Diversity and conservation of interactions for binding heme in b-type heme proteins

Diversity and conservation of interactions for binding heme in b-type heme proteins†‡

Sabine Schneider, Jon Marles-Wright,§ Katherine H. Sharp and Max Paoli*

Received (in Cambridge, UK) 17th January 2007
First published as an Advance Article on the web 28th March 2007
DOI: 10.1039/b604186h

Covering: up to 2006

The heme prosthetic group is vital to many cellular processes and is therefore widespread throughout organisms of different phylogenetic origin. Heme is used in proteins involved in cellular respiration, acts as a chemical mediator in ligand binding and signalling proteins, and is the key co-factor in many enzymes. Strikingly, there are over 20 different folding topologies of b-type heme proteins that are able to incorporate the same, chemically identical heme ligand. Comparisons of structures show that heme–protein interactions are generally diverse, though a degree of conservation exists at contacts with the pyrrole rings, the propionate groups and the proximal ligand. These interaction "hot spots" presumably define major determinants for binding heme and provide guidelines for the future design of novel heme proteins.

1 Introduction

Heme (iron-protoporphyrin IX) is an extremely versatile prosthetic group widespread in biological systems and vital to aerobic life. It is an essential co-factor for oxygen binding and transport functions carried out by the globins, and the electron transfer and redox reactions of respiration and photosynthesis. Heme groups are also found in proteins involved in catalysis, such as the catalases and mono-oxygenases, as well as in proteins that carry out a great variety of processes, including signal transduction and the control of gene expression.[1,2]

The structure of the heme co-factor can be modified by derivatization of the pyrroles' methyl and vinyl side chains which extend from the edge of the porphyrin. In b-type heme (Fig. 1), iron-protoporphyrin IX is non-covalently bound to the protein; this is the prosthetic group of the familiar structures of hemoglobin, myoglobin and cytochrome b5. Most other cytochromes, however, contain c-type heme which differs from a b-heme in the vinyl side chains that are covalently linked to two cysteine residues by thioether bonds. Other, less common heme derivatives include the a-type heme, present in cytochrome c oxidase, and the d_1-type heme identified in nitrite reductase.[3] A summary of heme derivatives and associated heme proteins is reported in the Promise database.[4]

In the past decade, the explosive growth in structural biology led to the elucidation of the molecular architecture of a striking number a novel b-type heme proteins, such as the insect protein nitrophorin,[5] the enzyme NO synthase,[6] the bacterial siderophore HasA,[7] the mammalian transport protein hemopexin,[8] the binding protein albumin[9] and the molecular sensors CooA[10] and EcDos.[11] These and other heme-proteins show a spectacular range of distinct folding topologies all associated with the same, chemically identical heme ligand; b-type heme is found in over 20 different folds as highlighted in Table 1 and depicted in Fig. 2. Evidently, the evolutionary pressure of generating various biochemical functions, combined with the reactivity and versatile chemical properties of heme, have led to the rise of very different structural associations between heme and proteins. Even in the case of the evolution of a particular function, multiple paths can be followed that result in proteins with distinct folds, structures and mechanisms. This is true for the mammalian blood serum glycoprotein hemopexin

School of Pharmacy and Centre for Biomolecular Sciences, University of Nottingham, University Park, Nottingham, NG7 2RD, England, UK. E-mail: max.paoli@nottingham.ac.uk; Fax: +44 (115) 8468002; Tel: +44 (115) 8468015

† This paper was published as a part of a special issue on the chemistry and biochemistry of heme proteins.

‡ Electronic supplementary information (ESI) available: Supplementary Table 1: Solvent accessibility and protein–propionate contacts, and Table 2: Protein–heme-face and protein–heme-edge contacts. See DOI: 10.1039/b604186h

§ Present address: Institute for Cell and Molecular Biosciences, Medical School, Newcastle University, Framlington Place, Newcastle upon Tyne, NE2 4HH, UK.

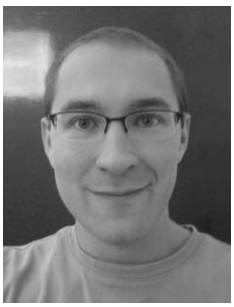
Sabine Schneider received her Diploma in Biology, with an emphasis on molecular biology, from the Ludwig-Maximillians-University in Munich in 2002. Currently, she is pursuing her PhD under the supervision of Dr Max Paoli at the University of Nottingham and is working on the structure–function relationships of bacterial heme-transport proteins.

Jon Marles-Wright received his BSc in Biochemistry from the University of Cambridge and then obtained a PhD at the University of Oxford where he worked in the field of structural biology. He is currently doing post-doctoral research at the University of Newcastle working on stress response in Bacillus subtilis.

Katie Sharp earned a Master of Chemistry with Drug Design and Toxicology from the University of Hull in 2001. She gained a PhD in Chemistry at the University of Leicester in 2005 working on the molecular evolution of heme proteins for enhanced biocatalysis. Her current research work as a research associate at the University of Nottingham involves structural studies of bacterial heme proteins.

Max Paoli studied at the University of York for both his BSc in Biochemistry and PhD in Chemistry. Since his post-graduate work with Professor Guy Dodson, he has worked on several heme proteins including hemoglobin, hemopexin, FixL, and HemS. His current research programme at the University of Nottingham focuses on the structural biology of protein–ligand and protein–protein complexes.

Sabine Schneider



Jon Marles-Wright

Katherine H. Sharp

Max Paoli

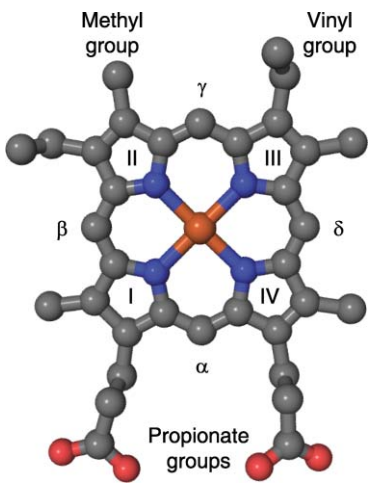


Fig. 1 Ball-and-stick model of b-type heme. Heme (protopophyrin IX) consists of four pyrrole rings (I–IV) linked by methyl bridges (α, β, γ, δ) forming a so-called tetrapyrrole ring. Pyrroles I and IV carry the two propionate groups which can engage with their carboxy termini in electrostatic interactions with the protein environment. In the centre of the porphyrin plane a ferric ($^{3+}$) or ferrous ($^{2+}$) iron (shown in orange) is coordinated by the four pyrrole nitrogens (shown in blue). The apparent symmetry of the heme breaks down because of the four methyl and two vinyl groups which line the heme edge protruding from the plane of each of the pyrrole rings.

and the bacterial siderophore HasA which have broadly equivalent roles of sequestration, transport and receptor-mediated release of heme but are structurally unrelated.

