literature citation	CATH class, architecture, and homologous superfamily code[a]	Number and type of heme	Axial ligands	Heme binding mode[b]	Heme solvent accessible surface area (Å^2)	Heme E° (mV)[c]
Cytochrome *b* subunit of the mitochondrial cytochrome bc_1 complex from cow, 1pp9 (chain c)[142]	Mainly alpha, up-down bundle, 1.20.810.10	2 *b*-type	Heme 1 His, His	Buried	7	
			Heme 2 His, His	Buried	10	
Cytochrome *b* subunit of fumerate reductase from *Wolinella succinogenes*, 1qla (chain C)[108]	Mainly alpha, up-down bundle, 1.20.950.10	2 *b*-type	Heme 1 His, His	Buried	3	
			Heme 2 His, His	Buried	13	
Cytochrome c oxidase (subunit I) from cow, 1v54 (chain A)[64]	Mainly alpha, Up-down bundle, 1.20.210.10	2 heme *a*	Heme 1 His, His	Buried	4	276.0
			Heme 2 His, vacant	Buried	12	290.0
Hydroxylamine oxidoreductase from *Nitrosomonas europaea*, 1fgj[63]	Mainly alpha, domain 1: orthogonal bundle, 1.10.780.10; domain 2: up-down bundle, 1.20.850.10	7 *c*-type and one heme P460. Heme 4 (P460) has covalent link from Heme CHC to Tyr Cε1 in adjacent subunit in trimer	Heme 1 His, His	1+ring C	157	
			Heme 2 His, His	1	50	
			Heme 3 His, His	Buried	13	
			Heme 4 His, vacant	3+ring C	108	−260.0
			Heme 5 His, His	1	36	
			Heme 6 His, His	Buried	6	
			Heme 7 His, His	Buried	9	
			Heme 8 His, His	1	217	
Cytochrome c_3 from *Desulfovibrio desulfuricans*, 2cy3[61]	Alpha beta, complex, 3.90.10.10	4 *c*-type	Heme 1 His, His	3	130	
			Heme 2 His, His	1+ring C	320	
			Heme 3 His, His	1+ring C	202	
			Heme 4 His, His	3	126	
Function—Other						
Cystathionine beta-synthase from human, 1jbq[127]	Alpha beta, 3-layer (aba) sandwich, 3.40.50.1100	1 *b*-type	Cys, His	1	147	
Putative melanin biosynthesis protein Tyr A, 2iiz[128]	*Alpha and beta proteins, ferrodoxin-like*	1 *b*-type	His, vacant	Buried	24	

[a]For proteins that are not yet included in CATH, the SCOP class and fold are given in italics (see Methods).
[b]For the most exposed heme groups the overall binding mode is listed followed by any other additional parts of the heme group that have a significant exposure.
[c]Heme reduction potential values were taken from the Heme Protein Database[33]. Only values which can be assigned to an individual heme group in multi-heme proteins are listed. For 2i5n the assignments to individual heme groups were taken from Ref. 143.

Twenty-seven of the structures in the heme protein data set have a resolution of 2 Å or better. For the remaining seven proteins (CO-sensing protein from *Rhodospirillum rubrum*, rabbit hemopexin, human cystathionine β-synthase, the putative melanin biosynthesis protein Tyr A, human indoleamine 2,3-dioxygenase, hydroxylamine oxidoreductase from *Nitrosomonas europaea*, and the cytochrome *b* subunit of the mitochondrial cytochrome bc_1 complex) only lower resolution structures are available. To assess the effects that using lower resolution structures might have on the analysis reported in this article, we have compared the heme binding sites in the structures we have used, with those in other structures that are available for the same proteins. Some small variations are seen in the heme group solvent accessibility, the residues involved in making hydrogen bonds or nonbonded contacts with the heme group and the number of these contacts in different structures. However, the differences are small compared to the differences seen between proteins and are averaged out when comparisons are made across the data set. We have though chosen to take a conservative approach and have omitted the seven lower resolution structures from the detailed shape analysis and clustering reported in Figure 6.

Table II
Summary of the Characteristics of Proteins Within the Data Set of Heme Binding Proteins

Number of proteins in data set	
Total	34
Electron transfer subset	6
Binds small molecules subset	6
Enzymes subset	9
Transient heme binding subset	5
Multi-heme protein subset	6
Other protein subset	2
Protein structure adopted	
CATH Class Mainly α	15
CATH Class Mainly β	7
CATH Class αβ	12
Number of heme groups	
Total	50
Heme *a*	2
Heme *b*	27
Heme *c*	17
Heme d_1	1
Heme P460	1
Siro heme	1
Heme from myeloperoxidase	1
Axial ligands for heme group	
His	61
Met	8
Cys	6
Tyr	3
Pro	1
N-terminal amino	1
Vacant/water	18
Changes on oxidation/reduction	2

RESULTS

Data set of proteins with heme groups

Although the protein data bank contains over 2300 entries that are heme proteins, many of these are multiple structures of the same protein or structures of closely related proteins. Our approach therefore has been to identify all the different types of protein structure in the protein data base[35] with one or more heme group bound and to select one example of each of these for detailed analysis. Using this method we have assembled a data set of 34 heme proteins, each of which belongs to a different CATH homologous super family[36,37] and a different SCOP family[38,39] (Tables I and II and Supplementary material). These proteins contain a total of 50 heme groups. The data set has been divided into six subsets on the basis of the function of the heme group in the protein (Table I). We have identified three main roles for heme (subsets 1–3), namely heme groups that are involved in electron transfer, heme groups that bind small molecules (e.g. oxygen or nitric oxide) and heme groups that are part of an enzyme active site usually catalyzing a reaction involving oxidation or reduction. The fourth subset of proteins in the data set consists of heme complexes to proteins where the heme group is bound as a ligand and is not permanently attached to the protein in its functional form. These are proteins which act as heme transporters such as hemopexin (1qjs)[55] or are enzymes that catalyze the degradation of heme groups (e.g. heme oxygenase 1iw0).[56] We refer to these proteins in this article as those that show transient heme binding. In addition, we have in the data set six proteins that bind more than one heme group (Table I). All the heme groups within these proteins do not necessarily play the same role so we have grouped these proteins in a fifth subset of multiheme proteins. For two of the proteins in the data set the function of the heme group is not clear. These proteins are classed together in a final sixth subset (other proteins).

Most heme binding proteins are rich in α-helical secondary structure. The predominate CATH Classes and Architectures in the data set of heme binding proteins are the mainly α orthogonal bundle and the mainly α up-down bundle (11 and 7 examples of each respectively). However, heme groups are found in proteins with a wide range of structure. These include the β-barrel of nitrophorin 4 (1d2u),[57] the β-propeller of cytochrome cd_1 nitrite reductase (1qks),[58] and the immunoglobulin-like β-sandwich of iron regulated surface determinant protein C (2o6p)[59] (Table I).

Variants of the heme group and axial ligands

In addition to the variety of protein structures, there are also differences in the chemical structure of the heme group and its mode of attachment to the protein (Tables I and II). Within the data set heme *b* and heme *c* groups predominate with 27 and 17 examples, respectively (Fig. 1).

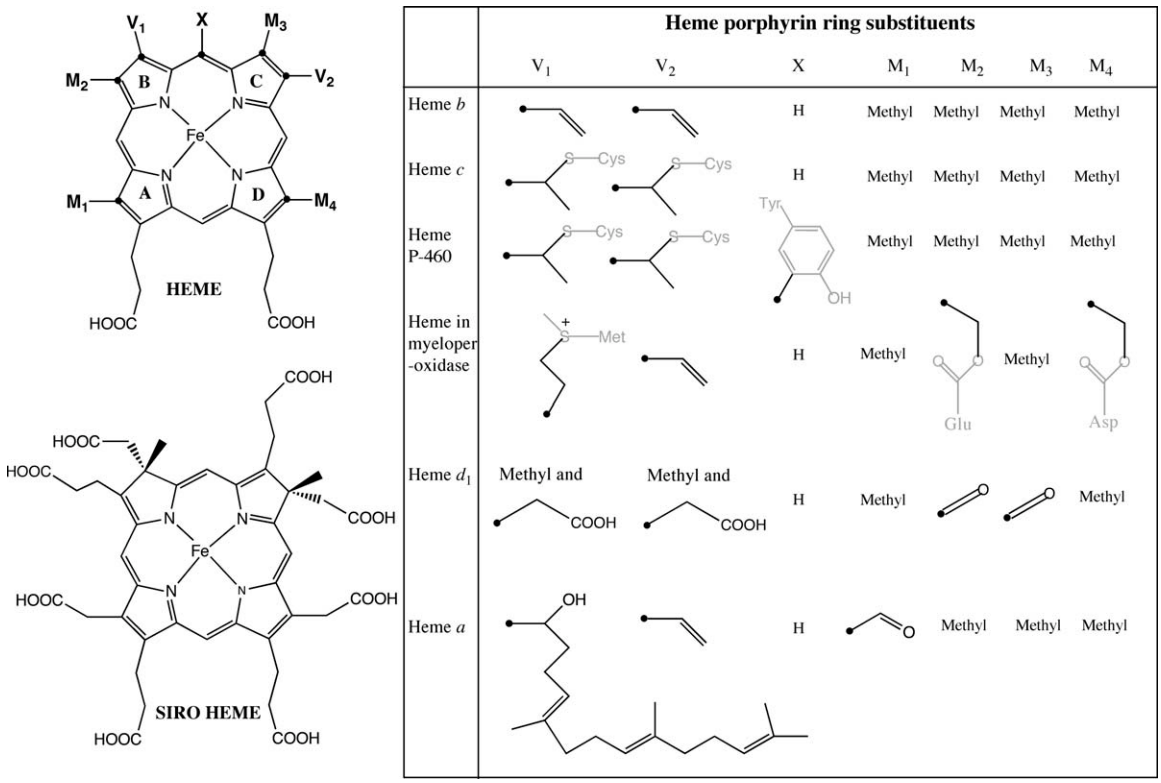

Figure 1

Structures of the different heme group variants found in the data set of heme proteins. The heme group porphyrin ring skeleton is shown at the top left and the various porphyrin ring substituents found in the different heme variants are given in the table on the right. Black spheres indicate which atoms in the substituents bond to the porphyrin ring. Any atoms which are part of an amino acid side chain within the attached protein are shown in gray.

The most common form of attachment for heme *c* is via two thioether bonds from the V_1 and V_2 groups to a CXXCH amino acid motif in the protein sequence (Fig. 1).[60] This arrangement is seen in 16 of the *c*-type hemes in the data set. However, in heme 2 in cytochrome c_3 from *Desulfovibrio desulfuricans* (2cy3) the two thioether linkages from the heme group are instead to a CXXXXCH motif.[61] Other forms of attachment that are known include linkages via two thioether bonds to a CXXCK or a CXXXCH sequence or via just one thioether bond.[5,62]

The data set also contains examples of proteins that contain the heme P460,[63] heme d_1,[58] heme *a*,[64] and siro heme groups,[65] and an example of the heme group from myeloperoxidases[66] (Fig. 1 and Tables I and II). A number of other heme variants are known including heme *o* that is seen in ubiquinol oxidase,[67] heme *d* that is seen in *E. coli* catalase HPII,[68] and chlorocruorohaem that is seen in chlorocruorins (giant oxygen-binding proteins in polychaeta).[69] However, these are not included in the data set and analyzed here because of a close homology of the protein with which the heme is associated to another member of the data set, or because no structures of proteins with these heme groups are available.

