

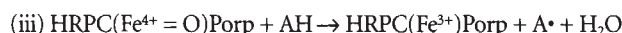
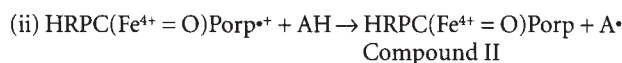
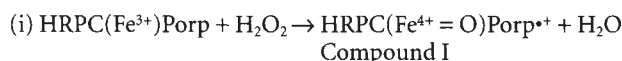
Crystal structure of horseradish peroxidase C at 2.15 Å resolution

Michael Gajhede¹, David J. Schuller², Anette Henriksen¹, Andrew T. Smith³ and Thomas L. Poulos²

The crystal structure of horseradish peroxidase isozyme C (HRPC) has been solved to 2.15 Å resolution. An important feature unique to the class III peroxidases is a long insertion, 34 residues in HRPC, between helices F and G. This region, which defines part of the substrate access channel, is not present in the core conserved fold typical of peroxidases from classes I and II. Comparison of HRPC and peanut peroxidase (PNP), the only other class III (higher plant) peroxidase for which an X-ray structure has been completed, reveals that the structure in this region is highly variable even within class III. For peroxidases of the HRPC type, characterized by a larger FG insertion (seven residues relative to PNP) and a shorter F' helix, we have identified the key residue involved in direct interactions with aromatic donor molecules. HRPC is unique in having a ring of three peripheral Phe residues, 142, 68 and 179. These guard the entrance to the exposed haem edge. We predict that this aromatic region is important for the ability of HRPC to bind aromatic substrates.

Classical horseradish peroxidase isozyme C (HRPC) is the archetypal enzyme for the biochemical study of peroxidases and is probably the most thoroughly studied member of the plant peroxidase superfamily¹. Within the superfamily, HRPC belongs to class III, the plant extracellular peroxidases².

The resting ferric enzyme reacts rapidly with hydrogen peroxide to form compound I (ref. 3), an oxy-ferryl species in which one electron has been withdrawn from the haem group to form a porphyrin π cation radical. This intermediate is reduced in two sequential one electron steps through compound II (in which the porphyrin cation radical has been reduced):



The role of the key catalytic residues of the distal haem pocket in mediating peroxide activation and compound I formation is well understood in terms of a general acid base mechanism^{4–8} but less is known about the factors that determine the nature and rate at which substrates are oxidized. In general the rates of reaction of compound I with substrates correlate well with the reduction potentials of their respective cation radical products⁹ and the highest occupied and lowest unoccupied molecular orbitals energy levels of substrates¹⁰. Class III peroxidases and HRPC in particular are notable in that the resting enzyme forms stable 1:1 complexes with reducing substrates, several of which have been extensively characterized by 2D NMR spectroscopy^{11–13}. These and more recent studies^{14,15} have shown that aromatic donor molecules interact with the haem C18 methyl and a key unassigned Phe residue thought to be either Phe 179 or Phe 221. The

topological characteristics of the haem-linked architecture associated with aromatic substrate oxidation have also been examined using alkyl hydrazines as suicide substrates^{16,17}. In the light of the structure of yeast cytochrome *c* peroxidase, these results were interpreted as suggesting that electron transfer is mediated at or close to a solvent accessible, reactive haem edge, and that direct access by bulky substrates to the ferryl intermediate is restricted by a closed haem architecture^{17,18}.

More recently, the availability of recombinant expression systems^{19,20} has resulted in a renewed interest in HRPC as a vehicle for studying the effect of protein environment on the modulation of haem-protein reactivity^{21–23}. A serious limitation in our ability to interpret both kinetic^{7,8,24,25} and spectroscopic^{14,26,27} data for this model system has been the lack of a crystal structure for this classical peroxidase.

Schuller *et al.*²⁸ published a 2.7 Å resolution structure of peanut peroxidase (PNP), which is the first complete structure of a class III peroxidase. In this paper we report the structure of recombinant HRPC to a resolution of 2.15 Å. Although the structure of HRPC is similar to PNP, sequence alignments of over 60 members of class III suggest that sequence variability is confined to several 'hot spots' in the structure which include part of the region bounding the aromatic donor molecule interaction site and substrate access channel. These features exhibit small but distinct differences from those of peanut peroxidase, affecting the accessibility of the reactive haem edge to aromatic substrates. Taken together, the structures of horseradish and peanut peroxidases, which belong to opposite ends of the plant peroxidase phylogenetic tree²⁹, will provide a close structural representative for many of the class III plant peroxidase sequences currently identified.

Overall structure and comparison to other peroxidases
HRPC was expressed in *E. coli* from the synthetic gene comprising 309 amino acid residues including the additional N-terminal

¹Department of Chemistry, University of Copenhagen, Universitetsparken 5, DK-2100, Copenhagen, Denmark. ²Department of Molecular Biology and Biochemistry, University of California, Irvine, California 92697-3900, USA. ³School of Biological Sciences, University of Sussex, Falmer, Brighton, BN1 9QG, UK.