Despite the remarkably different architecture of heme proteins it is tempting to speculate whether any interactions are shared given that they are all associated with heme. Comparisons of the wide range of heme proteins known to date may throw light on the structural requirements of a heme binding site.

2 What makes a heme binding site?

Historically, it has been presumed that the bond(s) between the heme iron and the amino acid(s) coordinating the iron is/are the major force holding the heme into the protein. However, experiments on globin and cytochrome mutants in which the proximal histidine was changed into a glycine and the side chain replaced by an imidazole,[12,13] showed that the protein could still incorporate heme even without a coordinate/covalent bond attachment. This therefore suggests that the protein framework provides sufficient interactions for binding within a relatively heme-specific binding pocket. The importance of these non-covalent interactions is highlighted by the work on many "designer" heme proteins such as molecular maquettes[14] and others.[15,16] In the engineering of these proteins the focus was on the positioning of histidine residues, resulting in heme binding at the expense of structural stability. Several other design experiments have yielded molecules with low heme affinity, in addition to their molten globule-like characteristics and lacking tightly packed interiors. It was previously recognised that heme-contacting residues other than the histidine ligands are important for heme binding[17] thus one can ask if any common, key structural features or re-occurring heme–protein packing interactions might exist in distinct folding environments.

Table 1 Selected representatives of b-type heme proteins with distinct folding topologies. In addition to their different folds, the table highlights the diversity of their functions and phylogenetic origins as well as the ligation and coordination state of the heme-iron

Fold	Representative proteins	Function	Origin	Ligand	PDB code
4-Helix bundle	Cytochrome b_{562}	Electron transport	Bacteria (*Escherichia coli*)	His/Met 6c	1QPU
β-Propeller	Hemopexin	Heme binding and transport	Rabbit (*Oryctolagus cuniculus*)	bis-His 6c	1QHU
CAP	CooA	CO sensing	Bacteria (*Rhodospirillum rubrum*)	His/Pro 6c	1FT9
Catalase	Catalase	Hydrogen peroxide decomposition	Bacteria (*Helicobacter pylori*)	Tyr 5c	1QWL
Cupredoxin-like	Ubiquinol oxidase	O_2 reduction to water	Bacteria (*Escherichia coli*)	His 5c	1FFT
Cytochrome b5	Cytochrome b5	Electron transfer	Rat (*Rattus norvegicus*)	bis-His 6c	1EUE
Dioxygenase	Indoleamine 2,3-dioxygenase	L-Tryptophan catabolism	Man (*Homo sapiens*)	His 5c	2D0T
Globin-like	Hemoglobin	O_2 binding and transport	Man (*Homo sapiens*)	His 5c	1A3N
Heme-oxygenase	Heme oxygenase	Heme degradation	Man (*Homo sapiens*)	His/H_2O 6c	1N45
Heme-peroxidase	Cyt c peroxidase	Biosynthetic and catabolism	Yeast (*Saccheromyces cerevisia*)	His 5c	2CYP
HemS	HemS	Heme transport	Bacteria (*Yersinia enterocolitica*)	His 5c	2J0P
H-NOX/SONO	SONO	NO sensing	Bacteria (*Thermoanaerobacter tengcongensis*)	His 5c	1XBN
Immunoglobulin-like	Cellobiose dehydrogenase	Lignin and cellulose degradation	Fungus (*Phanerochaete chrysosporium*)	His/Met 6c	1D7B
Lipocalin	Nitrophorin	NO transfer	Insect (*Rhodnius prolixus*)	His 5c	1NP4
Meander	Has A	Heme binding and transport	Bacteria (*Serratia marcescens*)	His/Tyr 6c	1B2V
NO	NO synthase	Catalytic	Mouse (*Mus musculus*)	Cys 5c	1NOS
P450	P450 mono-oxygenase	Oxidation of organic substrates	Fungus (*Streptomyces coelicolor*)	Cys 5c	1ODO
PAS	FixL	O_2-sensor/signalling	Bacteria (*Bradyrhizobium japonicum*)	His 5c	1DRM
Serum albumin-like	Albumin	Regulation of the colloidal osmotic pressure of blood	Man (*Homo sapiens*)	Tyr 5c	1N5U
Vitamin B6 family	Cystathione β-synthase	Redox-controlled PLP-dependent synthesis of cystathione	Man (*Homo sapiens*)	Cys/His 6c	1JBQ

The b-type heme proteins listed in Table 1 (and Supplementary Table 1 and 2‡) are functionally dissimilar. Any shared interactions should therefore play a primarily structural role in the association with the heme. We report a comparative analysis of heme-proteins from structurally unrelated families by means of multiple structural overlays. The study shows that, while a wide variety of folds can bind heme and differences in heme environments obviously exist as dictated by the different functions, a specific molecular pattern emerges for some contact points between heme and the protein binding pocket.

3 Comparative analysis of heme environments in distinct structures

Structures of 68 b-type heme proteins with less than 60% sequence identity and representative of over 20 different folds (see Supplementary Table 1) have been compared by least-squares superposition of the heme atoms. In some cases where significant sequence divergence exists between two heme proteins despite their equivalent folding topologies, multiple representatives of the same fold were analysed (see Supplementary Table 1). For instance, molecules such as sulfite oxidase (for the PDB code of the exact coordinate file referred to here and for proteins mentioned anywhere else in the analyses reported in this review, please see Supplementary Tables 1 and 2) were included because its cytochrome b5 domain is structurally equivalent to the rat cytochrome b5, despite the lack of significant sequence identity.[18] Another example is given by the structure of soluble guanylate cyclase which is not included in Table 1 because it has the same fold as SONO (sensor of nitric oxide) but was considered in our analysis given the lack of significant sequence similarity.[19,20]

Atomic coordinates from the Protein Data Bank[21] were superimposed using the atoms of the porphyrin ring as a reference frame. Firstly a rotation/translation matrix was calculated from the best least-squares fit of the porphyrin atoms with the addition of, in some cases, the proximal histidine imidazole. Secondly the matrix was applied to the whole protein. The computation of the overlays was carried out using the fitting procedures implemented in *DeepView*[22] and visual inspections were carried out using both *DeepView* and the program *O*.[23] The analysis was guided by a list of all amino acids within 4.5 Å of any of the heme atoms, produced with the program *DISTANG*.[24] In this way the folding