In general, two protein side chain groups coordinate to the iron of the heme group when it is bound to the protein as axial ligands (Tables I and II). Histidine predominates as an axial ligand (61 occurrences in the data set) especially in multiheme proteins. A previous detailed analysis of histidine axial ligands that are part of the CXXCH amino acid motif in *c*-type heme groups shows that bonding constraints mean they are orientated so they point either between the heme propionates or between the NA and NB heme nitrogens.[34] Other side chains that act as axial ligands in the data set are methionine, cysteine, and tyrosine (8, 6, and 3 occurrences in data set, respectively). In two cases a change in axial ligand is observed on reduction. In CO-sensing protein (1ft9) in the Fe^{3+} oxidized state Cys 75 is an axial ligand but this is replaced by His 77 on reduction to Fe^{2+},[70] while in cytochrome cd_1 nitrite reductase (1qks) one of the axial ligands, Tyr 25, moves out of the coordination sphere of the heme d_1 iron on reduction leaving a vacant site.[58]

Heme solvent accessibility, heme binding modes, and heme group distortions

One of the initial characteristics of a heme group bound to a protein is its level of exposure to solvent and, related to this, the mode by which it binds in the heme binding pocket.[32,71] The solvent accessibility of a free

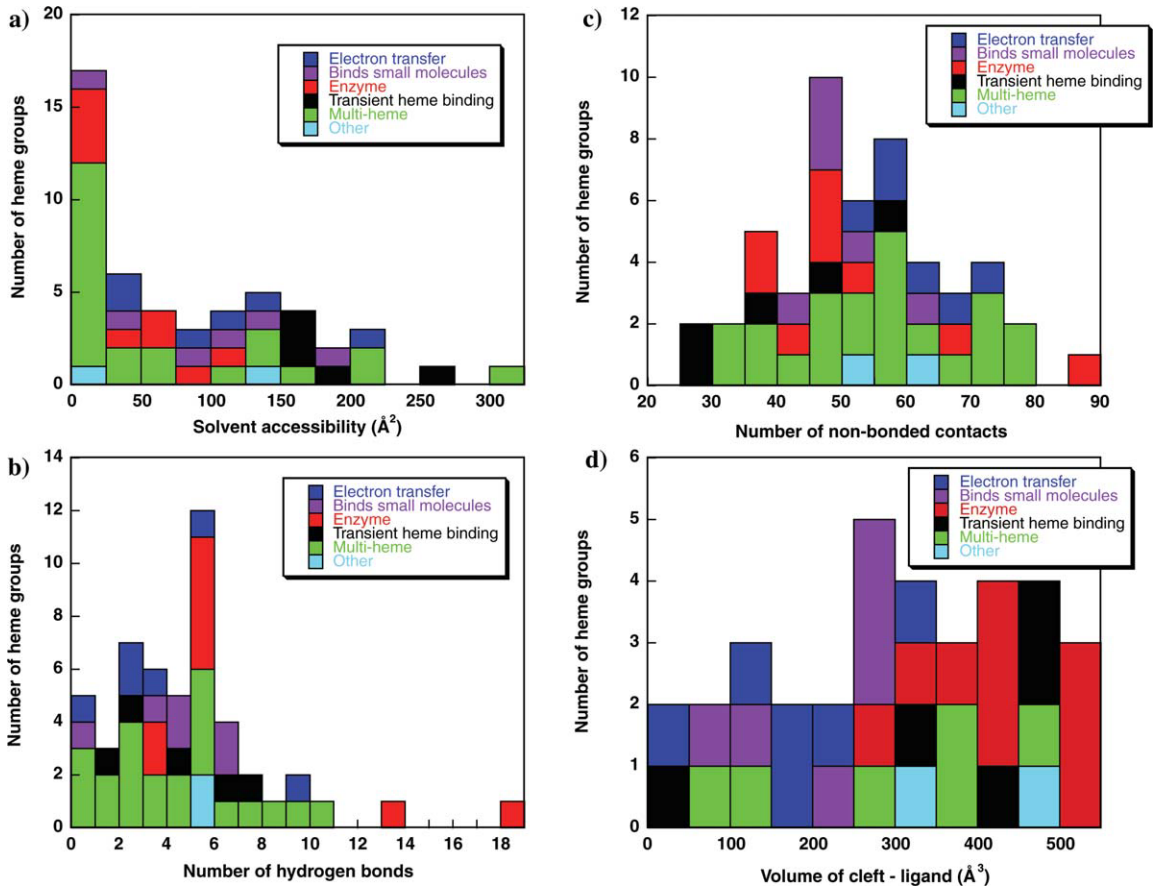

Figure 2

The distribution in the heme protein data set of (**a**) heme group solvent accessibility, (**b**) number of hydrogen bonds between the heme group and the protein, (**c**) number of nonbonded contacts between the heme group and the protein, (**d**) additional volume in the heme binding cleft. The data points for each heme group are colored according to the functional subset to which the protein concerned belongs. The additional volume plotted in panel (d) is the volume of the heme binding cleft minus the volume of the heme ligand. For multiheme proteins there is overlap in the heme binding cleft of the individual heme groups in some cases. Here the volume per heme group has therefore been determined by calculating the cleft volume of all heme groups together and dividing by the total number of heme groups present. The mean volume of the heme ligands present has then been subtracted to give the additional volume value.

heme group is 830 Å^2 while, with two exceptions, all the heme groups in the proteins analyzed are significantly buried with an accessibility of 225 Å^2 or less [Fig. 2(a)]. The two exceptions with higher heme solvent accessibility are in iron-regulated surface determinant protein C (2o6p; 258 Å^2) and cytochrome c_3 (2cy3 heme 2; 320 Å^2). The high accessibility of heme 2 in cytochrome c_3 may result from the high heme-to-polypeptide chain ratio in this protein. Cytochrome c_3 is a small protein (118 residues) but contains four heme groups[61] and there may be insufficient polypeptide chain to provide a buried heme environment.[72] Indeed, all four heme groups in the protein are significantly exposed (solvent accessibility 126–320 Å^2). Iron-regulated surface determinant protein C binds heme as a ligand.[59] All the five proteins in the data set that show transient heme binding have a relatively high heme solvent accessibility (158–258 Å^2). This may reflect the different requirements for a binding site where heme is readily bound and released rather than permanently associated with the protein. These different requirements are considered in more detail in later sections.

Seventeen of the 50 heme groups, particularly those in multiheme proteins and some enzymes, are almost completely buried (solvent accessibility $<$25 Å^2). Of the more exposed heme groups, the majority of them (24) are bound to the protein so that both their propionate side chains, attached to heme rings A and D, are solvent exposed.[71] We call this orientation heme binding mode 1 (Fig. 3 and Table I). For five of these heme groups with binding mode 1 the heme group is particularly solvent exposed and additional parts of the heme molecule also have a significant exposure. For example, in the cytochrome domain of cellobiose dehydrogenase (1d7b),[73] the substituents to heme rings A, B, and D are all solvent accessible, while for heme 2 in cytochrome c_3

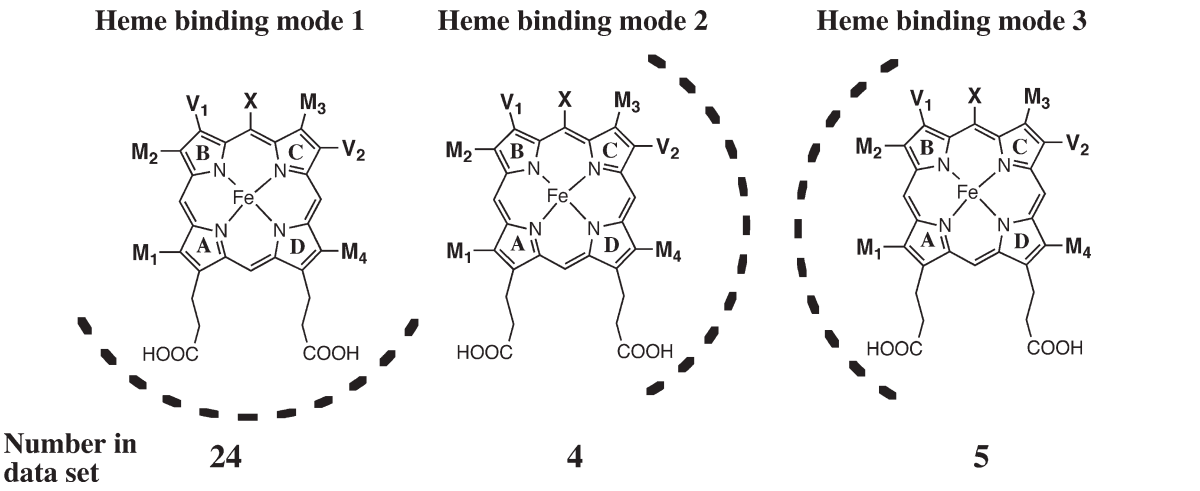

Figure 3

Heme binding modes observed in the data set of heme binding proteins. The parts of the heme group that are solvent exposed are shown by the dashed line. The substituents V_1, V_2, X, M_1, M_2, M_3, and M_4 are defined for the different heme variants in Figure 1.

(2cy3; heme with the maximum accessibility) rings A, C, and D are solvent accessible.

Although heme binding mode 1 predominates, for eight heme groups in the data set different heme binding modes are observed (Table I). In particular we define heme binding mode 2, in which the substituents of heme rings C and D have a significant exposure, and heme binding mode 3, in which the substituents of heme rings A and B have a significant exposure (Fig. 3). Binding mode 2 and 3 are seen for four and five heme groups respectively in the data set. Interestingly, in the heme complex of heme oxygenase binding mode 2 is observed in the crystal structure[56] but in solution, as probed by NMR,[74] two isomers are observed where the heme groups is rotated about the α-δ meso axis. This means that the heme binds in either mode 2 or 3. Heme binding modes 2 or 3 are particularly seen in multiheme proteins and in proteins that show transient heme binding.

Although an isolated heme group is essentially planar most heme groups in proteins are distorted. In the data set these distortions include bowl shaped heme groups, seen for example in chloroperoxidase (1cpo),[75] saddle-shaped heme groups, observed in mitochondrial cytochrome c (1ycc)[76] and cytochome c peroxidase (2cyp),[77] and heme groups with extensive ruffling such as that observed in nitrophorin 4 (1d2u).[57,78] Distortions to a heme group are recognized to affect its reduction potential and ligand binding affinities, as well as the spectroscopic properties of the heme group.[78–82] For example, for the H-NOX domain (1u56), recent studies have shown that a Pro115Ala mutation to the protein results in a significant reduction in the distortion in the heme group of the protein. An increased oxygen binding affinity and a decreased reduction potential is observed for the flattened heme group in this mutant protein.[82]

Heme-protein contacts

Figure 2(b,c) show the number of hydrogen bonds and nonbonded contacts between the heme group and the protein respectively. The largest numbers of protein-heme hydrogen bonds are seen with siro heme (1aop; 19 hydrogen bonds) and heme d_1 (1qks; 13 hydrogen bonds). Both of these heme groups have extra carboxylate side chain substituents compared to hemes *a*, *b*, or *c*. All the other heme groups with six or more hydrogen bonds are completely buried or bind in modes 2 or 3. In these cases, hydrogen bonds are seen from the buried propionate side chains to the protein, particularly to lysine and arginine side chains and to backbone amide groups. For example, in human catalase I (1dgf; heme buried with solvent accessibility 2 $Å^2$)[83] there are hydrogen bonds between the propionate groups and the side chains of Arg 72, Arg 112, His 362, and Arg 365.