Correspondence should be addressed to M.G. email: gajhede@jerne.ki.ku.dk and A.T.S. email: A.T.Smith@sussex.ac.uk

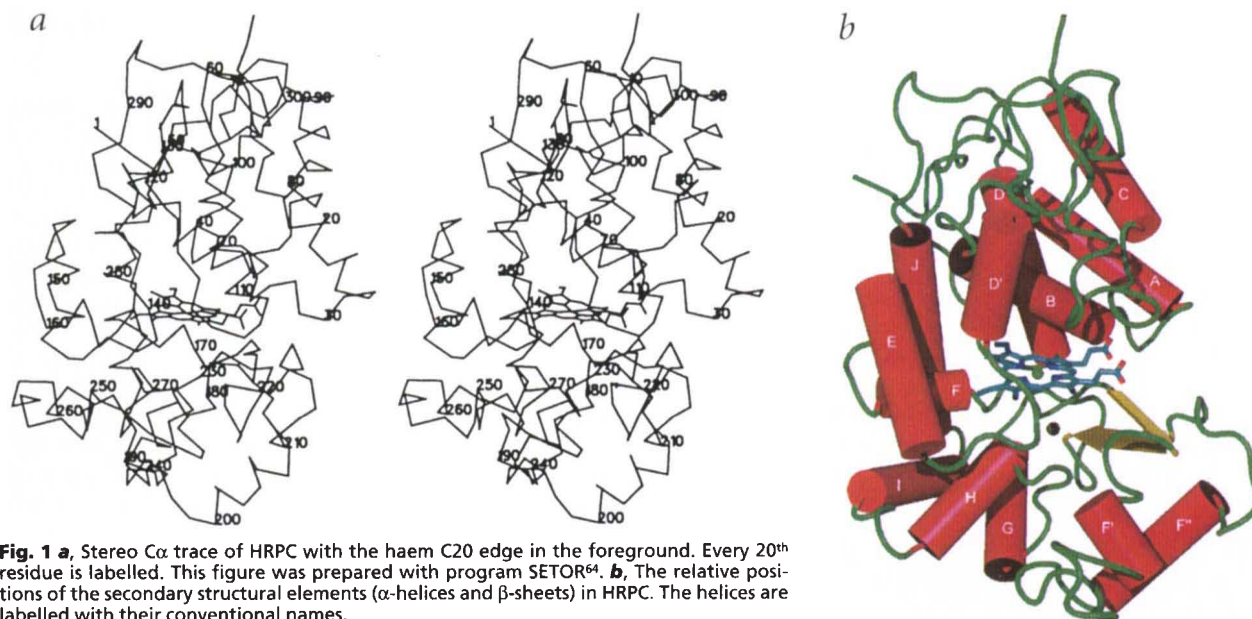


Fig. 1 **a**, Stereo C α trace of HRPc with the haem C20 edge in the foreground. Every 20th residue is labelled. This figure was prepared with program SETOR⁶⁴. **b**, The relative positions of the secondary structural elements (α -helices and β -sheets) in HRPc. The helices are labelled with their conventional names.

Met required to initiate translation¹⁹. Only 306 residues are visible in the crystal structure. The N-terminal Met and the extreme C terminus (residues Asn 307–Ser 308) are disordered. A numbered C α trace of HRPc is shown in Fig. 1a, and the overall fold in Richardson representation is shown in Fig. 1b. HRPc has the same overall fold typical of the peroxidase superfamily, with the haem group and calcium binding sites in similar positions.

Structurally derived amino acid sequence alignments of representative members of the various peroxidase classes are shown in Fig. 2. Representative peroxidases from the three classes have been included: cytochrome *c* peroxidase (CCP)³⁰ and ascorbate peroxidase (APX)³¹ from class I, *arthromyces ramosus* peroxidase (ARP)³² and lignin peroxidase (LIP)^{33,34} from class II and finally PNP²⁸ and HRPc from class III. The ten prominent helices, A–J, are all found in similar positions, and there are 13 structurally conserved regions (SCRs; Fig. 2) that have a root-mean-square (r.m.s.) deviation with respect to the C α position in the six structures of less than 2.5 Å. Notably not all the SCRs are helices; for example the extended strand connecting helices D and D' is conserved in all six structures, while the short B' helix found in class I and class II peroxidases has become a 3_{10} -helix in HRPc and PNP.

A pair-wise r.m.s. comparison for six representative peroxidase structures using only the SCRs identified in the structural alignment (Fig. 2) is shown in Table 1. This shows that APX has the most average core fold of any peroxidase, accounting for its success as a search model for the molecular replacement solution of PNP.

A superposition of the ribbon representations of HRPc and PNP is shown in Fig. 3. The regions exhibiting the greatest degree of structural variation are labelled. PNP contains a glycosylation site located at Asn 60, which is right at the most exposed end of a loop. The nearest glycosylation site in wild type HRPc is at Asn 57 which is positioned to one side of the loop, thus probably causing the difference in structure. HRPc has a single insertion relative to PNP at residue Leu 157, leading to differences in main chain conformation. HRPc and PNP both contain three helices, the D', F' and F'' helices, which are additional to the core peroxidase fold and are not found in class I or II. The D' helix is

identical in length to that seen in PNP. The class III peroxidases contain a long insertion between helices F and G maintained by a disulphide bridge, Cys 176–Cys 209 (HRPc numbers). This insertion is seven residues longer in HRPc than in PNP and is generally hypervariable within class III. The insertion contains the F' and F'' helices, as suggested by secondary structure predictions³⁵. The F'' helices are of identical length in PNP and HRPc, whereas in HRPc the F' helix is slightly shorter and the surface loop connecting these two helices is seven residues longer. The shortened F' helix and altered side chain composition relative to PNP modulate the character and accessibility of the substrate access channel in this vicinity. HRPc also contains a short antiparallel β -sheet, immediately preceding the disulphide bridge Cys 176–Cys 209. The loop region Pro 210–Asn 214 in HRPc differs significantly from the equivalent region in PNP, Pro 202–Asn 206. This variation was unexpected since Pro 210 and Gly 213 occur at the corresponding positions in PNP. A three residue insertion in HRPc between helices H and I (Fig. 1) is partly accommodated by a relative prolongation of the H helix, which is eight residues long in HRPc, but only five in PNP. Apart from these differences, the structures of these two class III peroxidases are quite similar. A structural alignment using 292 C α atoms carried out with the program O³⁶ gives an overall r.m.s. deviation of 1.02 Å.