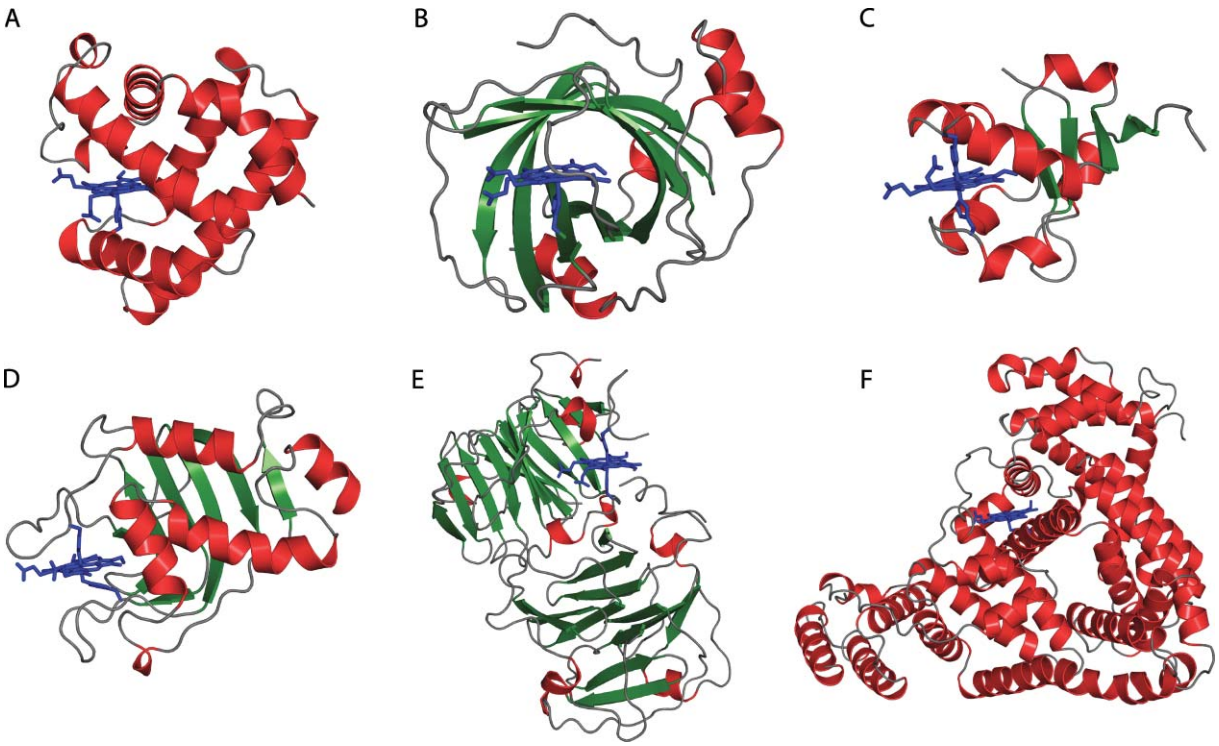


Fig. 2 Schematic diagrams showing the structures of six representative b-type heme proteins highlighting the striking variety of their folding topologies; the heme environment can be formed by any element of secondary structure or any combination of these, as well as turns and loop regions. (**A**) Tuna myoglobin (1MYT). (**B**) Insect nitrophorin (1NP4). (**C**) Cow cytochrome b5 (1CYO). (**D**) Bacterial HasA (1DK0). (**E**) Rabbit hemopexin (1QHU). (**F**) Human serum-albumin (1N5U).

Table 2 Average solvent accessible area for the heme in different protein folds. The solvent accessible area was calculated using the program *AREAIMOL*.[24] The table reports values for the exposure of the heme calculated as a percentage exposure relative to free heme. Where two or more distant representatives of a given fold are analysed, an average value is reported; only in the case of the PAS domain the heme turned out to have a large difference in solvent accessibility between the protein members of this group, EcDos (20%) and FixL (5%). For a detailed list see Supplementary Table 1

		Average heme solvent accessibility	
Protein	No. of structures	Å^2	%
Free heme	—	829	100
4-Helix bundle	3	166	20
β-Propeller	1	210	25
CAP	1	233	28
Catalases	8	11	1
Cupredoxin-like	1	32	4
Cytochrome b5	4	184	22
Dioxygenase	1	92	11
Globin like	12	135	16
Heme-oxygenase	6	160	19
Heme-peroxidase	6	72	8
HemS	1	155	19
H-NOX/SONO	2	19	2
Immunglobulin-like	1	179	22
Lipocalin	1	130	16
Meander	1	191	23
NO	4	128	15
P450	14	22	3
PAS	2	111	12
Serum albumin-like	1	66	8
Vitamin b6 family	1	148	18

topologies and structural environments around the heme group were compared. All figures were prepared using *PyMol* (DeLano Scientific LLC).

4 Diversity of heme–protein packing contacts

The comparative analysis mentioned above highlights differences and similarities between various structurally unrelated and diverse heme-proteins. Evolutionary constraints imposed by the specific function of each protein presumably dictate many differences in the heme environments, the structure of which has been "sculpted" to modulate the reactivity of the heme. The range of strikingly different folding topologies is associated with an entire spectrum of heme arrangements, as suggested by the set of structures shown in Fig. 2. For example, the heme can either be buried inside the molecule, like in catalases with an average solvent accessibility of 1.4%, or is bound in a pocket near the surface, often relatively exposed to solvent, like in hemopexin with an average solvent accessibility of 25% (see Table 2 and Supplementary Table 1). Its orientation is not fixed, even in proteins with equivalent biological functions; the propionate groups can point either towards the outside (*e.g.* HasA) or the interior (*e.g.* HemS) of the molecule. In the repertoire of proteins surveyed, the heme environment is formed by either α-helical or β-extended structures, or both, as well as loop regions and excursions from secondary elements. Although the packing of amino acid residues around the heme is generally tight, there are examples of unusual and relatively open heme pockets, such as those found in hemopexin, cellobiose dehydrogenase, cystathione β-synthase and CooA.

5 Re-occurring contacts: protein–heme interaction "hot spots"

The structural superpositions also reveal that some key contacts are shared between distinct proteins, as discussed in more detail below and illustrated in Fig. 3 (see also Tables 2–5). In spite of the great diversity of folds and the general diversity of binding pockets imposed by the different functions, visual inspection of the overlays revealed a definite preference for specific amino acids in the interactions with heme. These contacts were categorised in three sets depending on whether they interact with i) the propionate groups, ii) the plane of the pyrrole rings or heme face, and iii) the heme edge defined by the perimeter of the porphyrin including the atoms of the methyl and vinyl groups.