The maximum number of nonbonded contacts is seen for the heme group in catalase I (1dgf) that has 86 nonbonded contacts and very low heme solvent accessibility of 2 $Å^2$. The lowest number of nonbonded contacts are seen for two proteins that show transient heme binding, hemopexin (1qjs; 25 contacts) and heme transport protein HemS (2j0p; 27 contacts). The heme groups in both these proteins are very solvent exposed (176 $Å^2$ and 167 $Å^2$ for hemopexin and HemS, respectively) but bind in a mode that buries one of the propionate groups (binding modes 2 and 3 for hemopexin and HemS, respectively) with a significant number of protein-heme hydrogen bonds (7 and 6 for hemopexin and HemS, respectively). The different characteristics of the heme-protein contacts in these cases may reflect the specific requirements when heme is binding transiently as a ligand rather than becoming permanently associated with the protein. Hydrogen bonds, rather than nonbonded

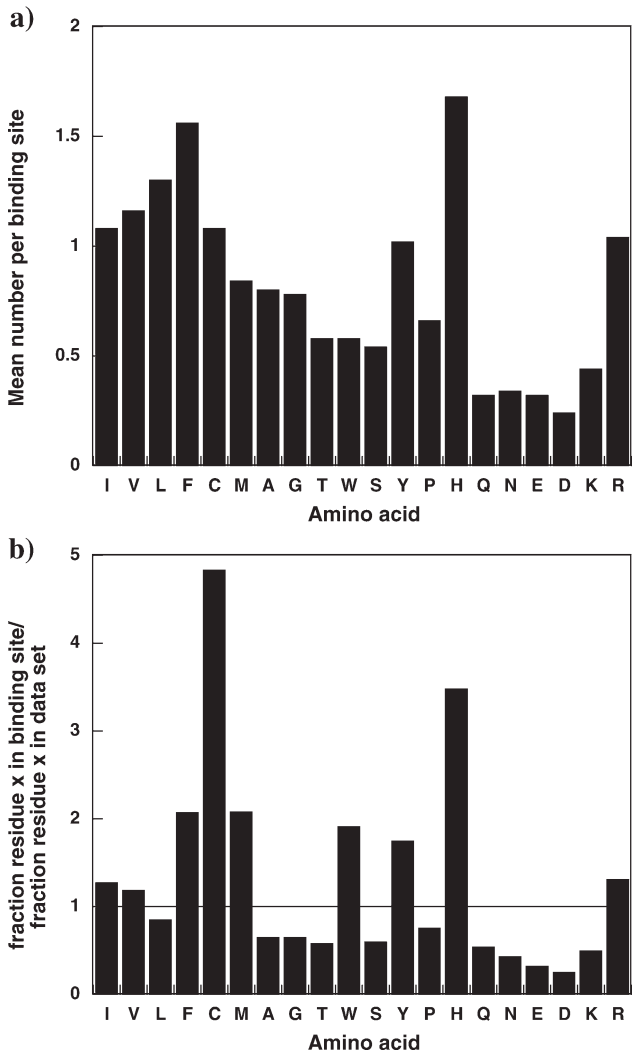

Figure 4

Amino acid profiles of the binding sites of the heme groups in the data set. (**a**) The mean number of each amino acid present per binding site. (**b**) The normalized amino acid profiles calculated as

$$\text{Normalized population for amino acid x} = \frac{\%\text{ of residues of type x in heme site}}{\%\text{ of residues of type x in the data set}}.$$

contacts, can be more easily modulated by changes such as pH. Indeed hemopexin is reported to bind heme with a very high affinity ($K_d < 1$ pM) but *in vitro* heme release occurs below pH 5.[55,84]

Composition and conservation of heme binding sites

To assess the chemical composition of the heme binding sites and their level of hydrophobicity, the amino acid profile of the binding sites have been determined using the residues making contacts with the heme group identified by LIGPLOT.[45] Figure 4(a) shows the mean number of each type of amino acid found in the binding sites of the 50 heme groups in the data set. In Figure 4(b), the amino acid profiles are normalized to take into account the relative abundance of the different amino acid types. The normalized profile shows clear maxima for cysteine (seen in the CXXCH motif with *c*-type heme) and histidine (the most common axial ligand in the data set). These residues are particularly seen in the heme binding sites of multiheme proteins which mostly contain *c*-type heme groups. Methionine, the second most frequent axial ligand, also shows an enhanced normalized frequency. The normalized profiles also show a preference in the heme binding sites for the aromatic residues phenylalanine, tryptophan, and tyrosine and for positively charged arginine side chains.[85] The aromatic side chains tend to adopt either an off-set parallel staggered or an edge-to-face orientation relative to the heme group.[21] These orientations have been recognized to give rise to favorable electrostatic interactions between aromatic π systems.[86–88] In chloroperoxidase (1cpo),[75] for example, five aromatic residues make contacts to the heme group. Three of these, Phe 57, Phe 103, and Phe 186, are arranged in a parallel staggered arrangement relative to the heme while the other two, Trp 213 and Phe 214, are positioned perpendicular to the heme with their faces orientated relative to the edge of the heme group. Arginine side chains bind to the carboxylic side chains on the heme groups.[32] and these are especially seen in the heme binding sites of the enzymes sulfite reductase (1aop) and cytochrome cd_1 nitrite reductase (1qks) (five arginine side chains in each binding site) that contain the siro heme[65] and heme d_1 groups[58] respectively with additional carboxylic acid groups.

Although the focus of the analysis reported in this article is on a data set of nonhomologous proteins, we have also been interested in assessing the similarity of heme binding sites within a given protein family. This analysis has used the ConSurf-HSSP database[46] with multiple sequence alignments taken from the homology-derived secondary structure of proteins (HSSP) database[47] and has considered the residues that make nonbonded contact with the heme group in the binding site as identified by LIGPLOT (see Methods section).[45] In general, for proteins where the heme binding site is very nonpolar, the character of the amino acids within the binding site is conserved, rather than their actual identity. For example, Figure 5(a) shows the conservation of aromatic and hydrophobic residues within the heme binding site of the electron transfer protein cytochrome b_5 (1cyo).[89] Phe 35 is found as a phenylalanine, tyrosine or tryptophan in the multiple sequence alignment while Leu 25, Leu 32, and Leu 45 are found predominantly as leucine, isoleucine, or valine (also some Cys, Ala, and Gly for Leu 32). In the heme binding sites of enzymes there is also a significant level of conservation of hydrophobic groups. However, here the conservation of charged or polar residues in the heme binding site that

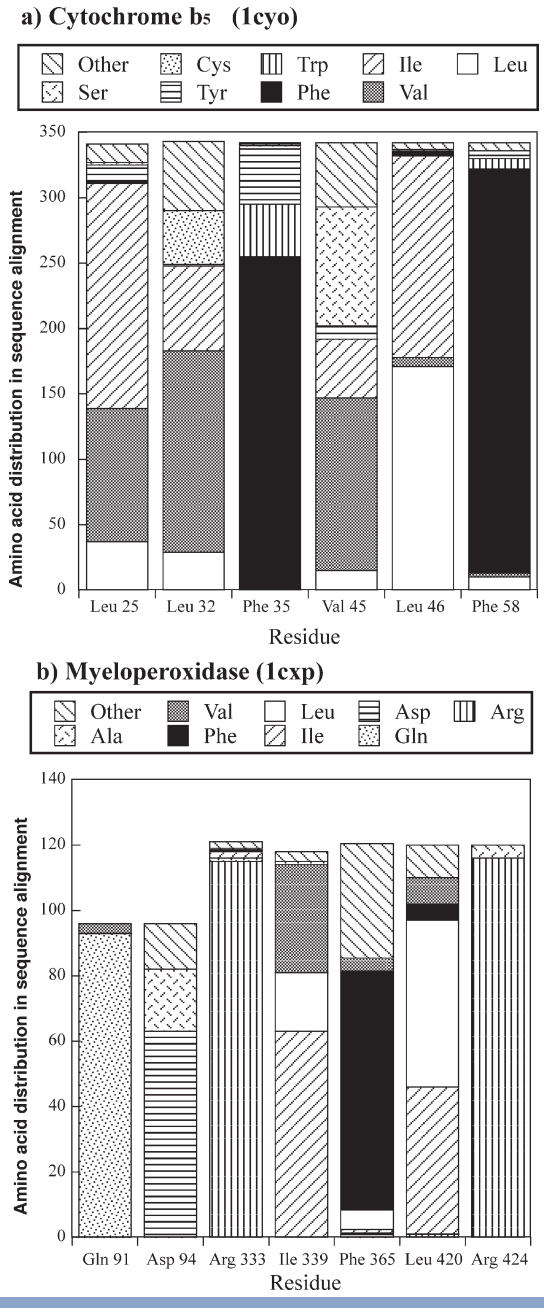

Figure 5

Conservation of residue type in the heme binding sites of sets of proteins that are homologous to cytochrome b_5 (1cyo; panel **a**) and myeloperoxidase (1cxp; panel **b**). For cytochrome b_5, the distribution of amino acid type across the alignment of 342 sequences at positions corresponding to Leu 25, Leu 32, Phe 35, Val 45, Leu 46, and Phe 58 is shown. For myeloperoxidase the distribution of amino acid type across the alignment at positions corresponding to Gln 91, Asp 94 in chain A (98 sequences) and Arg 333, Ile 339, Phe 365, Leu 420, and Arg 424 in chain C (139 sequences) is shown.

are involved in substrate binding, in the enzyme mechanism or in coordination to heme is striking. This is seen, for example, for myeloperoxidase (1cxp),[66] where Gln 91, Arg 333, Arg 424, and to some extent Asp 94, are highly conserved across the multiple sequence alignment [Fig. 5(b)]. In myeloperoxidase Gln 91 hydrogen bonds to water molecules in the distal side of the heme cavity and is thought to be involved in hydrogen bonding to the halide ion substrate. Arg 33 and Arg 424 both hydrogen bond to the heme propionate groups while Asp 94 forms an ester linkage to the heme.[66]

Shape of heme binding sites

Another important feature of the binding site, for both heme recognition and function, is its shape. A global shape descriptor was used to analyze the shape of the heme binding pockets in each of the protein in the data set as described in the Methods section.[49,51] The volumes of the heme binding pockets varies from 735 Å^3 for cytochrome b_5 (1cyo) through to 1642 Å^3 in cytochrome *c* oxidase (1v54). Some of this variation reflects the different volumes of the various heme ligands, the greatest binding site volumes being seen for cytochrome *c* oxidase and sulfite reductase (1aop; 1552Å^3) with the larger siro heme and heme *a* groups. To take into account the different sizes of the various heme ligands, we have subtracted the ligand volume from the volume of the binding pocket to give the additional volume present in the heme binding cleft. Figure 2(d) shows the distribution of these additional volume values within the heme binding site clefts identified within the proteins in the data set. In most of the multiheme proteins, there is some overlap between the clefts identified for the different heme groups. It is therefore difficult to define a volume for the binding sites of individual heme groups. Instead for multiheme proteins a single binding site cleft for all the heme groups in the protein has been determined. This has then been used to calculate the volume per heme group from which the additional volume has been calculated using the average volume of the heme groups present.

For most electron transfer proteins, there is a particularly good fit between the shape of the heme and the binding site cleft (Euclidean distance from the comparison of the coefficients describing the shape of the ligand and the cleft 5.54–7.96) and the binding site cleft has little additional volume. This contrasts with the binding site clefts in enzymes. Here there is a poorer fit between the heme group and the binding site cleft (Euclidean distance from the comparison of the coefficients describing the shape of the ligand and the cleft 7.36–10.96 excluding 1aop with the siro heme group) and a larger additional volume that enables the substrate to be accommodated adjacent to the heme group.