The glycosylation sites of wild type HRPc are also shown in Fig. 3 (ref. 37). All sites are N-glycosylation sites, positioned in loop regions that are exterior to the core, except for Asn 268 which, although solvent accessible, is located right at the end of helix I. All glycosylation sites point away from the molecule, and are evenly distributed over the entire surface suggesting that the main purpose of glycosylation is to increase the solubility of the enzyme in water and perhaps to increase resistance to free radical induced protein crosslinking. This may account for the enhanced sensitivity of the recombinant enzyme to inactivation by high hydrogen peroxide concentration³⁸.

Distal haem pocket and distal Ca²⁺ site

The distal pocket of HRPc has a very strong resemblance to that of PNP. The disposition of the key conserved catalytic residues

articles

Arg 38, Phe 41 and His 42 are very similar. The His 42 dihedral angles defined by atoms C α -C β -C γ -N δ 1 differ by $\sim 30^\circ$ but this leaves the N ϵ 2's in the same positions relative to the iron. The distal pocket found in HRPC conforms to expectations from the proposed three-step reaction mechanism for compound I formation^{4,30,39}. The Fe(III) ion of resting HRPC has been shown by resonance Raman spectroscopy to be in a five-coordinate, high spin state⁴⁰. This is in agreement with results obtained for wild-type plant HRPC⁴¹ where it was reported that the Fe(III) ion was a mixture of five-coordinated and six-coordinated high spin states with the five-coordinate state predominating. As indicated on Fig. 4a the closest water molecule in the distal pocket is 3.2 Å from the Fe(III) position. This distance indicates that the iron of the haem is pentacoordinate. The closest water molecule, Wat 92, is hydrogen bonded to the N ϵ of Arg 38 (3.0 Å), the N ϵ 2 of His 42 (2.9 Å) and also to another water molecule Wat 123. Wat 123 makes a hydrogen bond to N η 2 of Arg 38 (2.75 Å) and probably also to a disordered water, not included in the model. The distal cavity of PNP has a water molecule at the same position as Wat 92 in HRPC; Wat 123 of HRPC is absent in PNP. The Fe(III)-ion does not seem to be significantly displaced from the haem plane.

Mutagenesis studies at Asn 70 (ref. 25) have highlighted the role of this residue in maintaining the basicity of the distal His. A structural explanation for this can be seen in the HRPC crystal structure. The O δ 1 of N70 is hydrogen bonded to N δ 1 of the distal histidine (His 42) and the N δ 2 is hydrogen bonded to the backbone oxygen of Glu 64. N δ 2 of N70 is also hydrogen bonded through Wat 14 to Glu 64 O ϵ 1. Glu 64, which is conserved amongst all members of the plant peroxidase family, hydrogen bonds to Wat 15, which is also a ligand of the distal Ca²⁺ site. This hydrogen bond network is shown in Fig. 4a.

The Ca²⁺ site is seven-coordinate, and the ligands are carbonyl and side chain oxygens. Additionally, Ca²⁺ is structurally coupled to the active site through Asp 43, which provides two ligands to the ion, and which is adjacent in sequence to His 42. Note that Ca²⁺ is required for formation of the structural environment of the haem, and during *in vitro* folding of the denatured apoprotein^{19,42}. Extremes of pH or Ca²⁺ depletion is expected to affect

the position of the distal histidine and affect catalytic performance of the enzyme⁴³.

Proximal haem pocket and hydrogen bonding networks

A view of the surroundings of the haem in HRPC, including the hydrogen-bonding network, is shown in Fig. 4b. His 170 (the proximal His) is covalently bonded to the iron as in the other known haem peroxidase structures. Direct hydrogen bonds exist between the protein and the haem propionates through Gln 176, Ser 73, Ser 35 and Arg 31. The proximal His is indirectly coupled to the proximal Ca²⁺ site through the side chain of an adjacent residue, Thr 171, which is a Ca²⁺ ligand.

Both Ca²⁺ sites are in the same relative positions with respect to the haem group as in the other class II and class III peroxidases and are identical to those found in PNP²⁸. The presence of the two Ca²⁺ sites, although not obviously required to achieve a stable peroxidase fold (see, for example, APX and CCP), has been suggested to relate to a gene duplication at an early stage in the evolution of the plant peroxidase superfamily⁴⁴.

The distal pocket is coupled to the proximal side by a hydrogen-bonding network involving Arg 38, which is hydrogen-bonded through Wat 41 to a haem propionate (Fig. 4b). Gln 176 appears to play an analogous role to that of His 181 in CCP by connecting the two halves of the extended hydrogen-bonding network above and below the haem group. Confirmation of this structural connection explains the complex chemical shift perturbations exhibited by protons of haem-linked proximal