6 Anchoring of heme propionates by arginine residues

In most of the 20 folds surveyed, at least one of the propionate groups is involved in electrostatic interactions, principally salt-bridges where 38% are with an arginine side chain. This is not only seen in proteins sharing some sequence similarity such as cytochromes $P450_{cam}$, $P450_{terp}$ and $P450_{BM-3}$,[25] but also in functionally unrelated molecules. Whilst the heme is more often oriented with the propionates pointing towards the outside of the protein molecule or located at its surface, the propionate groups are partially excluded from the solvent by their interactions with amino acids such as arginines. In those cases in which the heme is buried (*e.g.* cytochrome P450 and catalase) or when it directs the acidic groups towards the protein interior (*e.g.* hemopexin), arginine residues are still found to engage in interactions with propionates. Regardless of the orientation of the heme, arginine residues appear to be universal partners for the propionate groups, providing them with anchoring points to the protein (see Table 3 and Supplementary Table 1). Being positively charged, lysine and histidine residues are also used to interact with the negatively charged propionates, as seen for instance in the globin-fold and P450 peroxidases, but this is less common with only 10% and 17%, respectively, of all observed contacts. Surprisingly, tyrosine residues are involved in electrostatic interactions with the propionate groups almost as often as lysines (see Table 3 and Supplementary Table 1). In some cases, as in the peroxidase domain of prostaglandin H2 synthase-1, nitrous oxide synthase and heme oxygenase (Table 1), the propionates are in contact with the backbone amides of loop regions that embrace part of the prosthetic group.

7 Conserved interactions at the "heme face" involving leucine

The majority of heme-face contacts involves leucine and isoleucine (11%). Valine is also common in heme face interactions (11%), and, like leucine and isoleucine, may be found on both proximal and distal sides, making non-polar contacts to yield a close fit between heme and protein atoms (see Table 4). The most common heme-face contact, seen in at least twelve distinct folding topologies, involves leucine and isoleucine residues located approximately over the bridge between pyrrole rings II and III, as depicted by the cluster of side chains shown in Fig. 3a. It is striking that twelve diverse, unrelated structures feature side chains that occupy the same three-dimensional space in relation to the heme.

Table 3 Frequencies of residues in contact with the heme propionate groups in b-type heme proteins with distinct folding topologies. Distances were calculated using *DISTANG*[24] and residues with a hydrogen bonding distance of 2.5–3.2 Å[41] to the propionate groups were taken into account. Percentages are relative to the number of contacts in a particular fold. A detailed list can be found in Supplementary Table 1

	Propionate contacts (2.5–3.2 Å) in %				
Fold	Arg	His	Lys	Tyr	Other
4-Helix bundle	100.0	—	—	—	—
β-Propeller	33.3	16.7	—	50.0	—
CAP	—	—	—	—	100.0
Catalases	82.9	8.6	—	—	8.6
Cupredoxin-like	75.0	—	—	—	25.0
Cytochrome b5	—	10.0	20.0	10.0	60.0
Dioxygenase	75.0	—	—	—	25.0
Globin like	2.9	17.1	34.3	11.4	34.3
Heme-oxygenase	37.5	—	31.3	25.0	6.3
Heme-peroxidase	10.3	24.1	10.3	—	55.2
HemS	57.1	—	14.3	14.3	14.3
H-NOX/SONO	70.0	—	—	10.0	20.0
Immunglobulin-like	—	—	—	40.0	60.0
Lipocalin	—	—	66.7	—	33.3
Meander	—	16.7	—	16.7	66.7
NO fold	20.0	—	—	80.0	—
P450	47.4	28.2	2.6	3.8	17.9
PAS	28.6	57.1	—	—	14.3
Serum albumin-like	—	50.0	50.0	—	—
Vitamin b6 family	100.0	—	—	—	—
Overall %	**38.4**	**17.2**	**10.4**	**9.0**	**25.0**

In other proteins, such as the globins, the side chains of leucine or isoleucine are seen at different positions but still contributing similar hydrophobic contacts onto the heme face. Site-directed mutagenesis to alanine of the leucine in contact with the heme in the globin fold results in an increased rate of hemin dissociation.[26] This has mainly been attributed to hydration of the environment around the proximal ligand, but the simple removal of favourable van der Waals interactions between leucine and heme is likely to contribute to the stability of the heme–protein association.

8 Conserved interactions at the "heme face" involving phenylalanine/tyrosine side chains

Phenylalanine, tyrosine and tryptophan are responsible for a great deal of the other contacts to the heme face (15%, 4% and 4%, respectively; see Table 4 and Supplementary Table 2), engaging predominantly in stacking interactions with the porphyrin. Most of these phenylalanine–porphyrin contacts have distances in the range of 3.6–5.0 Å and have an offset face-to-face parallel geometry relative to one of the four pyrrole rings. Similar observations have been made with regard to the pairing of aromatic side chains in proteins.[27,28] Some of the phenylalanine–porphyrin contacts clearly involve aromatic–aromatic π-stacking interactions and are likely to provide significant stabilisation for heme binding. In addition, surface representations, examination of van der Waals radii and shapes of side chains within the pocket show how these contacts often result in excellent steric complementarity between heme and protein. Strikingly, sets of phenylalanine residues from different structures cluster at specific sites close to the porphyrin surface. Figs. 3b and 3d highlight structurally conserved heme–phenylalanine contacts. Phenylalanine residues are located over the bridge between pyrrole rings III and IV in the case of