The shape of the heme binding clefts in the proteins in the data set have been compared using a standard Euclidean distance metric calculated from the shape coefficients from the spherical harmonic expansion description of each pair of clefts. A Euclidean distance matrix for the

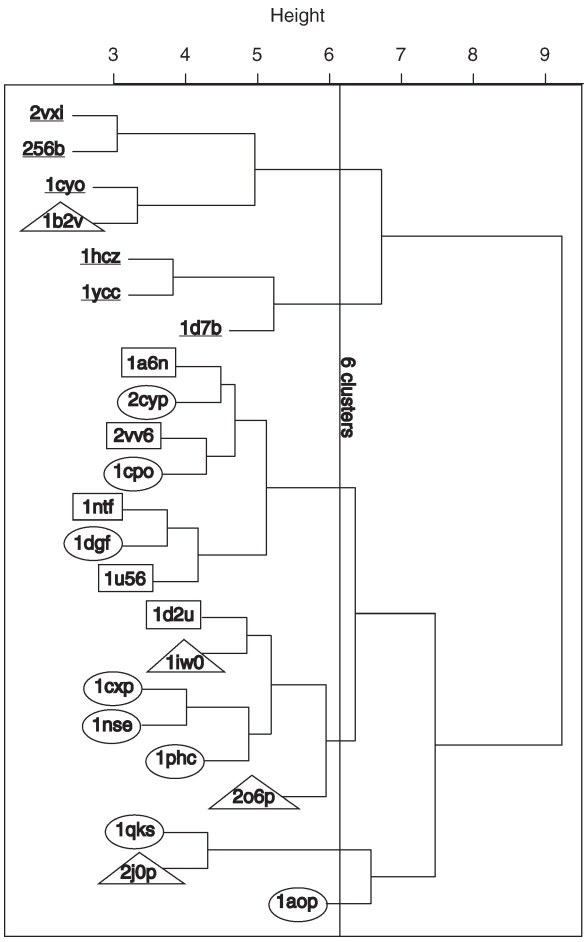

Figure 6

Clustering of the heme binding site clefts present in the data set of heme binding proteins dependent on their shape. Different markers are used to distinguish the pdb codes of the different heme group functions: electron transfer subset, underlined; binds small molecules subset, rectangle; enzyme subset, oval; transient heme binding subset, triangle. Full details of the approaches used to define the binding cleft shapes and in the clustering are given in the Methods section. Multiheme proteins have been excluded due to overlap between the binding sites of individual heme groups and seven lower resolution structures have also been excluded.

all-against-all comparison of binding clefts in proteins with single heme groups has been produced and from this the heme binding sites have been clustered according to their shape (see details in Methods section; Fig. 6). Multiheme proteins have been excluded from this part of the analysis due to difficulties in defining individual heme binding sites because of their overlap. The seven proteins with lower resolution (>2 Å) have also been excluded from this analysis (see Methods).

Compared to some other ligands that are much more flexible, the shapes of the binding sites for heme are all very similar. However, within the set six clusters of shape can be identified. Two of the clusters contain the heme binding sites of all six electron transfer proteins and the binding site of one protein showing transient heme binding (1b2v). In all the proteins in these clusters there is a close fit between the shape of the heme group and the binding site (Fig. 7 shows the binding sites for the proteins 1cyo, 1d7b, and 1b2v in these clusters).

The other four clusters contain all the proteins where the heme group is part of an enzyme active site or where the heme group binds small molecules together with three of the proteins that show transient heme binding (2o6p, 1iw0, and 2j0p). These are in general larger binding sites. The additional volume in the cleft is sometimes found localized above the iron atom on one face of the heme group. In other proteins the extra volume is to the edges of the heme group or, in the case of the largest clefts, all around the heme group (Fig. 7 shows example heme binding site for proteins 1a6n, 1d2u, 1iw0, 1cpo, and 1nse). In these proteins water molecules often fill empty space around the ligand molecules, thereby inducing an increase in geometrical complementarity between protein and

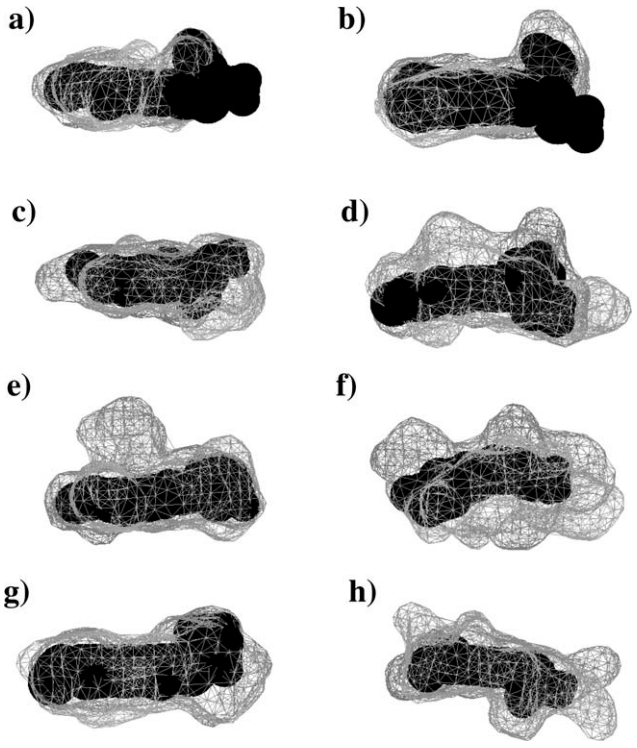

Figure 7

The shape of the heme binding cleft in the following proteins: (**a**) cytochrome b_5 (1cyo); (**b**) cellobiose dehydrogenase (1d7b); (**c**) myoglobin (1a6n); (**d**) nitrophorin 4 (1d2u); (**e**) chloroperoxidase (1cpo); (**f**) nitric oxide synthase (1nse); (**g**) heme binding protein A (1b2v); (**h**) heme oxygenase (1iw0). The heme groups shown in (a) and (b) are in the electron transfer subset, those in (c) and (d) are in the subset of proteins whose heme groups bind small molecules, those in (e) and (f) are in the enzyme subset, and those in (g) and (h) are in the subset of proteins that show transient heme binding. In each case the cleft is shown as a gray mesh and the heme group is shown in black in a space filling representation.

ligand.[49] In addition, the shape of the cleft often plays an important functional role. For example, in the enzyme chloroperoxidase (1cpo) the extra space is localized on one side of the heme group where it forms part of a channel which allows the substrate to access the buried heme active site [Fig. 7(e)][75] and in nitric oxide synthase (1nse) space in the binding cleft to the edge of the heme group allows a tetrahydrophterin cofactor to bind [Fig. 7(f)].[90] One of the clusters contains only a single heme binding site that of sulfite reductase (1aop).[65] This protein binds the siro heme group, which has quite different shape requirements due to the additional carboxylate substituents attached to its heme group (Fig. 1).

Heme reduction potentials

Reduction potentials values are available for 18 heme groups in the heme protein data set (Table I).[33] The precise tuning of the reduction potential of a heme group is of crucial importance in heme proteins, especially those where the heme is involved in electron transfer or enzyme mechanisms. For many years intense effort has been put into both experimental and theoretical studies aimed to understand the determinants of heme reduction potentials in proteins.[33,91–94] Reduction potential values depend on differences in free energy between the oxidized and reduced forms resulting from changes in bonding interactions at the redox center, from modulations in electrostatic interactions with the protein or solvent, from conformational changes to the protein or solvent structure and from differential ligand binding effects.[93,95–97] Reduction potentials also show a dependence on the extent to which the heme group is distorted from planarity[78–82] and on the protonation states of the heme propionate groups.[98–100] In the light of such a complex interplay of factors it is not surprising that no clear correlations are observed between the reduction potential values for the proteins in the heme data set and the structural parameters analyzed in this article (heme solvent accessibility, number of protein-heme hydrogen bonds, and nonbonded contacts and volume of heme binding site—data shown in Supporting Information). However, we note that when groups of very closely related heme proteins are compared some correlations can be observed, particularly between the heme reduction potential and the solvent accessibility of the heme group.[71,101]

Proteins with multiple heme groups

Six proteins in the data set are multiheme proteins, three proteins having two heme groups, two proteins having four heme groups and one protein having six heme groups in the structure. The majority of these heme groups have an electron transfer role, although in cytochrome *c* oxidase (1v54) and hydroxylamine oxidoreductase (1fgj) one of the heme groups is in the enzyme active site. Multiple heme groups are recognized to be important in enabling the proteins to participate in metabolic pathways that require multiple electron transfer steps.[102–104] In some cases, the proteins are involved in a series of one electron reductions while in other cases pairs of electrons may be transferred For example, two electron step reactions are thought to occur in the reaction catalyzed by hydroxylamine oxidoreductase (catalyses the conversion of hydroxylamine into nitrite).[63]

The relative orientations of heme groups within proteins in the hydroxylamine oxidoreducatse and cytochrome c_3 families have been extensively analyzed.[9,30–31,105] The *c*-type heme groups in proteins in the hydroxylamine oxidoreductase family, including hydroxylamine oxidoreductase itself,[63] octaheme tetrathionate reductase,[106] and pentaheme cytochrome *c* nitrite reductase,[107] have been found to have a conserved arrangement despite the differing 3D structures of the proteins themselves.[30,105] Pairs of heme groups which are close in space have either a parallel stacked arrangement (heme stacking motif) or a perpendicular arrangement (diheme elbow motif). The perpendicular arrangement of heme groups is also seen in proteins in the cytochrome c_3 family.[31]

Here we have extended this analysis of heme orientation to include all the multiheme proteins in the data set, including the cytochrome *b* subunit from fumerate reductase (1qla),[108] the cytochrome subunit of the photosynthetic reaction center (2i5n),[109] and cytochrome *c* oxidase (1v54) that has two *a*-type heme groups.[64] In all the multiheme proteins, except for the cytochrome *b* subunit from cytochrome bc_1 (1pp9), there are pairs of heme groups with an Fe-Fe distance of less than 15.5 Å. Concentrating on these pairs, Figure 8(a) shows that an approximately perpendicular arrangement between the two heme groups predominates, with the parallel stacked arrangement being seen only in hydroxylamine oxidoreductase [Fig. 8(b)]. It has been suggested that these heme geometries may help facilitate rapid electron transfer between the groups, and that interactions between these heme groups which are close in space may be important in tuning their redox potential values.[30,31,110] The parallel and perpendicular arrangements of heme groups may also be an energetically favorable way of packing together a number of hemes in quite a small protein allowing favorable electrostatic interactions between their π systems.[88]

A further feature, apparent from the range of proteins studied here, is that when a pair of heme groups are in the perpendicular arrangement then the relative orientation of the two heme groups can either be face-to-edge or edge-to-edge. The face-to-edge orientation is seen particularly in the pairs of heme groups in hydroxylamine oxidoreductase, and also in cytochrome *c* oxidase and between heme groups 1 in 2 in cytochrome c_3. In contrast, the edge-to-edge arrangement is seen in the cytochrome subunit from photosynthetic reaction center (hemes 1 and 2, and hemes 3 and 4), in the cytochrome *b* subunit from fumerate reductase and between heme groups 3 and 4 in cytochrome c_3 [Fig. 8(a)].