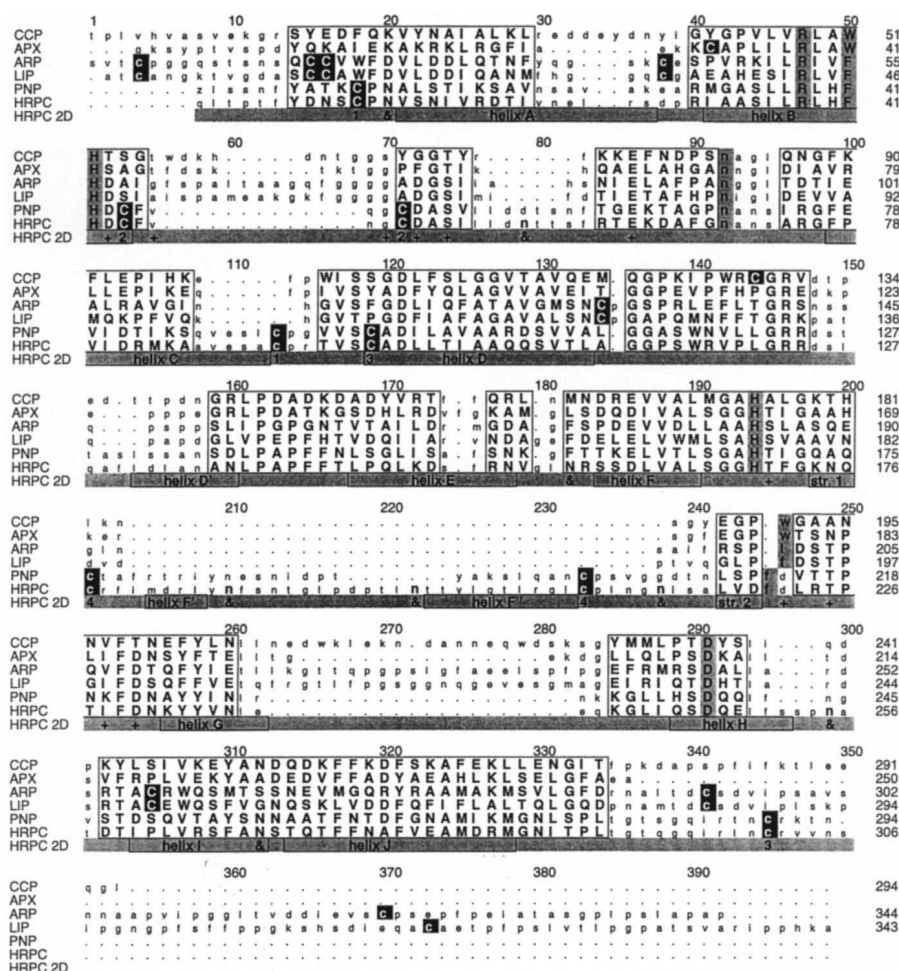


Fig. 2 Structural alignment of representative members of the peroxidase superfamily. Coordinate files were obtained from the Protein Data Bank at Brookhaven. The alignment was made with the programs STAMP⁶⁵, AMP5⁶³ and ALSCRIPT⁶⁶. Boxed, capitalized regions exhibit less than 2.5 Å r.m.s. deviation for C α in all six structures. The last line, in gray, indicates secondary structure of HRPC. Cys residues are shown in black boxes, with pairings for HRPC (and PNP) indicated by numerals. Some important active site residues are highlighted in gray boxes. Calcium ligands are indicated by '+', and glycosylation sites of native HRPC by '&'.

Table 1 a, Pairwise r.m.s distances¹ and b, % sequence identities²

a	CCP	APX	ARP	LIP	PNP
APX	1.06				
ARP	1.78	1.68			
LIP	1.76	1.69	0.75		
PNP	2.02	1.93	2.29	2.31	
HRPC	1.94	1.88	2.24	2.29	0.81
b	CCP	APX	ARP	LIP	PNP
APX	38.7				
ARP	17.4	27.5			
LIP	19.5	24.6	43.9		
PNP	16.6	25.5	19.2	15.9	
HRPC	20.8	24.9	19.6	18.5	49.3

¹Pairwise r.m.s distances for six structures representative of the superfamily based on the 183 structurally conserved residues (C α atoms only) boxed in the amino acid sequence alignment in Fig. 2 (CCP4 program POLYPOSE⁵⁸). Notes for Table legend RMS distances between all structures: 1.83, from mean structure: 1.18 note that once again APX is the best 'average' or 'representative' structure.

²Pure sequence based % amino acid identity reported by the AMPS program with default parameters⁶³.

residues detected in the 1D- and 2D-NMR spectra of cyanide-ligated HRPC distal cavity mutants⁴⁵ and the weakening of the Fe-imidazolate bond detected by Raman spectroscopy in mutants where this hydrogen-bond network is likely to be disrupted²⁶.

The N δ 1 of the proximal haem ligand His 170 is seen to be hydrogen bonded to Asp 247 (2.8 Å). The presence of a strong hydrogen bond is believed to increase the basicity of the His 170 proximal ligand relative to the globins, thus helping to stabilize high oxidation state intermediates by a charge relay system while helping to maintain the haem in a five-coordinate state. The functional significance of an imidazolate proximal ligand in the formation of compound I has been questioned by the finding that proximal ligand mutations in CCP are largely functional⁴⁶. His 170 is seen to be 'stacking' in a hydrophobic interaction with Phe 221. The corresponding aromatic residue in the CCP sequence is Trp 191; it also stacks against the proximal histidine. CCP Trp 191 is known to be the free radical site in compound I^{47,48} and is held in a local hydrogen-bonding network involving the conserved aspartate residue (Asp 235) and the axial ligand (His 175). These alternate structural arrangements plus the lower oxidation potential of Trp relative to Phe could explain the properties of CCP. However, APX also has a Trp at this position, and yet APX forms a compound I porphyrin radical^{31,49}. Hence CCP's unique compound I radical site has been attributed to the lack of a proximal Ca²⁺ binding site in CCP³¹. Other class III peroxidases contain either a Leu or occasionally a Met at this position, suggesting that an aromatic interaction at this site is not essential.

Substrate access channel

The general electrostatic character of the substrate access channel can be described as having a peripheral hydrophobic layer, contributed by Phe residues and the haem C18 methyl, while the inner channel and lining of the haem cavity has an overall positively charged character due primarily to the distal Arg

38. Phe 68, Phe 142 and Phe 179 form a distinctive hydrophobic patch near the exposed haem edge (Fig. 5a). Two of these, Phe 68 and Phe 179, are either absent or shifted in position in PNP. Several NMR studies have long implicated two Phe residues situated towards the entrance of the aromatic donor-interaction site^{11,12}. One of these in particular exhibits distinct NOEs to *para*-cresol and ferulic acid while both exhibit NOEs to indole-3-propionic acid¹¹. The former exhibits strong NOEs to the haem C18 methyl. The HRPC crystal structure strongly suggests that the critical Phe residue is Phe 179, a suggestion confirmed by recent analysis of Phe 179 mutants⁵⁰. It seems reasonable to suggest that this proposed site is the site of the pre-electron transfer complex formed between substrates and high oxidation state intermediates. It is precisely this region of the structure that has proven intractable to molecular modelling due to the lack of a homologous template^{27,35}.