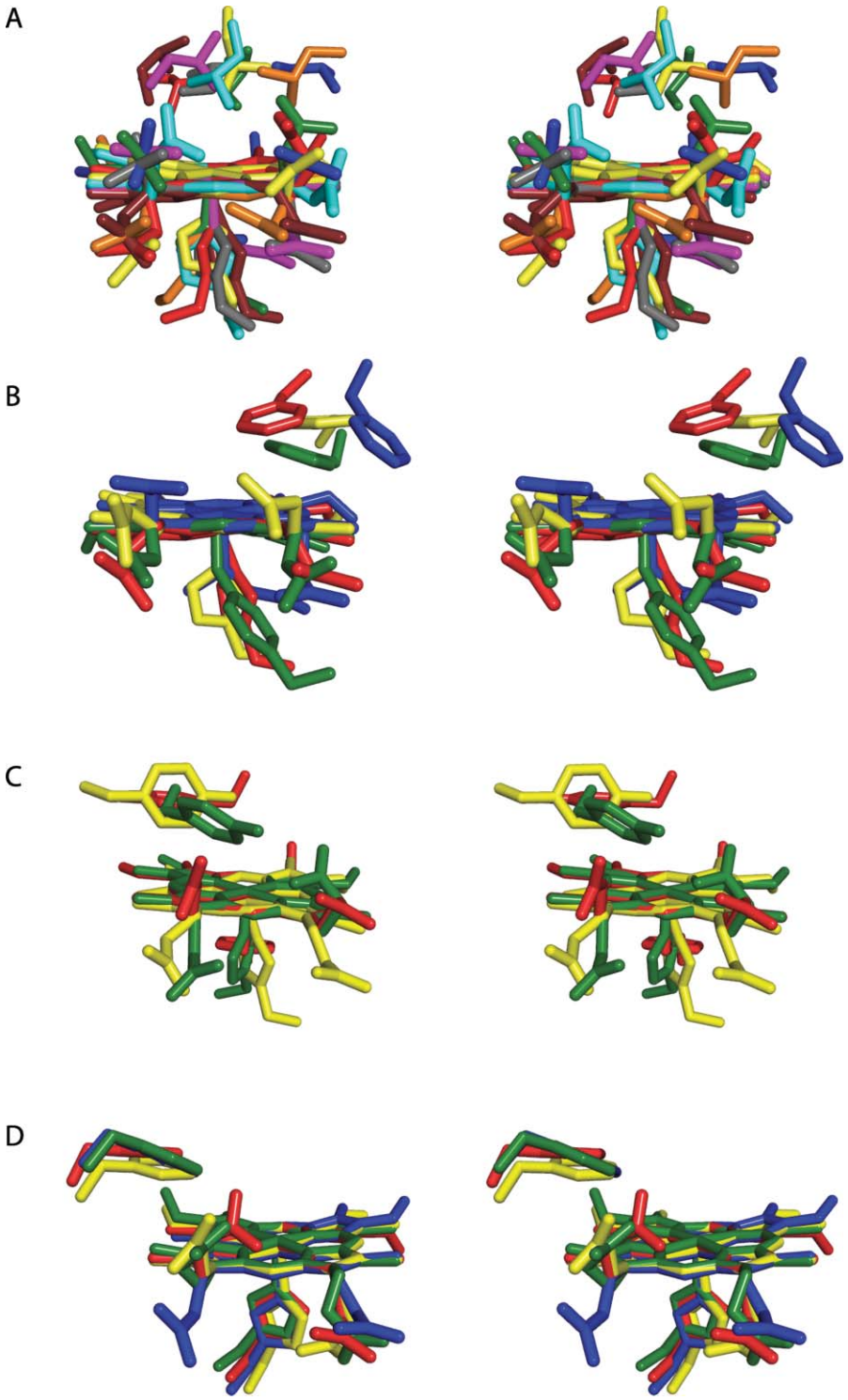


Fig. 3 Panel highlighting structurally conserved packing interactions to the "heme-face" involving leucine/isoleucine and phenylalanine/tyrosine side chains. (**A**) Clustering of leucine/isoleucine residues over the heme occurs in eleven distinct proteins; for clarity only nine of these structures are shown in the figure: CooA (1FT9) Leu 112 in red, bacterioferritin (1BFR) Ile 22 in purple, cytochrome b5 (1AWP) Ile 25 in green, nitrophorin (1NP4) Leu 132 in yellow, ubiquinol oxidase (1FFT) Ile 425 in grey, FixL (1DRM) Leu 236 in cyan, serum-albumin (1N5U) Leu 139 in blue, heme oxygenase (1N45) Leu 147 in orange, SONO (1XBN) Ile 75 in dark red [not shown: Ile 57 HasA (1DK0), Ile 75 of guanylate cyclase (1U55), Leu 92 and Leu 94 of HemS (2J0P)]. (**B**) Conserved phenylalanine residues over the bridge between pyrrole rings III and IV in distinct heme proteins: SONO (1XBN) Phe 78 in red, cytochrome b_{562} (1QPN) Phe 65 in yellow, catalase (8CAT) Phe 160 in green and bacterioferritin (1BFR) Phe 26 in blue. (**C**) Conserved tyrosine residues packing over pyrrole I: human serum-albumin (1N5U) Tyr 138 in red, cellobiose dehydrogenase (1D7B) Tyr 90 in green, bacterial SONO (1XBN) Tyr 140 in yellow. (**D**) Conserved phenylalanine side chains packing in proximity of pyrrole I in the globin fold: rice hemoglobin (1D8U) Phe 54 in green, tuna myoglobin (1MYT) Phe 43 in yellow, mouse neuroglobin (1Q1F) Phe 42 in red, *Escherichia coli* flavohemoglobin (1GVH) Phe 43 in blue. Least-square superposition of the heme-atoms was carried out in *DeepView*.[22]

Table 4 Frequencies of residues in van der Waals contact (3.6–4.1 Å[41]) with the heme-face in different folding topologies of b-type heme proteins. Distances were calculated using *DISTANG*[24] and percentages are relative to the number of contacts in a particular fold. A more detailed list can be found in Supplementary Table 2

	Heme-face contacts (3.6–4.1 Å) in %							
Fold	Ala	Ile	Leu	Phe	Trp	Tyr	Val	Other
4-Helix bundle	—	30.0	20.0	10.0	10.0	—	—	30.0
β-Propeller	—	—	—	—	33.3	33.3	—	33.3
CAP	—	14.3	28.6	—	—	—	—	57.1
Catalases	—	—	—	20.9	—	—	32.6	46.5
Cupredoxin-like	—	22.2	—	33.3	—	—	11.1	33.3
Cytochrome b5	7.9	—	10.5	15.8	5.3	5.3	7.9	47.4
Dioxygenase	16.7	—	33.3	33.3	—	—	16.7	—
Globin like	2.4	9.8	20.7	22.0	—	6.1	28.0	11.0
Heme-oxygenase	—	—	4.3	17.4	—	—	4.3	73.9
Heme-peroxidase	2.5	—	12.5	—	15.0	—	7.5	62.5
HemS	—	—	—	66.7	—	—	33.3	—
H-NOX/SONO	—	14.3	35.7	14.3	—	14.3	—	21.4
Immunglobulin-like	25.0	—	—	—	25.0	25.0	—	25.0
Lipocalin	16.7	—	50.0	16.7	—	16.7	—	—
Meander	—	—	20.0	—	—	20.0	20.0	40.0
NO	8.3	4.2	—	16.7	16.7	4.2	8.3	41.7
P450	21.2	3.0	8.1	15.2	—	—	6.1	46.5
PAS	—	38.9	22.2	5.6	—	11.1	11.1	11.1
Serum albumin-like	—	20.0	—	20.0	—	20.0	—	40.0
Vitamin b6 family	14.3	—	14.3	—	14.3	—	—	57.1
Overall %	**7.3**	**6.2**	**12.2**	**15.4**	**3.8**	**4.0**	**12.9**	**38.1**

Table 5 Frequencies of residues in van der Waals contact (3.6–4.1 Å[41]) with the heme-edge in different folding topologies of b-type heme proteins. Distances were calculated using *DISTANG*[24] and percentages are relative to the number of contacts in a particular fold. A more detailed list can be found in Supplementary Table 2