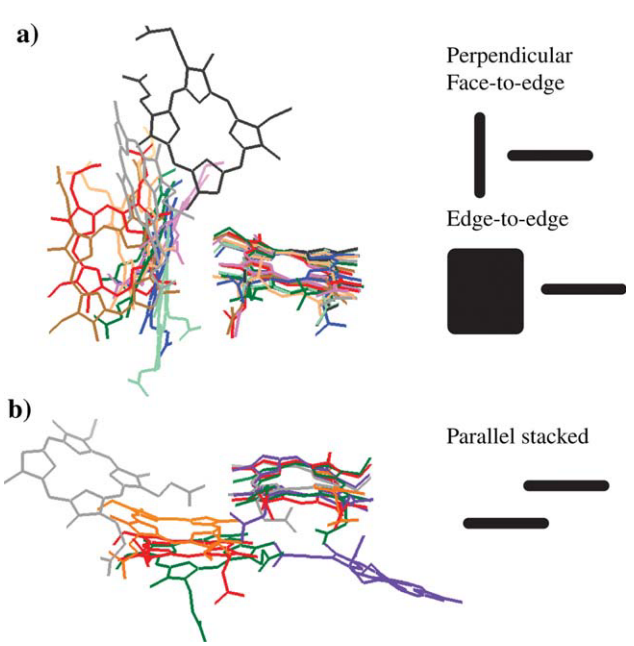

Figure 8

The relative orientation of pairs of heme groups in proteins with multiple heme groups. Panels (**a**) and (**b**) shows a superposition of pairs of heme group with an approximately perpendicular or parallel arrangement, respectively. In each case one of the heme groups in the pair has been superimposed on hemes 1 (panel b) and 2 (panel a) of hydroxylamine oxidoreductase to give the best match. The following heme groups are shown: (a) hydroxylamine oxidoreductase (1fgj) hemes 2 and 3 (blue), hemes 5 and 6 (magenta), hemes 7 and 8 (dark green), cytochrome c_3 (2cy3) hemes 1 and 2 (black), hemes 1 and 3 (light green), hemes 3 and 4 (orange), cytochrome subunit from photosynthetic reaction center (2i5n) hemes1 and 2 (red), hemes 3 and 4 (yellow), fumerate reductase cytochrome b subunit (1qla) hemes 1 and 2 (brown), cytochrome c oxidase (1v54), and hemes 1 and 2 (light gray). (b) Hydroxylamine oxidoreductase (1fgj) hemes 1 and 2 (red), hemes 3 and 5 (green), hemes 4 and 6 (purple), hemes 4 and 7 (gray), hemes 6 and 7 (orange). The different relative orientation of pairs of heme groups seen are shown schematically on the right.

Apo state characteristics

Significant differences are seen between heme proteins on removal of the heme group(s) and forming the apo state. Experimental data are available regarding the apo states of nine of the proteins in the data set (either for the identical protein or a close homologue) (Table III). The most disordered apo state is reported for apo mitochondrial cytochrome *c* which at low ionic strength is essentially an unfolded random coil.[111,112] In *c*-type cytochromes formation of the covalent thioether bonds between the heme group and the protein is a catalyzed posttranslational modification process.[5,113] These covalent thioether bonds are thought to form while the protein is largely unfolded. Indeed, it has been shown that the cytochrome *c* biogenesis system can even attach heme groups to small peptide fragments containing a CXXCH motif.[114] Once the heme group is attached to a polypeptide chain it can then provide a nucleation site for folding.[115] This is likely to be particularly important in multiheme proteins such as cytochrome c_3 where there is a high heme-to-polypeptide chain ratio.

When the heme group is not covalently attached to the protein, as with *b*-type heme groups, recognition between the heme group and the binding site in the protein is required. Here the data available show that the apo states contain at least a nascent heme binding site. In some cases, such as cytochrome P-450$_{cam}$, the apo state has molten globule like characteristics[116] while in others, such as HemS, the apo state is fully folded and closely similar to the holo form.[117] Differences are seen between the proteins that bind heme permanently in their active forms and those that show transient heme binding (Fig. 9).

Data are available for five proteins that bind a heme group permanently, namely cytochromes b_{562} and b_5 (electron transfer subset),[118–121] myoglobin[122] and histidine kinase FixL[123] (subset of proteins that bind small molecules) and cytochrome P-450$_{cam}$[116] (enzyme subset) (Table III). Here a significant disorder is seen for at least part of the heme binding pocket in the apo state, and in some cases this disorder extends throughout the structure. In apomyoglobin, for example, at pH 6.1 the disorder is localized to part of the heme binding pocket. NMR data show that overall apomyoglobin adopts a well-defined structure, very similar to that of the holo protein, but with conformational disorder for residues in the EF loop, F helix, FG loop and at the start of helix G in the heme binding site.[122] A similar situation is seen in cytochrome b_{562}, where the holo and apo states have similar structures overall but the loop connecting helix II and III and the last section of helix IV, which form part of the heme binding site, are disordered in the apo state.[118,119] However, the apo states of cytochrome P-450$_{cam}$ and histidine kinase FixL are more disordered overall with some molten globule-like characteristics.[116,123]

This situation is very different in the proteins in the subset that show transient heme binding. Here there are data available for three proteins, heme oxygenase,[124] heme binding protein A[125] and heme transport protein HemS.[117] In each case both the apo and holo states are fully folded but changes are seen in the relative orientation of folded units on heme binding (Fig. 9). For example, in HemS on heme binding the protein structure changes from an open to a closed form through global rigid body-like movements of the folded N- and C-terminal domains.[117] In a similar manner, heme binding protein A swaps from an open to closed form on binding heme by movement of loop L1 (which has a well-defined conformation in both the holo and apo states).[125]

DISCUSSION

The analysis of heme groups in the data set of 34 different proteins reported here has demonstrated the diver-

Table III
Characteristics of the Apo States of Heme Binding Proteins

Protein	Functional subset	Type of heme	Apo state characteristics
Horse heart cytochrome *c*[111,112]	Electron transfer	1 *c*-type	Apo state is an unfolded random coil at low ionic strength.
Cytochrome b_{562} from *E. coli*[118,119]	Electron transfer	1 *b*-type	Apo state has same topology as holo state but residues 42–55 in the heme binding site, and residues at the protein termini, are disordered.
Rat cytochrome b_5[120,121]	Electron transfer	1 *b*-type	Apo state has a folded core but the region around the heme binding pocket is disordered (residues 43–69). ^{15}N relaxation data show that apo state has more mobility overall than the holo state.
Sperm whale myoglobin[122]	Binds small molecules	1 *b*-type	Apo state has well-defined structure at pH 6.1 which is closely similar to that of the holo state. However, conformational disorder is seen for residues 82–101 in heme binding site
Histidine kinase FixL heme domain from *Bradyrhizobium japonicum*[123]	Binds small molecules	1 *b*-type	Apo state is partially folded with reduced secondary structure compared to the holo state. The NMR spectrum of the apo state shows broad lines and a lack of chemical shift dispersion.
Cytochrome P-450$_{cam}$ from *Pseudomonas putida*[116]	Enzyme	1 *b*-type	Apo state has molten globule-like characteristics. It is susceptible to proteolysis and is prone to aggregation.
Heme complex of human heme oxygenase-1[124]	Transient heme binding	1 *b*-type	Apo state is fully folded. Some small differences from the holo state in the heme binding pocket, particularly relating to the relative positions of the proximal and distal helices.
Heme binding protein A from *Serratia marcescens*[125]	Transient heme binding	1 *b*-type	Apo state is fully folded. On heme binding residues 29–45 in loop L1 fold onto the protein core.
Heme transport protein HemS from *Yersinia enterocolitica*[117]	Transient heme binding	1 *b*-type	Apo state is fully folded. There is a rigid body-like movement of the N- and C-terminal domains on heme binding.

sity of heme prosthetic groups in proteins. Heme itself is not one ligand but a family of related ligands all of which share the same central skeleton but differ in their side chain substituents and in their mode of attachment and coordination to the protein chain. Nature uses this variety of heme groups, together with a diversity in the heme binding site clefts in the proteins to which they bind, to enable these related groups to perform a range of functions. Despite the differences between the proteins analyzed various important principles have been highlighted in this work. In particular, all heme groups are significantly buried within the protein. In the data set the mean heme solvent accessibility is 84.1 Å^2 compared to a solvent accessibility of 830 Å^2 for a free heme group. The heme group binds in a pocket that contain significant numbers of nonpolar side chains, particularly those of aromatic residues, making substantial numbers of nonbonded contacts to the protein. In the data set for heme groups that are permanently associated with a protein on average 54.4 $\pm$ 12.5 nonbonded contacts are seen. Heme groups generally bind in a mode that exposes the propionate side chains to solvent (binding mode 1). However, if these propionate groups are buried there are polar groups in the protein, particularly arginine side chains, positioned in the binding cavity to hydrogen bond to the propionate groups.

These overall properties provide an important framework for the design of *de novo* heme binding proteins. Much progress is being made in this area at present.[15–20] However, it is still proving challenging to determine a high-resolution structure of a *de novo* heme binding proteins in the holo state presumably due to the difficulties of crystallising these proteins in the heme bound state. However, Zou *et al.*[28] have reported in the literature a model from molecular dynamics simulations for a *de novo* designed four helix bundle maquette that has been studied extensively experimentally.[26,126] There are two *b*-type heme groups bound to the protein in this model in a parallel staggered arrangement. Both heme groups adopt binding mode 1 but have a high solvent accessibility (239 Å^2 and 236 Å^2), a very low number of nonbonded contacts (28 and 25) and no protein-heme hydrogen bonds. The heme binding properties of the groups in the model structure are therefore very different from those seen in the naturally occurring heme proteins analyzed. It is likely that this protein shows much less specificity in its heme binding and it is not surprising that it has not been possible to obtain a crystal structure of the holo state of this protein.

The analysis reported here shows that there are recognizable characteristics typical of the heme groups and binding site properties for heme groups playing the dif-

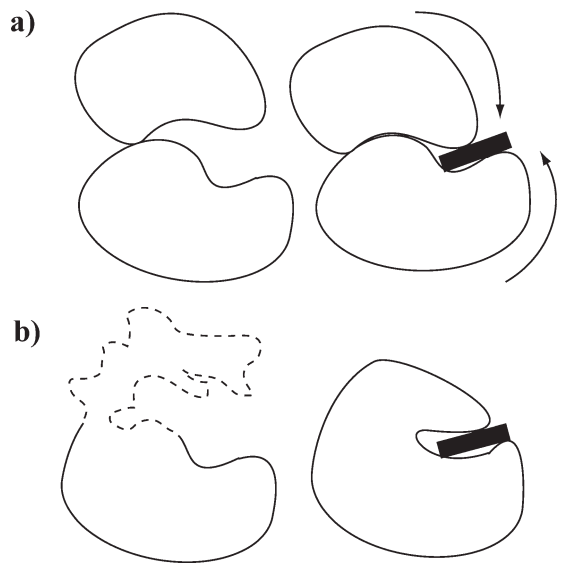

Figure 9

A schematic comparison of the characteristics of the apo and holo states of proteins which show transient heme binding (**a**) and those which bind heme permanently (**b**). Proteins which show transient heme binding tend to have a fully folded apo state. Changes are seen in the relative orientations of folded units on heme binding. In contrast proteins that bind heme permanently show significantly disorder in the heme binding region in their apo states. These proteins become fully folded only on heme binding.

ferent functional roles in biology. In particular, in proteins where the heme group plays an electron transfer role, heme *b* or *c* binds in a small hydrophobic pocket whose shape is closely complementary to the shape of the heme group. Similar characteristics are seen for proteins with heme groups that bind small molecules although in some cases the binding site volumes are slightly larger, presumably to allow entry and binding of the small molecules. In multiheme proteins, heme *c* predominates, with a preference for a perpendicular arrangement between heme groups that are close to each other. In cases such as cytochrome c_3, where there is a high heme-to-polypeptide chain ratio, heme *c* may be necessary to ensure that all heme groups stay permanently associated with the protein. Indeed for heme 2 in cytochrome c_3 there are only 38 nonbonded contacts and no hydrogen bonds between the heme group and the protein (compared to mean values in the data set of 52.8 and 4.4 for the number of protein-heme nonbonded contacts and hydrogen bonds respectively). Again the heme binding pocket is very nonpolar in multiheme protein, in keeping with the electron transfer role seen for most of the heme groups in this subset. However, a much greater variety of heme binding modes, heme group solvent accessibility and the proportion of nonbonded contacts compared to hydrogen bonds between the heme and protein is seen compared to the single heme electron transfer proteins.