HRPC and ARP show many common enzymatic properties including a broad pH optimum for phenolic compounds⁵¹. As shown in the solvent accessible surface diagrams in Fig. 5b, the haem periphery of both peroxidases is well exposed to solvent. However, Phe 179 in HRPC is positioned close to the haem edge sterically hindering the access channel. ARP contains no such structurally analogous Phe residue, since class II enzymes lack the additional F' helix. For the majority of class III peroxidases, like for example PNP, the Phe at position 179 is displaced one position in sequence where it points away from the channel and does not occlude it. This region of the protein is also largely responsible for the superior ability of HRPC to form stable 1:1 complexes with reducing-substrates and presumably product radicals.

Conclusion

Horseradish peroxidase (HRPC) has been the archetypal example of a class III, higher plant peroxidase, partly due to its availability and worldwide use in immunodiagnostic kits and assays as a reporter enzyme. Within class III, HRPC and PNP (peanut peroxidase) occupy distant branches of the phylogenetic tree and therefore provide a close structural representative for the many class III sequences currently identified, which should facilitate future rationalization of structure-function relationships. The availability of this new structure for HRPC and a number of site-directed mutants will give a fresh impetus to the investigation of haem protein structure-function relationships using this model system. For example, it has been possible to engineer peroxyge-

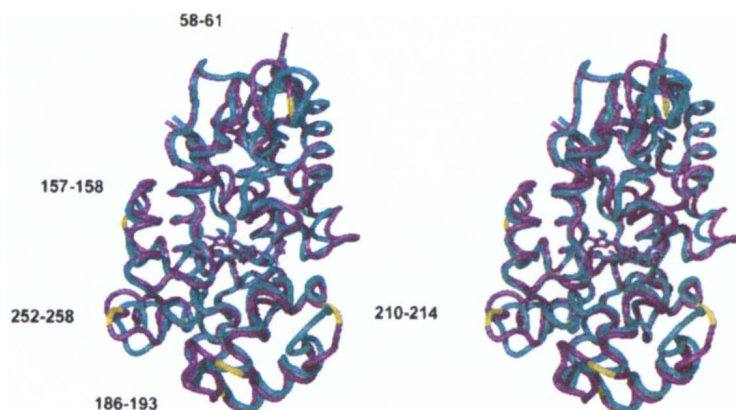
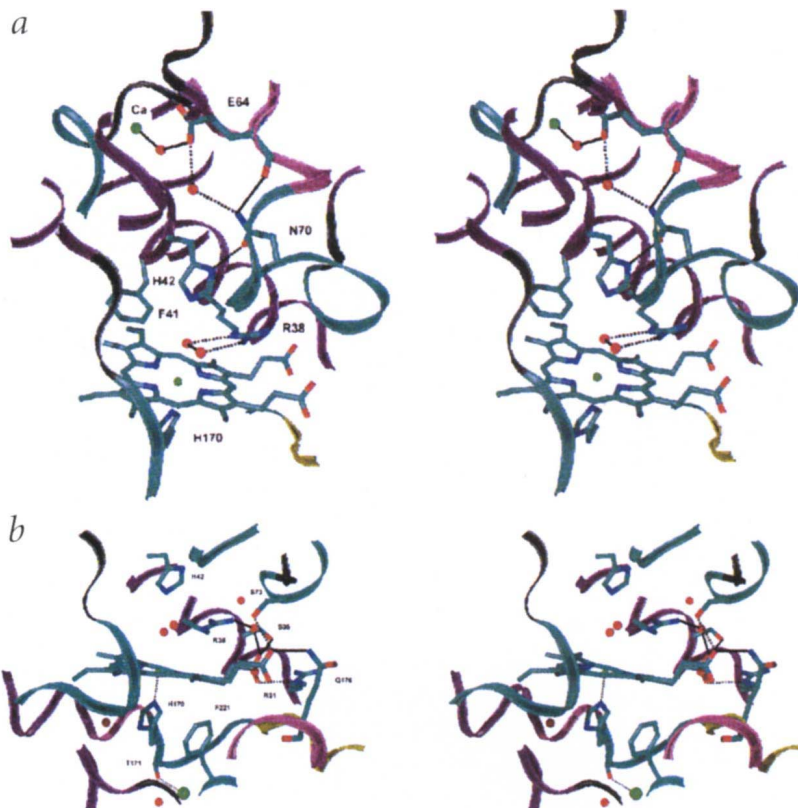


Fig. 3 Superposition of the ribbon representations of HRPC (magenta) and PNP (cyan). Regions of greatest structural deviation are labelled. Glycosylation sites identified in native HRPC are coloured yellow. This figure was made with program VMD⁶⁷.