	Heme-edge contacts (3.6–4.1 Å) in %							
Fold	Ala	Ile	Leu	Phe	Trp	Tyr	Val	Other
4-Helix bundle	7.1	14.3	28.6	—	7.1	7.1	—	35.7
β-Propeller	—	—	—	33.3	—	—	—	66.7
CAP	—	11.1	22.2	22.2	11.1	—	11.1	22.2
Catalases	14.9	1.4	10.8	9.5	—	1.4	5.4	56.8
Cupredoxin-like	—	50.0	—	25.0	—	—	—	25.0
Cytochrome b5	14.0	6.0	32.0	10.0	6.0	4.0	8.0	20.0
Dioxygenase	—	28.6	—	28.6	—	14.3	14.3	14.3
Globin like	12.5	7.5	23.8	7.5	—	12.5	10.0	26.3
Heme-oxygenase	6.5	2.2	8.7	19.6	—	10.9	13.0	39.1
Heme-peroxidase	1.9	13.0	20.4	20.4	7.4	3.7	9.3	24.1
HemS	33.3	—	33.3	33.3	—	—	—	—
H-NOX/SONO	—	—	15.4	38.5	—	30.8	—	15.4
Lipocalin	—	16.7	16.7	16.7	—	33.3	16.7	—
Meander	12.5	12.5	12.5	25.0	—	12.5	—	25.0
NO	3.1	3.1	12.5	9.4	12.5	6.3	6.3	46.9
P450	8.4	9.3	27.1	15.9	—	0.9	4.7	33.6
PAS	10.0	20.0	20.0	10.0	10.0	10.0	10.0	10.0
Serum albumin-like	9.1	—	27.3	27.3	—	—	9.1	27.3
Vitamin b6 family	—	—	—	—	—	20.0	—	80.0
Overall %	**8.8**	**7.6**	**19.9**	**14.4**	**2.6**	**6.3**	**7.4**	**33.1**

cytochrome b_{562}, catalase, bacterioferritin and SONO (Fig. 3b), all proteins with distinct topologies and unrelated functions. Fig. 3c shows another set of structurally conserved contacts where three tyrosines pack onto pyrrole I. A similar interaction occurs in the globin fold and in the two-over-two helical fold (Fig. 3d); in these cases, interestingly, equivalent phenylalanine contacts are made to different pyrrole rings. Fig. 3d shows the structures after superposition on both the heme and proximal histidine. The hemes are essentially flipped 180°, so that pyrrole I of truncated hemoglobin is overlaid onto pyrrole IV of myoglobin (see later section about heme flipping). Regardless of the heme orientation, these structures clearly exhibit the same aromatic interaction.

9 Contacts at the heme edge

More examples of steric complementarity are provided by amino acids that pack at the heme edge, though in this case there is limited conservation in the position of interacting groups across various folding environments (Table 5 and Supplementary Table 2). There is a striking preference for aromatic residues such

as phenylalanines which are frequently used in van der Waals contacts with the edge of the porphyrin and are found in a variety of different positions on the non-polar sides of the heme. Leucine side chains are re-occurring, though isoleucine is rarely observed possibly because its numerous rotamer conformations make it too flexible to keep the heme in a fixed position (Table 5 and Supplementary Table 2). Leucines often pack at the edge of the heme between pyrroles II and III. Other residues packing at these sites include alanine and valine (Table 5 and Supplementary Table 2). All these amino acids contribute to building relatively rigid and complementary surfaces for binding.

10 Interactions with the proximal ligand

It appears that residues next to the proximal histidine may in some proteins contribute to the association with heme. In both hemopexin and cytochrome c peroxidase, a tryptophan packs face-to-face onto the histidine imidazole, also making contact with the porphyrin, as shown in Fig. 4a. Surveys of aromatic interactions in proteins previously revealed that tryptophan and histidine side chains preferentially interact with a horizontally displaced stacked geometry.[28] The tryptophan-proximal imidazole packing provides a favourable interaction which may help to position the histidine side chain and hence contribute to the limited stability of the histidine–iron bond. In some cases another amino acid residue is present also on the other side of the proximal ligand, making close contacts (see Fig. 4) providing further packing onto the proximal imidazole and thus reinforcing the idea of its positioning with respect to the heme iron. Spectroscopic evidence from work on the globins indicates that this bond is relatively weak with an estimated energy of about 10 kcal mol^{-1}.[29]

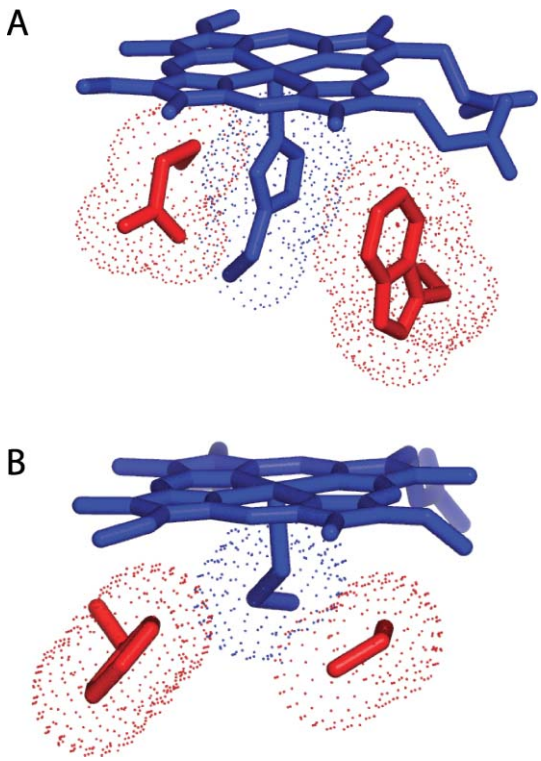


Fig. 4 Packing interactions with the proximal ligand. (**A**) In hemopexin (1QHU) His 265 is sandwiched between Glu 225 (left) and Trp 267 (right). (**B**) In cytochrome P450 (1PQ2) Cys 435 is sandwiched between Phe 428 (left) and Gly 437 (right).

In both lignin peroxidase and nitrophorin, a phenylalanine packs against the imidazole of the proximal histidine suggesting that π-stacking is also taking place in this case. Leucine and valine residues are found at this position in, respectively, fungal peroxidase and in the peroxidase domain of prostaglandin H_2 synthase-1. More extensive stacking interactions exist in both catalase and ubiquinol oxidase, where the proximal histidine is sandwiched between two phenylalanine side chains. Finally, in the case of cytochrome P450 the cysteine heme ligand packs between a phenylalanine and a glycine, as shown in Fig. 4b. The presence of these side chains presumably still provides a stabilising packing interaction with the proximal ligand and the heme.