In contrast to the proteins where the heme group is involved in electron transfer or binding small molecules, a significant diversity is seen in the proteins where the heme groups is part of an enzyme active site. This subset of proteins contains almost all the different heme variants (heme P460, siro heme, heme d_1, heme in myeloperoxidase) with a range of groups in the protein acting as axial ligands for the heme (including His, Tyr, and Cys). The heme binding pockets tend to be larger with a variety of shapes and contain more polar and charged side chains. The subset of proteins that show transient heme binding are particularly interesting. There is again a significant diversity within the class but in many cases the heme groups have a higher solvent accessibility and show heme binding mode 2 or 3 which leaves at least one propionate group buried. Overall the heme groups make fewer nonbonded contacts (mean number 38.6) but a significant number of hydrogen bonds (mean number 4) to the protein within the binding site. These differences presumably reflect the different requirements when heme is bound as a ligand which can dissociate from the protein. Protein-heme hydrogen bonds, particularly when at least one of the heme propionate side chains is buried, allow for modulation in heme binding with pH or ionic strength. The differences between the proteins in this class are likely to reflect the fact that they function in different environments *in vivo* and require different heme dissociation constants for their roles.

The distinguishable characteristics of heme groups playing different functions in biology identified here provide possibilities for identifying the role of heme groups in proteins of unknown function. There are two proteins in the data set where the role of the heme group is not certain. These are human cystathionine β-synthase (1jbq)[127] and the structural genomics target, putative melanine biosynthesis protein Tyr A (2iiz).[128] Cystathionine β-synthase is an enzyme that catalyzes the condensation of serine and homocysteine to give cystathionine. The heme motif is approximately 20 Å from the active site of enzyme and so it is unlikely that the heme group plays a direct role in the enzyme mechanism.[129,130] Furthermore, mutational studies have shown that the enzyme retains activity in the absence of the heme group, although the heme is needed for maximal activity.[131] It has been suggested that the heme may provide structural support in folding and that it could have a regulatory role via redox changes, or by the binding of small molecules at heme prompting conformational changes.[127,129,130] Comparison with the shapes of the heme binding clefts of the other proteins in the heme data set shows that the heme binding site in cystathionine β-synthase is most similar to that from the H-NOX oxygen binding domain of the chemotaxis receptor from *Thermoanaerobacter tengcongensis* (1u56) (binds small molecules subset). The other properties of the heme binding site in cystathionine β-synthase would also be consistent with the proposed regulatory role for this heme group via

small molecule binding or redox changes. For example, the heme binding site has a relatively low volume (1117 Å^3) and there are a substantial number of nonbonded contacts between the heme group and the protein (61), especially involving nonpolar side chains.

The characteristics of the heme binding site in the structural genomics target putative melanine biosynthesis protein Tyr A (2iiz) are quite different to those of cystathionine β-synthase. The most notable feature is the significant number of charged and polar side chains in the heme binding pocket (Asp 145, Glu 154, His 187, Arg 206, His 224, Asn 229, Arg 242) including two that have negatively charged side chains. This together with the larger volume of the heme binding site (1291 Å^3) would be consisting with the proposal that this heme group forms part of an enzyme active site.

The analysis reported in this article has demonstrated not only the diversity in heme groups and their binding sites in proteins, but also the differences in apo state characteristics that different heme proteins display. The characteristics of the apo state of a protein have thermodynamic consequences as well as affecting its susceptibility to degradation. A very important feature for the proteins that show transient heme binding is that their apo states, as well as their holo states, need to be active in biological systems resisting proteolysis or aggregation. This may be a key reason why the apo states of proteins that show transient heme binding are fully folded. For proteins which bind a heme group permanently in their active form, an apo state with some disorder may enable ready degradation of this inactive form. Disorder in the apo state also allows the subtle modulation of heme binding affinity, by a balance between entropy and enthalpy effects.[123] When a significant degree of folding occurs on heme binding, the unfavorable entropy changes may be compensated for, at least in part, by enthalpically favorable heme-protein contacts. Here it is interesting that the analysis reported in this paper has shown that proteins that bind heme groups permanently, and have more disordered apo states, in most cases have more nonbonded protein-heme contacts than those which bind heme as a ligand and have very well-defined apo states. This observation shows again the complex interplay between protein folding and dynamics, the characteristics of the heme binding site and the redox characteristics of the groups involved which are finely tuned together to give an optimal system for the required biological role.

ACKNOWLEDGMENTS

We thank Professor Geoff Moore for helpful discussions, Julia Fischer for helping with the clustering analysis, Professor Kent Blasie and Ms. Hongling Zou for providing the coordinates of their MD model for the *de novo* four bundle protein and Professor Brian Gibney for help with accessing the Heme Protein Database.

REFERENCES

1. Edsall JT. Blood and hemoglobin: the evolution of knowledge of functional adaption in a biochemical system. J History Biol 1972;5:205–257.
2. Fischer H, Zeile K. Synthese des haematoporphyrins, protoporphyrins und haemins. Liebigs Ann Chem 1929;468:98–116.
3. Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wyckoff H, Phillips DC. 3-Dimensional model of the myoglobin molecule obtained by X-ray analysis. Nature 1958;181:662–666.
4. Perutz MF, Rossmann MG, Cullis AF, Muirhead H, Will G, North ACT. Structure of haemoglobin—3-dimensional Fourier synthesis at 5.5-Å resolution, obtained by X-ray analysis. Nature 1960;185:416–422.
5. Stevens JM, Daltrop O, Allen JWA, Ferguson SJ. *c*-type cytochrome formation: chemical and biological enigmas. Accounts Chem Res 2004;37:999–1007.
6. Reedy CJ, Gibney BR. Heme protein assemblies. Chem Rev 2004;104:617–649.
7. Bradley AL, Chobot SE, Arciero DM, Hooper AB, Elliott SJ. A distinctive electrocatalytic response from the cytochrome c peroxidase of *Nitrosomonas europaea*. J Biol Chem 2004;279:13297–13300.
8. Anderson JLR, Chapman SK. Ligand probes for heme proteins. Dalton Trans 2005;1:13–24.
9. Bertini I, Cavallaro G, Rosato A. Cytochrome *c*: occurrence and functions. Chem Rev 2006;106:90–115.
10. Ow YLP, Green DR, Hao Z, Mak TW. Cytochrome c: functions beyond respiration. Nature Rev Mol Cell Biol 2008;9:532–542.
11. Caillet-Saguy C, Turano P, Piccioli M, Lukat-Rodgers GS, Czjzek M, Guigliarelli B, Izadi-Pruneyre N, Rodgers KR, Delepierre M, Lecroisey A. Deciphering the structural role of histidine 83 for heme binding in hemophore HasA. J Biol Chem 2008;283:5960–5970.
12. Berry RE, Shokhirev MN, Ho AYW, Yang F, Shokhireva TK, Zhang HJ, Weichsel A, Montfort WR, Walker FA. Effect of mutation of carboxyl side-chain amino acids near the heme on the midpoint potentials and ligand binding constants of nitrophorin 2 and its NO. histamine, and imidazole complexes. J Am Chem Soc 2009;131:2313–2327.
13. Krieg S, Huché F, Diederichs K, Izadi-Pruneyre N, Lecroisey A, Wandersman C, Delepelaire P, Welte W. Heme uptake across the outer membrane as revealed by crystal structures of the receptor-hemophore complex Proc Natl Acad Sci USA 2009;106:1045–1050.
14. Gibney BR, Dutton PL. De novo design and synthesis of heme proteins. Adv Inorg Chem 2001;51:409–455.
15. Monien BH, Drepper F, Sommerhalter M, Lubitz W, Haehnel W. Detection of heme oxygenase activity in a library of four-helix bundle proteins: Towards the de novo synthesis of functional heme proteins. J Mol Biol 2007;371:739–753.
16. Cordova JM, Noack PL, Hilcove SA, Lear JD, Ghirlanda G. Design of a functional membrane protein by engineering a heme-binding site in glycophorin A. J Am Chem Soc 2007;129:512–518.
17. Isogai Y, Ishida M. Design of a novel heme protein with a non-heme globin scaffold. Biochemistry 2009;48:8136–8142.
18. Koder RL, Anderson JLR, Solomon LA, Reddy KS, Moser CC, Dutton PL. Design and engineering of an O_2 transport protein. Nature 2009;458:305–309.
19. Yeung N, Lin YW, Gao YG, Zhao X, Russell BS, Lei LY, Miner KD, Robinson H, Lu Y. Rational design of a structural and functional nitric oxide reductase. Nature 2009;462:1079–1082.
20. Lu Y, Yeung N, Sieracki N, Marshall NM. Design of functional metalloproteins. Nature 2009;460:855–862.
21. Robertson DE, Farid RS, Moser CC, Urbauer JL, Mulholland SE, Pidikiti R, Lear JD, Wand AJ, DeGrado WF, Dutton PL. Design and synthesis of multi-heme proteins. Nature 1994;368:425–432.
22. Ghirlanda G, Osyczka A, Liu WX, Antolovich M, Smith KM, Dutton PL, Wand AJ, DeGrado WF. De novo design of a D-2-symmetrical protein that reproduces the diheme four-helix bundle in cytochrome bc_1. J Am Chem Soc 2004;126:8141–8147.