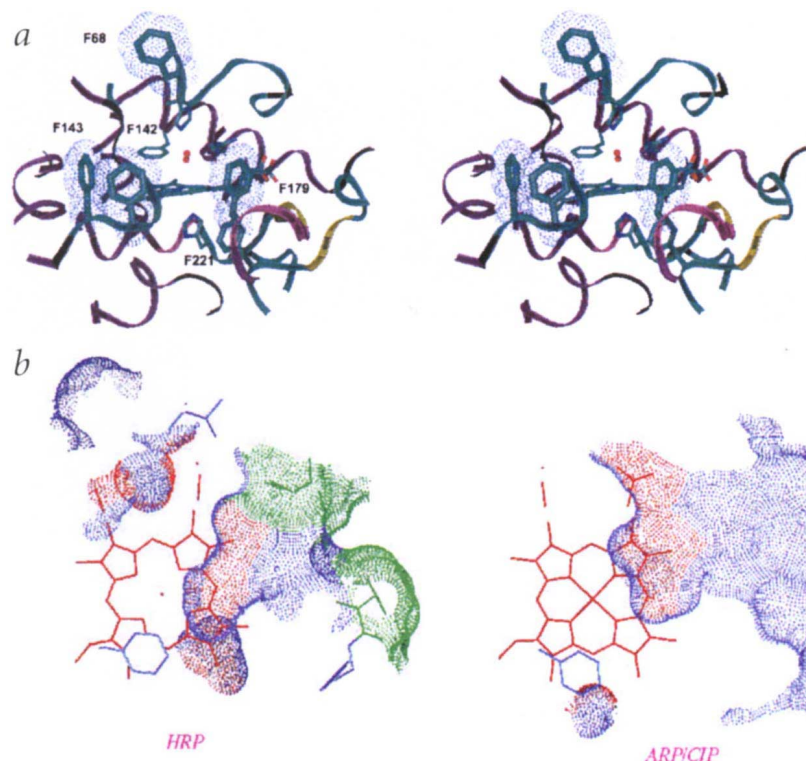
articles

Fig. 4 a, Distal haem cavity and associated hydrogen bonding network. The protein backbone is shown in a ribbon representation, coloured after secondary structure classification: purple for α -helix, yellow for β -strand, green for turns and gray for coil. Side chains of distal pocket residues (above haem in this view) from helix A, Phe 41, Arg 38 and His 42, and proximal histidine His 170 are also shown along with the distal waters. The extended hydrogen-bonding network from His 42 through Asn 70, water, Glu 64 and coordinating water to distal Co (green) is also shown. An only partially occupied distal water molecule, also hydrogen bonding to the distal Pro 139 O, is not shown. This figure was made with program VMD⁶⁷. **b**, Anchoring of the haem group, linkage from distal pocket to proximal pocket and linkage from proximal pocket to proximal calcium site. This figure shows the hydrogen-bonding network from the distal Arg, through water, to the propionate of the haem. From helix A, in the background, Arg 31 and Ser 35 make hydrogen bonds to the propionate deepest in the cavity. In the foreground, Ser 73 and Gln 176 make hydrogen bonds to the other propionate. The backbone oxygen of Gln 176, making the link to the hydrogen-bonding network of the proximal pocket is not shown. The two water molecules of the proximal pocket are shown. Phe 221 is seen to 'stack' with the proximal histidine His 170. Thr 171 which is a ligand to the proximal calcium is also shown. This figure was made with program VMD⁶⁷.



nase-like activity by increasing the accessibility of the ferryl intermediate to substrates such as styrene or substituted thioanisoles²¹. Substitutions of the distal Arg have allowed the detection of transient intermediates formed during peroxide activation and decay

of oxy peroxidase²⁴. The key aromatic residue involved in the complex formed between aromatic donors and the resting enzyme has been identified⁵⁰. This structure will facilitate the construction of models of aromatic donor-HRPC complexes, which should guide future attempts to tailor peroxidase specificity.



Methods

Purification. HRP is a monomeric haem protein of 308 residues containing two structural Ca^{2+} ions and eight N-linked neutral glycans³⁷. The enzyme was initially crystallized by Braithwaite⁵², but since the crystals had 32 molecules in the unit cell, detailed X-ray analysis was not pursued. The plant enzyme has been shown to contain 21 components, most of which are glycovariants⁵³. In contrast, the recombinant enzyme produced by *in vitro* refolding of material isolated from *E. coli* inclusion bodies¹⁹, is 96% homogeneous. However, the recombinant enzyme also crystallized in a unit cell with a large number of symmetry independent molecules⁵⁴.

Fig. 5 a, Critical Phe residues guarding the substrate access channel with their accessible surfaces contoured. The entrance to the reactive haem edge is viewed edge on showing the presumptive substrate interaction site. Phe 68 is situated above the distal pocket, Phe 142 to the left and Phe 179 to the right of the substrate access channel. This figure was made with program VMD⁶⁷. **b**, Comparison of the substrate access channels of HRP (left) and ARP (right) showing a view of the molecular surfaces. The peripheral Phe residues that guard access to the reactive haem edge are shown in green. The view is orthogonal to the view in (a) with Phe 142 toward the bottom of the panel and Phe 179 at the top. Note the more open nature of the access channel in ARP.

Table 2 Progress of molecular replacement solution¹ and data collection statistics

#mols	CC	R-factor
1	0.113	0.693
2	0.455	0.580
3	0.531	0.546
4	0.590	0.517
6	0.697	0.444
Data collection statistics		
Maximum resolution (Å)	2.15	
Outer shell (Å)	2.19–2.15	
No. of observations	142,833	
No. of unique reflections	73,216	
Completeness (%)		
All data	82.6	
Outer shell	65.4	
R_{sym}		
All data	0.070	
Outer shell	0.401	

¹Numbers reported are following RBR. Molecular replacement and RBR carried out with AMoRe⁵⁶, using data in the 20.0–4.0 Å resolution range.

Crystallization. Crystals were prepared as described⁵⁴. The space group was determined to be P3₁12 or the enantiomorphic P3₂12 and it was suggested that the unit cell contained up to six molecules related by non-crystallographic symmetry (NCS). A lack of non-crystallographic peaks on the self-rotation function, along with centering peaks in the native Patterson function led to the conclusion that the non-crystallographic symmetry consisted of translations only. This knowledge proved to be crucial during pursuit of the molecular replacement solution. In addition, the lattice was sparsely populated at low resolution due to an approximate superlattice, a phenomenon related to the translation-only NCS.]

Data collection. Data were collected from a single native crystal using a Rigaku R-axis IIC image plate system with a Rigaku RU200 rotating anode operating at 180 mA and 50 kV as the X-ray source. The system was equipped with a graphite monochromator and a 0.5 mm collimator. The crystal was rotated through 52 frames, each with an oscillation of 1.2°. The distance between crystal and detector was 140 mm and a 2θ offset of 20° was used. Data integration and reduction were performed with Denzo and Scalepack⁵⁵. X-ray data statistics are listed in Table 2.