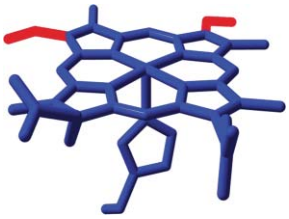
11 Heme orientation/"flipping" relative to the proximal ligand

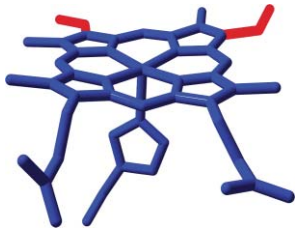
When structures are overlaid using a frame of reference which includes the proximal ligand in addition to the central atoms of the porphyrin, the heme is effectively rotated 180° in some proteins relative to others. So, interestingly, it appears that the heme can be bound in either of two flipped orientations defined by the asymmetry in the porphyrin due to the positions of the vinyl substituents. The graphics in Table 6 help to visualise these two distinct arrangements. Whilst here we do not try to make a comprehensive survey of the heme orientations in all known protein structures, some examples are highlighted below. These "flipped arrangements" were first observed in NMR studies on rat cytochrome b_5.[30–32] Interestingly, the 1.5 Å resolution structure of neuroglobin revealed a mixed population of heme groups bound in both the conformations shown in Table 6.[33] Some protein structures or folds appear to select, through the specific packing of residues around the heme, a particular heme orientation, such as for instance in human hemoglobin and in the hemoglobin from the ciliate *Paramecium caudatum* (reported in bold typescript in Table 6). It is therefore apparent that the orientation of the heme, in terms of the positions of the vinyl groups relative to the proximal ligand, does not affect function, as long as the protein scaffold has evolved to associate with heme in a given orientation.[30–33] Indeed, examination of the protein contacts at the heme edge revealed limited conservation of interactions across different folds which reflects the fact that the different packing requirements of binding heme in either of its distinct orientations.

12 Perspectives for the design of novel heme proteins

The observations on common heme-binding motifs reported above provide guidelines for the design of novel heme proteins. In many of the proteins examined here, aromatic interactions clearly play a key role in stabilising the heme–protein association. In some cases they are absent (CooA), indicating that aromatic contacts are not essential. Analysis of van der Waals surfaces strongly suggests that steric complementarity could be more important than aromatic stacking. It may not be possible however to estimate the relative contributions of the different interactions, especially in those cases in which the evolutionary functional pressures dominate the structure of the heme pocket. For instance, in hemopexin and HasA the heme is a true ligand, whereas in the

Table 6 "Flipped" heme orientations relative to the proximal ligand. Structural comparisons of heme pockets show that either of two possible "flipped" heme orientations with respect to the proximal ligand is observed in heme–protein associations. Interestingly, as highlighted in the top panel of the table, different orientations can exist in homologous or functionally related proteins from different species. The nomenclature "P" and "E" refers to proteins with prokaryotic and eukaryotic origins, respectively





Vinyl left + back			Vinyl back + right		
Fold	PDB code	Origin/ligand	Fold	PDB code	Origin/ligand
cyt b5 fold			**cyt b5 fold**		
cyt b558, *E. vacuolata*	1CXY	P/bis-His	cyt b5, *R. norvegicus*	1EUE	E/bis-His
HO fold			**HO fold**		
Hemoglobin, *H. Sapiens*	IN45	E/His	Heme oxygenase, *C. diphteria*	1IW0	P/His
Globin-like			**Globin-like**		
Heme oxygenase, *H. sapiens*	IA3N	E/His	Hemoglobin, *P. caudatum*	IDLW	P/His
Neuroglobin, *M. musculus*			Neuroglobin, *M. musculus*		
Conformation I	1Q1F	E/bis-His	Conformation II	1Q1F	E/bis-His
P450 fold			**P450 fold**		
P450, *H. sapiens*	1PQ2	E/Cys	P450, *S. coelicolor*	1ODO	E/Cys
H-NOX fold			**H-NOX fold**		
sGC, *T. tengcongensis*	1U55	P/His	SONO, *T. tengcongensis*	1XBN	P/His
β-Propeller	1QHU	E/bis-His	4-Helix bundle	1BFR	P/bis-Met
Catalase	1A4E	E/Tyr	Vitamin B6 family	1JBQ	E/His-Cys
			Immunoglobulin-like	1D7B	E/His-Met
HemS-fold	2J0P	P/His	CAP	1FT9	P/His-Pro
			Helical membrane	1FFT	P/bis-His
PAS	1V9Y	P/His	Heme peroxidase	2CYP	E/His
			Lipocalin	1NP4	E/His
Serum albumin-like	1N5U	E/Tyr	Meander	1DK0	P/Tyr-His
			NO fold	1NOS	E/Cys

globins and cytochromes the heme is a prosthetic group that becomes an integral part of the structure. The latter are in fact less stable in their apo-forms and the folding process itself may rely on heme incorporation.[34,35] When high-affinity binding, which is nevertheless reversible, is the major function, multiple factors combine to yield pico- and femto-molar range binding affinities, as epitomised by the structures of HasA and hemopexin.[7,8,36]

Substantial efforts have been made towards the design of heme-proteins (for example[15,37]). One study pointed to a likely role of arginine residues in anchoring the propionate groups, and of leucine side chains in making hydrophobic contacts with the porphyrin, but in the final design experiments a very minimalist approach was maintained.[17] Most other design strategies focused primarily on the positioning of histidine ligands. Presumably, the limited stability of the resulting designed proteins[15,16,38,39] is to be, at least in part, attributed to steric interactions and non-specific packing between porphyrin atoms and protein side chains. The identification, therefore, of heme-binding interactions involving arginine, leucine and phenylalanine, should be beneficial for new generations of designed heme-proteins, particularly when catalysis is one of the design aims.[40] In these cases, the specific heme-packing motifs presented here may be used to maximise the association of heme while allowing for a trade off in stability, often required by enzymes for scaffold flexibility and modulation of geometry at the active site.

13 Conclusions

The heme binding sites of 68 b-type proteins with 20 different folds and less than 60% sequence identity have been compared by means of structural superpositions. This analysis revealed that multiple structural solutions are possible for binding the same, chemically identical heme ligand. Considerable diversity generally exists in the interactions employed by proteins to bind heme. Non-polar interactions are made with both the edge and the flat face of the heme. The edge contacts are particularly varied but some remarkable conservation was noticed in contacts made with the heme-face. Strikingly, leucine and isoleucine side chains were found at a re-occurring position making van der Waals contacts with the heme face. Aromatic rings were also found, in distinct folds, frequently aligned at three sites over the porphyrin, engaged in stacking interactions. Over two thirds of the proteins surveyed use arginine side chains to anchor the heme's propionate groups through electrostatic bridges. Other interesting contacts include hydrophobic side chains that pack against the heme iron proximal ligand, presumably helping to maintain the coordinating side chain in a suitable orientation for interacting with the iron. Structural superpositions thus showed that residues from unrelated structures, despite the great diversity of folding topologies, cluster at particular interaction "hot-spots" defining some common structural heme-binding motifs.

These observations identify interactions and structural motifs that contribute to the association and incorporation of heme by proteins, relevant to the understanding of structure–function relationships in heme-proteins and useful to the efforts made to design proteins able to incorporate this versatile and ubiquitous prosthetic group.

14 Acknowledgements

This work was supported by funds provided by the University of Nottingham, the Royal Society and the BBSRC. We thank Drs A. Smith (University of Missouri-Kansas City), P. D. Barker and F. Hollfelder (University of Cambridge) for critical reading of the manuscript.