23. Huang SS, Koder RL, Lewis M, Wand AJ, Dutton PL. The HP-1 maquette: from an apoprotein structure to a structured hemoprotein designed to promote redox-coupled proton exchange. Proc Natl Acad Sci USA 2004;101:5536–5541.
24. Koder RL, Dutton PL. Intelligent design: the de novo engineering of proteins with specified functions. Dalton Trans 2006;25:3045–3051.
25. Ramanavicius A, Ramanaviciene A. Hemoproteins in design of biofuel cells. Fuel Cells 2009;9:25–36.
26. Ye SX, Strzalka JW, Discher BM, Noy D, Zheng SY, Dutton PL, Blasie JK. Amphiphilic 4-helix bundles designed for biomolecular materials applications. Langmuir 2006;22:512–512.
27. Ueno T, Yokoi N, Unno M, Matsui T, Tokita Y, Yamada M, Ikeda-Saito M, Nakajima H, Watanabe Y. Design of metal cofactors activated by a protein-protein electron transfer system. Proc Natl Acad Sci USA 2006;103:9416–9421.
28. Zou HL, Strzalka J, Xu T, Tronin A, Blasie JK. Three-dimensional structure and dynamics of a de novo designed, amphiphilic, metallo-porphyrin-binding protein maquette at soft interfaces by molecular dynamics simulations. J Phys Chem B 2007;111:1823–1833.
29. Lehmann A, Saven JG. Computational design of four-helix bundle proteins that bind nonbiological cofactors. Biotechnol Progr 2008;24:74–79.
30. Mowat CG, Chapman SK. Multi-heme cytochromes—new structures, new chemistry. Dalton Trans 2005;21:3381–3389.
31. Pereira IAC, Xavier AV. Multi-heme cytochromes and enzymes. In: King RB editor. Encyclopedia of inorganic chemistry, 2nd ed. New York: Wiley; 2005.
32. Schneider S, Marles-Wright J, Sharp KH, Paoli M. Diversity and conservation of interactions for binding heme in *b*-type heme proteins. Nat Prod Reports 2007;24:621–630.
33. Reedy CJ, Elvekrog MM, Gibney BR. Development of a heme protein structureelectrochemical function database. Nucl Acids Res 2008;36:D307–D313.
34. Fufezan C, Zhang J, Gunner MR. Ligand preference and orientation in *b*- and *c*-type heme-binding proteins. Proteins: Struc Func Bioinf 2008;73:690–704.
35. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. Nucleic Acids Res 2000;28:235–242.
36. Orengo CA, Michie AD, Jones DT, Swindells MB, Thornton JM. CATH: a hierarchic classification of protein domain structures. Structure 1997;5:1093–1108.
37. Cuff AL, Sillitoe I, Lewis T, Redfern OC, Garratt R, Thornton JM, Orengo CA. The CATH classification revisited-architectures reviewed and new ways to characterize structural divergence in superfamilies. Nucleic Acids Res 2009;37:D310–D314.
38. Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 1995;247:536–540.
39. Andreeva A, Howorth D, Chandonia JM, Brenner SE, Hubbard TJP, Chothia C, Murzin AG. Data growth and its impact on the SCOP database: new developments. Nucl Acid Res 2008;36:D419–D425.
40. Hubbard SJ, Thornton JM. 'NACCESS', computer program, Department of Biochemistry and Molecular Biology, University College London; 1993.
41. Lee B, Richards FM. Interpretation of protein structures—estimation of static accessibility. J Mol Biol 1971;55:379–400.
42. McDonald IK, Thornton JM. Satisfying hydrogen-bonding potential in proteins. J Mol Biol 1994;238:777–793.
43. Pauling L, Corey RB, Branson HR. The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain. Proc Nat Acad Sci USA 1951;37:205–211.
44. Momany FA, McGuire RF, Burgess AW, Scheraga HA. Energy parameters in polypeptides. VII. Geometric parameters, partial atomic charges, nonbonded interactions, hydrogen bond interactions, and intrinsic torsional potentials for the naturally occuring amino acids. J Phys Chem 1975;79:2373–2381.
45. Wallace AC, Laskowski RA, Thornton JM. LIGPLOT—a program to generate schematic diagrams of protein ligand interactions. Protein Eng 1995;8:127–134.
46. Glaser F, Rosenberg Y, Kessel A, Pupko T, Ben-Tal N. The ConSurf-HSSP database: the mapping of evolutionary conservation among homologs onto PDB structures. Proteins: Struct Func Bioinform 2005;58:610–617.
47. Sander C, Schneider R. Database of homology-derived protein structures and the structural meaning of sequence alignment. Proteins: Struct Funct Bioinform 1991;9:56–68.
48. Pupko T, Bell RE, Mayrose I, Glaser F, Ben Tal N. Rate4Site: an algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues. Bioinformatics 2002;18:S71–S77.
49. Kahraman A, Morris RJ, Laskowski RA, Thornton JM. Shape variation in protein binding pockets and their ligands. J Mol Biol 2007;368:283–301.
50. Laskowski RA. SURFNET—a program for visualizing molecular-surfaces, cavities, and intermolecular interactions. J Mol Graph 1995;13:323–330.
51. Morris RJ, Najmanovich RJ, Kahraman A, Thornton JM. Real spherical harmonic expansion coefficients as 3D shape descriptors for protein binding pocket and ligand comparisons. Bioinformatics 2005;21:2347–2355.
52. Ward JH, Jr. Hierarchical grouping to optimize an objective function. Am Stat Assoc 1963;58:236–244.
53. Struyf A, Hubert M, Rousseeuw PJ. Integrating robust clustering techniques in S-PLUS. Comput Stat Data Anal 1997;26:17–37.
54. Kelley LA, Gardner SP, Sutcliffe MJ. An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally related subfamilies. Protein Eng 1996;9:1063–1065.
55. Paoli M, Anderson BF, Baker HM, Morgan WT, Smith A, Baker EN. Crystal structure of hemopexin reveals a novel high-affinity heme site formed between two beta-propeller domains. Nature Struc Biol 1999;6:926–931.
56. Hirotsu S, Chu GC, Unno M, Lee DS, Yoshida T, Park SY, Shiro Y, Ikeda-Saito M. The crystal structures of the ferric and ferrous forms of the heme complex of HmuO, a heme oxygenase of *Corynebacterium diphtheriae*. J Biol Chem 2004;279:11937–11947.
57. Roberts SA, Weichsel A, Qiu Y, Shelnutt JA, Walker FA, Montfort WR. Ligand-induced heme ruffling and bent NO geometry in ultra-high-resolution structures of nitrophorin 4. Biochemsitry 2001;40:11327–11337.
58. Fulop V, Moir JWB, Ferguson SJ, Hajdu J. The anatomy of a bifunctional enzyme—structural basis for reduction of oxygen to water and synthesis of nitric-oxide by cytochrome cd_1. Cell 1995;81:369–377.
59. Sharp KH, Schneider S, Cockayne A, Paoli M. Crystal structure of the heme-IsdC complex, the central conduit of the Isd iron/heme uptake system in *Staphylococcus aureus*. J Biol Chem 2007;282:10625–10631.
60. Allen JWA, Barker PD, Daltrop O, Stevens JM, Tomlinson EJ, Sinha N, Sambongi Y, Ferguson SJ. Why isn't 'standard' heme good enough for *c*-type and *d*(1)-type cytochromes? Dalton Trans 2005;21:3410–3418.
61. Czjzek M, Payan F, Guerlesquin F, Bruschi M, Haser R. Crystal-structure of cytochrome c_3 from *Desulfovibrio-desulfuricans* Norway at 1.7 Å resolution. J Mol Biol 1994;243:653–667.
62. Stroebel D, Choquet Y, Popot JL, Picot D. An atypical haem in the cytochrome b_6 of complex. Nature 2003;426:413–418.
63. Igarashi N, Moriyama H, Fujiwara T, Fukumori Y, Tanaka N. The 2.8 Å structure of hydroxylamine oxidoreductase from a nitrifying chemoautotrophic bacterium, *Nitrosomonas europaea*. Nature Struc Biol 1997;4:276–284.
64. Yoshikawa S, Shinzawa-Itoh K, Nakashima R, Yaono R, Yamashita E, Inoue N, Yao M, Fei MJ, Libeu CP, Mizushima T, Yamaguchi H, Tomizaki T, Tsukihara T. Redox-coupled crystal structural changes in bovine heart cytochrome c oxidase. Science 1998;280:1723–1729.

65. Crane BR, Siegel LM, Getzoff ED. Sulfite reductase structure at 1.6 Å—evolution and catalysis for reduction of inorganic anions. Science 1995;270:59–67.
66. Fiedler TJ, Davey CA, Fenna RE. X-ray crystal structure and characterization of halide-binding sites of human myeloperoxidase at 1.8 Å resolution. J Biol Chem 2000;275:11964–11971.
67. Abramson J, Riistama S, Larsson G, Jasaitis A, Svensson-Ek M, Laakkonen L, Puustinen A, Iwata S, Wikstrom M. The structure of the ubiquinol oxidase from Escherichia coli and its ubiquinone binding site. Nature Struc Biol 2000:7;910–917.
68. Bravo J, Verdaguer N, Tormo J, Betzel C, Switala J, Loewen PC, Fita I. Crystal-structure of catalase HPII from *Escherichia-coli* structure 1995;3:491–502.
69. Vinogradov SN. The structure of invertebrate extracellular hemoglobins (erythrocruorins and chlorocruorins). Comp Biochem Physiol 1985;82B:1–15.
70. Lanzilotta WN, Schuller DJ, Thorsteinsson MV, Kerby RL, Roberts GP, Poulos TL. Structure of the CO sensing transcription activator CooA. Nature Struc Biol 2000;7:876–880.
71. Stellwagen E. Heme exposure as determinate of oxidation-reduction potential of heme proteins. Nature 1978;275:73–74.
72. Barker PD, Ferguson SJ. Still a puzzle: why is haem covalently attached in c-type cytochromes? Structure 1999;7:R281–R290.
73. Hallberg BM, Bergfors T, Backbro K, Pettersson G, Henriksson G, Divne C. A new scaffold for binding haem in the cytochrome domain of the extracellular flavocytochrome cellobiose dehydrogenase. Structure 2000;8:79–88.
74. Li YM, Syvitski RT, Chu GC, Ikeda-Saito M, La Mar GN. Solution H-1 NMR investigation of the active site molecular and electronic structures of substrate-bound, cyanide-inhibited HmuO, a bacterial heme oxygenase from *Corynebacterium diphtheriae*. J Biol Chem 2003;278:6651–6663.
75. Sundaramoorthy M, Terner J, Poulos TL. The crystal structure of chloroperoxidase: a heme peroxidase-cytochrome P450 functional hybrid. Structure 1995;3:1367–1377.
76. Louie GV, Brayer GD. High-resolution refinement of yeast iso-1-cytochrome-*c* and comparisons with other eukaryotic cytochromes-*c*. J Mol Biol 1990;214:527–555.
77. Finzel BC, Poulos TL, Kraut J. Crystal-structure of yeast cytochrome-c peroxidase refined at 1.7-Å resolution. J Biol Chem 1984;259:3027–3036.
78. Maes EM, Roberts SA, Weichsel A, Montfort WR. Ultrahigh resolution structures of nitrophorin 4: heme distortion in ferrous CO and NO complexes. Biochemistry 2005;44:12690–12699.
79. Ravikanth M, Chandrashekar TK. Nonplanar porphyrins and their biological relevance—ground and excited-state dynamics. Coordination Chem 1995;82:105–118.
80. Barkigia KM, Chantranupong L, Smith KM, Fajer J. Structural and theoretical-models of photosynthetic chromophores—implications for redox, light-absorption properties and vectorial electron flow. J Am Chem Soc 1998;110:7566–7567.
81. Shelnutt JA, Song XZ, Ma JG, Jia SL, Jentzen W, Medforth CJ. Nonplanar porphyrins and their significance in proteins. Chem Soc Rev 1998;27:31–41.
82. Olea C, Boon EM, Pellicena P, Kuriyan J, Marletta MA. Probing the function of heme distortion in the H-NOX family. Chem Biol 2008;3:703–710.
83. Putnam CD, Arvai AS, Bourne Y, Tainer JA. Active and inhibited human catalase structures: ligand and NADPH binding and catalytic mechanism. J Mol Biol 2000;296:295–309.
84. Mullereberhard U. Hemopexin. Methods Enzymol 1988;163:536–565.
85. Kahraman A Morris RJ, Laskowski RA, Favia AD, Thornton JM. On the diversity of physicochemical environments experienced by identical ligands in binding pockets of unrelated proteins. Proteins: Struc, Funct, Bioinf 2010;78:1120–1136.
86. Hunter CA, Singh J, Thornton JM. π-π-interactions—the geometry and energetics of phenylalanine-phenylalanine interactions in proteins. J Mol Biol 1991;218:837–846.
87. McGaughey GB, Gagne M, Rappe AK. π-stacking interactions—alive and well in proteins. J Mol Biol 1998;273:15458–15463.
88. Marsili S, Chelli R, Schettino V, Procacci P. Thermodynamics of stacking interactions in proteins. Phys Chem Chem Phys 2008;10:2673–2685.
89. Durley RCE, Mathews FS. Refinement and structural analysis of bovine cytochrome b_5 at 1.5 Å resolution. Acta Crystal 1996;52:65–76.
90. Raman CS, Li HY, Martasek P, Southan G, Masters BSS, Poulos TL. Crystal structure of constitutive endothelial nitric oxide synthase: a paradigm for pterin function involving a novel metal center. Cell 1998;95:939–950.
91. Kassner RJ. Theoretical model for effects of local nonpolar heme environments on redox potentials in cytochromes. J Am Chem Soc 1973;95:2674–2677.
92. Moore GR, Pettigrew GW, Rogers NK. Factors influencing redox potentials of electron-transfer proteins. Proc Natl Acad Sci USA 1986;83:4998–4999.
93. Mauk AG, Moore GR. Control of metalloprotein redox potentials: What does site-directed mutagenesis of hemoproteins tell us? J Biol Inorg Chem 1997;2:119–125.
94. Mao JJ, Hauser K, Gunner MR. How cytochromes with different folds control heme redox potentials. Biochemistry 2003;42:9829–9840.
95. Liu Y, Moenne-Loccoz P, Hildebrand DP, Wilks A, Loehr TM, Mauk AG, de Montellano PRO. Replacement of the proximal histidine iron ligand by a cysteine or tyrosine converts heme oxygenase to an oxidase. Biochemistry 1999;38:3733–3743.
96. Reddi AR, Reedy CJ, Mui S, Gibney BR. Thermodynamic investigation into the mechanisms of proton-coupled electron transfer events in heme protein maquettes. Biochemistry 2007;46:291–305.
97. de Lavalette AD, Barucq L, Alric J, Rappaport F, Zito F. Is the redox state of the c_i heme of the cytochrome b_6f complex dependent on the occupation and structure of the Q_i site and vice versa? J Biol Chem 2009;284:20822–20829.
98. Reid LS, Mauk MR, Mauk AG. Role of heme propionate groups in cytochrome-b_5 electron-transfer. J Am Chem Soc 1984;106:2182–2185.
99. Das DK, Medhi OK. The role of heme propionate in controlling the redox potential of heme: square wave voltammetry of protoporphyrinato IX iron(III) in aqueous surfactant micelles. J Inorg Biochem 1998;70:83–90.
100. Chen ZC, Ost TWB, Schelvis JPM. Phe393 mutants of cytochrome P450BM3 with modified heme redox potentials have altered heme vinyl and propionate conformations. Biochemistry 2004;43:1798–1808.
101. Tezcan FA, Winkler JR, Gray HB. Effects of ligation and folding on reduction potentials of heme proteins. J Am Chem Soc 1998;120:13383–13388.
102. Frausto da Silva JJR, Williams RJP. The biological chemistry of the elements—the inorganic chemistry of life, 2nd ed. Oxford: Oxford University Press; 2001.
103. Iverson TM, Arciero DM, Hooper AB, Rees DC. High-resolution structures of the oxidized and reduced states of cytochrome c-554 from *Nitrosomonas europaea*. J Biol Inorg Chem 2001;6:390–397.
104. Polyakov KM, Boyko KM, Tikhonova TV, Slutsky A, Antipov AN, Zvyagilskaya RA, Popov AN, Bourenkov GP, Lamzin VS, Popov VO. High-resolution structural analysis of a novel octaheme cytochrome *c* nitrite reductase from the haloalkaliphilic bacterium *Thioalkalivibrio nitratireducens*. J Mol Biol 2009;389:846–862.
105. Richardson DJ. Bacterial respiration: a flexible process for a changing environment. Microbiology 2000;146:551–571.
106. Mowat CG, Rothery E, Miles CS, McIver L, Doherty MK, Drewette K, Taylor P, Walkinshaw MD, Chapman SK, Reid GA. Octaheme tetrathionate reductase is a respiratory enzyme with novel heme ligation. Nature Struct Mol Biol 2004;11:1023–1024.
107. Einsle O, Stach P, Messerschmidt A, Simon J, Kroger A, Huber R, Kroneck PMH. Cytochrome *c* nitrite reductase from *Wolinella succinogenes*—structure at 1.6 Å resolution, inhibitor binding, and heme-packing motifs. J Biol Chem 2000;275:39608–39616.