Molecular replacement. Molecular replacement was carried out with AMoRe⁵⁶. The most successful search model consisted of a full molecule of peanut peroxidase (PNP)²⁸. No successful molecular replacement solutions were obtained with a resolution range of 20–5 Å in either space group P3₁12 or P3₂12, and the *R*-factor for translation function solutions was consistently above 0.75. The superlattice approximation described above was causing sparse population of the lattice at low-resolution⁵⁴, so the resolution range for molecular replacement was broadened. A rotation search at 20–3.5 Å resolution produced a top solution with a correlation coefficient (CC) of 20.8, with the next best solution being 12.5.

Translation searches were carried out at 20–4 Å resolution in space groups P3₁12, P3₂12, and P3₁12. The top solution for space group P3₁12 had a CC of 0.091 with an *R*-factor of 0.697. The best CC of the other space groups was 0.075 and the best *R*-factor 0.694. With rigid body refinement (RBR) the P3₁12 solution improved to a CC of 0.113 with an *R*-factor of 0.693, while the top solutions of the other space groups did not improve beyond a CC of 0.079 and an *R*-factor of 0.693. Subsequent molecules were added in space group P3₁12 according to one of two strategies: (i) We fixed the first solution and searched for

additional solutions with a translation function search. (ii) We positioned additional subunits according to the centering peaks of the native Patterson map. Happily these two strategies converged to the same solutions. As additional molecules were added the CC improved rapidly and the *R*-factor more slowly as detailed in Table 2. A total of six molecules were placed in the asymmetric unit with only slight rotational differences and these approximate translations: (0, 0, 0); (1/3, 2/3, 0); (2/3, 1/3, 0); (0, 0, 1/2); (1/3, 2/3, 1/2); (2/3, 1/3, 1/2). This explains the superlattice seen in the initial analysis of diffraction data. The calculated solvent content is approximately 0.392 (ref. 57).

The six molecules were stripped to poly-alanine chains and used to calculate model phases at 3.5 Å resolution with CCP4 programs SFALL and SIGMAA⁵⁸. These phases, along with phases from a monoclinic crystal form with two molecules in its asymmetric unit (a complex of HRP and benzydroxamic acid, results to be published separately), were refined and extended to 2.3 Å resolution using solvent leveling, histogram matching, eight-fold non-crystallographic symmetry (NCS) averaging and partial structural information with program MAGIC-SQUASH⁵⁹.

Model building and refinement. Electron density maps for model building were calculated either with the extended phases or using conventional $2F_o - F_c$ and $F_o - F_c$ methods (including simulated annealing omit-maps). Model building was done with the program O³⁶ and the OOPS procedure⁶⁰ for locating and correcting suspicious regions. Minimization and simulated annealing were performed using program X-PLOR⁶¹, with the Engh-Huber parameters⁶² and no formal charges on any residues. 10% of the data were set aside for a free-*R* test set. At all stages strict NCS constraints were applied. Apart from the translational components, rotations of up to approximately 10° are involved in the relationships of the NCS-related molecules. Water molecules were included in the model if they obeyed hydrogen-bonding criteria, displayed difference electron density above 3σ and returned $2F_o - F_c$ electron density. At later stages a bulk-solvent correction was included with the optimal scale factor and temperature factor was empirically determined to be 0.32 and 80 Å². With the correction, all low-order data measured (no σ-cutoff was applied) could be applied with no increase in the conventional *R*-factor and a decrease in the free *R*-factor, while the quality of the electron density map improved. A total of 306 amino acid residues were located.

During the strict NCS refinement cycles the OOPS procedure pointed out the following residues: Val 46, Asn 50 and Ala 51 have distorted geometry, all are involved in forming the proximal calcium binding site. Arg 75 and Leu 121 show high pep-flip⁶⁰ values but the electron density is consistent with the model. Phe 172 likewise has a high pep-flip indicator, but this residue is part of the distal calcium binding site. Asn 182 displays a real space (RS)⁶⁰ *R*-factor (0.31). The region Asn 189–Thr 190 has high RS *R*-factor and bad pep-flip values. This appears to be a mobile loop, consequently with a less well-defined electron density. Similarly the region Leu 211–Gly 213 has relatively high RS *R*-factors, also consistent with a mobile region. Thr 288 has poor phi/psi angles. Asp 300 is found to have a high pep-flip value, probably due to the hydrogen-bonding interaction with Wat 56. After the strict refinement an *R*-value of 21% and a free *R*-value of 23% were obtained. This model contains 150 water molecules.

A cluster of five very high peaks in the electron density was found between two symmetry related molecules. They formed a trigonal bipyramid with the three-fold axis parallel to the crystallographic three-fold screw axis. One peak was on a crystallographic two-fold giving a total of three symmetry independent peaks. The distances between the peaks were found to be around 2.2 Å. The peaks could be a cluster of heavy atoms (for example, zinc ions from the crystallization medium) or an experimental artifact. Since no satisfying model could be obtained, and the position indicates no biological significance, no description of these peaks was included in the model.

Acknowledgements

This work was supported by Danish National Research Foundation and the Biotechnology and Biological Research Council grants to A.T.S., D.J.S. and T.L.P. were supported by the U.S. National Science Foundation.

Received 7 August, 1997; accepted 31 October, 1997.