15 References

1 S. K. Chapman, S. Daff and A. W. Munro, *Struct. Bonding*, 1997, **88**, 39–70.
2 M. Paoli, J. Marles-Wright and A. Smith, *DNA Cell Biol.*, 2002, **21**, 271–280.
3 J. W. Allen, P. D. Barker, O. Daltrop, J. M. Stevens, E. J. Tomlinson, N. Sinha, Y. Sambongi and S. J. Ferguson, *Dalton Trans.*, 2005, 3410–3418.
4 K. N. Degtyarenko, A. C. North and J. B. Findlay, *Protein Eng.*, 1997, **10**, 183–186.
5 A. Weichsel, J. F. Andersen, D. E. Champagne, F. A. Walker and W. R. Montfort, *Nat. Struct. Biol.*, 1998, **5**, 304–309.
6 T. O. Fischmann, A. Hruza, X. D. Niu, J. D. Fossetta, C. A. Lunn, E. Dolphin, A. J. Prongay, P. Reichert, D. J. Lundell, S. K. Narula and P. C. Weber, *Nat. Struct. Biol.*, 1999, **6**, 233–242.
7 P. Arnoux, R. Haser, N. Izadi, A. Lecroisey, M. Delepierre, C. Wandersman and M. Czjzek, *Nat. Struct. Biol.*, 1999, **6**, 516–520.
8 M. Paoli, B. F. Anderson, H. M. Baker, W. T. Morgan, A. Smith and E. N. Baker, *Nat. Struct. Biol.*, 1999, **6**, 926–931.
9 M. Wardell, Z. Wang, J. X. Ho, J. Robert, F. Ruker, J. Ruble and D. C. Carter, *Biochem. Biophys. Res. Commun.*, 2002, **291**, 813–819.
10 W. N. Lanzilotta, D. J. Schuller, M. V. Thorsteinsson, R. L. Kerby, G. P. Roberts and T. L. Poulos, *Nat. Struct. Biol.*, 2000, **7**, 876–880.
11 H. Kurokawa, D. S. Lee, M. Watanabe, I. Sagami, B. Mikami, C. S. Raman and T. Shimizu, *J. Biol. Chem.*, 2004, **279**, 20186–20193.
12 D. Barrick, *Biochemistry*, 1994, **33**, 6546–6554.
13 D. E. McRee, G. M. Jensen, M. M. Fitzgerald, H. A. Siegel and D. B. Goodin, *Proc. Natl. Acad. Sci. U. S. A.*, 1994, **91**, 12847–12851.
14 F. Rabanal, W. F. DeGrado and P. L. Dutton, *J. Am. Chem. Soc.*, 1996, **118**, 473–474.
15 B. R. Gibney and P. L. Dutton, *Protein Sci.*, 1999, **8**, 1888–1898.
16 N. R. Rojas, S. Kamtekar, C. T. Simons, J. E. McLean, K. M. Vogel, T. G. Spiro, R. S. Farid and M. H. Hecht, *Protein Sci.*, 1997, **6**, 2512–2524.
17 D. E. Robertson, R. S. Farid, C. C. Moser, J. L. Urbauer, S. E. Mulholland, R. Pidikiti, J. D. Lear, A. J. Wand, W. F. DeGrado and P. L. Dutton, *Nature*, 1994, **368**, 425–432.
18 C. Kisker, H. Schindelin, A. Pacheco, W. A. Wehbi, R. M. Garrett, K. V. Rajagopalan, J. H. Enemark and D. C. Rees, *Cell (Cambridge, Mass.)*, 1997, **91**, 973–983.
19 P. Nioche, V. Berka, J. Vipond, N. Minton, A. L. Tsai and C. S. Raman, *Science*, 2004.
20 P. Pellicena, D. S. Karow, E. M. Boon, M. A. Marletta and J. Kuriyan, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 12854–12859.
21 F. C. Bernstein, T. F. Koetzle, G. J. Williams, E. F. Meyer, Jr., M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi and M. Tasumi, *Arch. Biochem. Biophys.*, 1978, **185**, 584–591.
22 N. Guex, A. Diemand and M. C. Peitsch, *Trends Biochem. Sci.*, 1999, **24**, 364–367.
23 T. A. Jones, J. Y. Zou, S. W. Cowan and Kjeldgaard, *Acta Crystallogr., Sect. A: Found. Crystallogr.*, 1991, **47 (Pt 2)**, 110–119.
24 CCP4, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 1994, **50**, 760–763.
25 C. A. Hasemann, R. G. Kurumbail, S. S. Boddupalli, J. A. Peterson and J. Deisenhofer, *Structure (London)*, 1995, **3**, 41–62.
26 E. C. Liong, Y. Dou, E. E. Scott, J. S. Olson and G. N. Phillips, Jr., *J. Biol. Chem.*, 2001, **276**, 9093–9100.
27 G. B. McGaughey, M. Gagne and A. K. Rappe, *J. Biol. Chem.*, 1998, **273**, 15458–15463.
28 U. Samanta, D. Pal and P. Chakrabarti, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 1999, **55**(Pt 8), 1421–1427.
29 P. Stein, M. Mitchell and T. G. Spiro, *J. Am. Chem. Soc.*, 1980, **102**, 7795–7797.
30 B. Dangi, S. Sarma, C. Yan, D. L. Banville and R. D. Guiles, *Biochemistry*, 1998, **37**, 8289–8302.
31 K. B. Lee, G. N. La Mar, K. E. Mansfield, K. M. Smith, T. C. Pochapsky and S. G. Sligar, *Biochim. Biophys. Acta*, 1993, **1202**, 189–199.
32 G. B. Mortuza and D. Whitford, *FEBS Lett.*, 1997, **412**, 610–614.
33 B. Vallone, K. Nienhaus, M. Brunori and G. U. Nienhaus, *Proteins: Struct., Funct., Bioinf.*, 2004, **56**, 85–92.
34 T. E. Huntley and P. Strittmatter, *J. Biol. Chem.*, 1972, **247**, 4641–4647.
35 W. Pfeil, *Protein Sci.*, 1993, **2**, 1497–1501.
36 A. Smith and R. C. Hunt, *Eur. J. Cell. Biol.*, 1990, **53**, 234–245.
37 Z. Xu and R. S. Farid, *Protein Sci.*, 2001, **10**, 236–249.
38 C. T. Choma, J. D. Lear, M. J. Nelson, P. L. Dutton, D. E. Robertson and W. F. DeGrado, *J. Am. Chem. Soc.*, 1994, **116**, 856–865.
39 B. R. Gibney, F. Rabanal, K. S. Reddy and P. L. Dutton, *Biochemistry*, 1998, **37**, 4635–4643.
40 I. Obataya, T. Kotaki, S. Sakamoto, A. Ueno and H. Mihara, *Bioorg. Med. Chem. Lett.*, 2000, **10**, 2719–2722.
41 G. Vriend, *J. Mol. Graphics*, 1990, **8**, 52–56, 29.