108. Lancaster CRD, Kroger A, Auer M, Michel H. Structure of fumarate reductase from *Wolinella succinogenes* at 2.2 Å resolution. Nature 1999;402:377–385.
109. Li L, Mustafi D, Fu Q, Tereshko V, Chen DL, Tice JD, Ismagilov RF. Nanoliter microfluidic hybrid method for simultaneous screening and optimization validated with crystallization of membrane proteins. Proc Natl Acad Sci USA 2006;103:19243–19248.
110. Iverson TM, Arciero DM, Hsu BT, Logan MSP, Hooper AB, Rees DC. Heme packing motifs revealed by the crystal structure of the tetra-heme cytochrome c554 from *Nitrosomonas europaea*. Nature Struc Biol 1998;5:1005–1012.
111. Fisher WR, Taniuchi H, Anfinsen CB. Role of heme in formation of structure of cytochrome-c. J Biol Chem 1973;248:3188–3195.
112. Damaschun G, Damaschun H, Gast K, Gernat C, Zirwer D. Acid denatured apo-cytochrome-c is a random coil—evidence from small-angle X-ray-scattering and dynamic light-scattering. Bioch Biophys Acta 1991;1078:289–295.
113. Ferguson SJ, Stevens JM, Allen JWA, Robertson IB. Cytochrome c assembly: a tale of ever increasing variation and mystery? Biochim Biophys Acta 2008;1777:980–984.
114. Braun M, Thony-Meyer L. Biosynthesis of artificial microperoxidases by exploiting the secretion and cytochrome c maturation apparatuses of *Escherichia coli*. Proc Natl Acad Sci USA 2004;101:12830–12835.
115. Wittung-Stafshede P. Role of cofactors in protein folding. Acc Chem Res 2002;35:201–208.
116. Pfeil W, Nolting BO, Jung C. Apocytochrome-$P450_{cam}$ is a native protein with some intermediate-like properties. Biochemistry 1993;32:8856–8862.
117. Schneider S, Sharp KH, Barker PD, Paoli M. An induced fit conformational change underlies the binding mechanism of the heme transport proteobacteria-protein HemS. J Biol Chem 2006;281:32606–32610.
118. Feng YQ, Sligar SG, Wand AJ. Solution structure of apocytochrome b_{562}. Nature Struct Biol 1994;1:30–35.
119. D'Amelio ,N; Bonvin ,AMJJ; Czisch M, Barker P, Kaptein R. The C terminus of apocytochrome b_{562} undergoes fast motions and slow exchange among ordered conformations resembling the folded state. Biochemistry 2002;41:5505–5514.
120. Falzone CJ, Mayer MR, Whiteman EL, Moore CD, Lecomte JTJ. Design challenges for hemoproteins: The solution structure of apocytochrome b_5. Biochemistry 1996;35:6519–6526.
121. Bhattacharya S, Falzone CJ, Lecomte JTJ. Backbone dynamics of apocytochrome b_5 in its native, partially folded state. Biochemistry 1999;38:2577–2589.
122. Eliezer D, Wright PE. Is apomyoglobin a molten globule? Structural characterization by NMR. J Mol Biol 1996;263:531–538.
123. Landfried DA, Vuletich DA, Pond MP, Lecomte JTJ. Structural and thermodynamic consequences of *b* heme binding for monomeric apoglobins and other apoproteins. Gene 2007;398:12–28.
124. Lad L, Schuller DJ, Shimizu H, Friedman J, Li HY, de Montellano PRO, Poulos TL. Comparison of the heme-free and -bound crystal structures of human heme oxygenase-1. J Biol Chem 2003;278:7834–7843.
125. Wolff N, Izadi-Pruneyre N, Couprie J, Habeck M, Linge J, Rieping W, Wandersman C, Nilges M, Delepierre M, Lecroisey A. Comparative analysis of structural and dynamic properties of the loaded and unloaded hemophore HasA: functional implications. J Mol Biol 2008;376:517–525.
126. Xu T, Wu SP, Miloradovic I, Therien MJ, Blasie JK. Incorporation of designed extended chromophores into amphiphilic 4-helix bundle peptides for nonlinear optical biomolecular materials. Nano Lett 2006;6:2387–2394.
127. Meier M, Janosik M, Kery V, Kraus JP, Burkhard P. Structure of human cystathionine beta-synthase: a unique pyridoxal 5′-phosphate-dependent heme protein. EMBO J 2001;20:3910–3916.
128. Zubieta C, Joseph R, Krishna SS, McMullan D, Kapoor M, Axelrod HL, Miller MD, Abdubek P, Acosta C, Astakhova T, Carlton D, Chiu HJ, Clayton T, Deller MC, Duan L, Elias Y, Elsliger MA, Feuerhelm J, Grzechnik SK, Hale J, Han GW, Jaroszewski L, Jin KK, Klock HE, Knuth MW, Kozbial P, Kumar A, Marciano D, Morse AT, Murphy KD, Nigoghossian E, Okach L, Oommachen S, Reyes R, Rife CL, Schimmel P, Trout CV, van den Bedem H, Weekes D, White A, Xu QP, Hodgson KO, Wooley J, Deacon AM, Godzik A, Lesley SA, Wilson IA. Identification and structural characterization of heme binding in a novel dye-decolorizing peroxidase. Tyr A Proteins: Struc Func Bioinf 2007;69:234–243.
129. Yamanishi M, Kabil O, Sen S, Banerjee R. Structural insights into pathogenic mutations in heme-dependent cystathionine-beta-synthase. J Inorg Biochem 2006;100:1988–1995.
130. Singh S, Madzelan P, Banerjee R. Properties of an unusual heme cofactor in PLP-dependent cystathionine beta-synthase. Nat Prod Reports 2007;24:631–639.
131. Evande R, Ojha S, Banerjee R. Visualization of PLP-bound intermediates in hemeless variants of human cystathionine beta-synthase: evidence that lysine 119 is a general base. Arch Biochem Biophys 2004;427:188–196.
132. Lederer F, Glatigny A, Bethge PH, Bellamy HD, Mathews FS. Improvement of the 2.5 Å resolution model of cytochrome-b_{562} by redetermining the primary structure and using molecular graphics. J Mol Biol 1981;148:427–448.
133. Willies SC, Isupov MN, Garman EF, Littlechild JA. The binding of haem and zinc in the 1.9 A X-ray structure of *Escherichia coli* bacterioferritin. J Biol Inorg Chem 2009;14:201–207.
134. Martinez SE, Huang D, Ponomarev M, Cramer WA, Smith JL. The heme redox center of chloroplast cytochrome *f* is linked to a buried five-water chain. Protein Sci 1996;5:1081–1092.
135. Vojtechovsky J, Chu K, Berendzen J, Sweet RM, Schlichting I. Crystal structures of myoglobin-ligand complexes at near-atomic resolution. Biophys J 1999;77:2153–2174.
136. Ayers RA, Moffat K. Changes in quaternary structure in the signaling mechanisms of PAS domains. Biochemistry 2008;47:12078–12086.
137. Pellicena P, Karow DS, Boon EM, Marletta MA, Kuriyan J. Crystal structure of an oxygen-binding heme domain related to soluble guanylate cyclases. Proc Natl Acad Sci USA 2004;101:12854–12859.
138. Weichsel A, Maes EM, Andersen JF, Valenzuela JG, Shokhireva TK, Walker FA, Montfort WR. Heme-assisted S-nitrosation of a proximal thiolate in a nitric oxide transport protein. Proc Natl Acad Sci USA 2005;102:594–599.
139. Poulos TL, Finzel BC, Howard AJ. Crystal-structure of substrate-free *Pseudomonas putida* cytochrome-P-450. Biochemistry 1986;25:5314–5322.
140. Sugimoto H, Oda SI, Otsuki T, Hino T, Yoshida T, Shiro Y. Crystal structure of human indoleamine 2,3-dioxygenase: catalytic mechanism of O-2 incorporation by a heme-containing dioxygenase. Proc Natl Acad Sci USA 2006;103:2611–2616.
141. Arnoux P, Haser R, Izadi N, Lecroisey A, Delepierre M, Wandersman C, Czjzek M. The crystal structure of HasA, a hemophore secreted by *Serratia marcescens*. Nature Struct Biol 1999;6:516–520.
142. Huang LS, Cobessi D, Tung EY, Berry EA. Binding of the respiratory chain inhibitor antimycin to the mitochondrial bc(1) complex: a new crystal structure reveals an altered intramolecular hydrogen-bonding pattern. J Mol Biol 2005;351:573–597.
143. Fritzsch G, Buchanan S, Michel H. Assignment of cytochrome hemes in crystallized reaction centers from *Rhodopseudomonas viridis*. Biochim et Biophys Acta 1989;977:157–162.