1. Dunford, H.B. Horseradish peroxidase: Structure and Kinetic Properties. In *Peroxidases in Chemistry and Biology* Vol. 2 (eds Everse, I., Everse, K.E. & Grisham, M.B.) 1–24 (CRC Press, Boca Raton; 1991).
2. Welinder, K.G. Plant peroxidases: Structure-function relationships. In *Plant Peroxidases 1980–1990. Topics and detailed literature on molecular, biochemical and physiological aspects.* (eds Penel, C., Gaspar, T. & Greppin, H.) 1–24 (University of Geneva, Switzerland; 1992).
3. Chance, B. The kinetics of the enzyme-substrate compound of peroxidase. *J. Biol. Chem.* **151**, 553–577 (1943).
4. Poulos, T.L. & Kraut, J. The stereochemistry of peroxidase catalysis. *J. Biol. Chem.* **255**, 8199–8205 (1980).
5. Erman, J. E. et al. Histidine 52 is a critical residue for rapid formation of cytochrome c peroxidase compound I. *Biochemistry* **32**, 9798–9806 (1993).
6. Vitello, L.B., Erman, J.E., Miller, M.A., Wang, J. & Kraut, J. Effect of Arginine-48 replacement on the reactions between cytochrome c peroxidase and hydrogen peroxide. *Biochemistry* **32**, 9807–9818 (1993).
7. Rodriguez-Lopez, J.N., Smith, A.T. & Thorneley, R.N.F. Recombinant Horseradish Peroxidase Isoenzyme C: The Effect of Distal Haem Cavity Mutations (His 42→Leu and Arg 38→Leu) on Compound I Formation and Substrate Binding. *J. Biol. Inorg. Chem.* **1**, 136–142 (1996).
8. Rodriguez-Lopez, J.N., Smith, A.T. & Thorneley, R.N.F. Role of Arg-38 in Horseradish peroxidase. A critical residue for substrate binding and catalysis. *J. Biol. Chem.* **271**, 4023–4030 (1996).
9. Candeias, L.P., Folkes, L.P., Porssa, M., Parrick, J. & Wardman, P. Rates of reaction of indoleacetic acids with horseradish peroxidase compound I and their dependence on redox potentials. *Biochemistry* **35**, 102–108 (1996).
10. Sakurada, J., Sekiguchi, R., Sato, K., & Hosoya, T. Kinetic and molecular orbital studies on the rate of oxidation of monosubstituted phenols and anilines by horseradish peroxidase compound I. *Biochemistry* **29**, 4093–4098 (1990).
11. Veitch, N.C. & Williams, R.J.P. Two dimensional ¹H-NMR studies of horseradish peroxidase C and its interaction with indole-3-propionic acid. *Eur. J. Biochem.* **189**, 351–362 (1990).
12. Veitch, N.C. & Williams, R.J.P. Two dimensional proton nuclear magnetic resonance studies of plant peroxidase interactions with aromatic donor molecules. In *Biochemical, molecular and physiological aspects of plant peroxidases* (eds Lobarzewski, J., Greppin, H., Penel, C. & Gaspar, T.) 99–109 (University of Geneva, Switzerland; 1991).
13. Sakurada, J., Takahashi, S. & Hosoya, T. Nuclear Magnetic Resonance Studies on the Spatial Relationship of Aromatic Donor Molecules to the Heme Iron of Horseradish Peroxidase. *J. Biol. Chem.* **261**, 9657–9662 (1986).
14. Veitch, N.C., Williams, R.J.P., Bone N.M., Burke, J.F. & Smith, A.T. *Eur. J. Biochem.* **233**, 650–658 (1995).
15. Veitch, N.C. & Williams, R.J.P. The use of methylsubstituted benzhydroxamic acids as structural probes of peroxidase substrate binding. *Eur. J. Biochem.* **229**, 629–640 (1995).
16. Ator, M.A. & Ortiz de Montellano, P.R. Protein control of prosthetic heme reactivity. *J. Biol. Chem.* **262**, 1542–1551 (1987).
17. Ortiz de Montellano, P.R. Catalytic sites of hemeprotein peroxidases. *Annu. Rev. Pharmacol. Toxicol.* **32**, 89–107 (1992).
18. Miller, V.P., DePillis, G.D., Ferrer, J.C., Mauk, A.G. & Ortiz de Montellano, P.R. Monooxygenase Activity of Cytochrome c Peroxidase. *J. Biol. Chem.* **267**, 8936–8942 (1992).
19. Smith, A.T. et al. Expression of a synthetic gene for horseradish peroxidase C in *Escherichia coli* and folding and activation of the recombinant enzyme with Ca²⁺ and heme. *J. Biol. Chem.* **265**, 13335–13343 (1990).
20. Hartmann, C. & Ortiz de Montellano, P.R. Baculovirus expression and characterization of catalytically active horseradish peroxidase. *Arch. Biochem. Biophys.* **297**, 61–72 (1992).
21. Newmyer, S.L. & Ortiz de Montellano, P.R. Horseradish peroxidase H42A, H42V and F41A, mutants—histidine catalysis and control of substrate access to the heme iron. *J. Biol. Chem.* **270**, 19430–19438 (1995).
22. Newmyer, S.L. & Ortiz de Montellano, P.R. Rescue of the catalytic activity of an H42A mutant of horseradish peroxidase by exogenous imidazoles. *J. Biol. Chem.* **271**, 14891–14896 (1996).
23. Newmyer, S.L., Sun J., Loehr, T.M. & Ortiz de Montellano, P.R. Rescue of the horseradish peroxidase H170A mutant activity by imidazole - importance of proximal ligand tethering. *Biochemistry* **35**, 12788–12795 (1996).
24. Rodriguez-Lopez, J.N., Smith, A.T. & Thorneley, R.N. Effect of distal cavity mutations on the binding and activation of oxygen by ferrous horseradish peroxidase. *J. Biol. Chem.* **272**, 389–395 (1997).
25. Nagano, S., Tanaka, M., Ishimori, K., Watanabe, Y. & Morishima, I. Catalytic roles of the distal site asparagine-histidine couple in peroxidases. *Biochemistry* **35**, 14251–14258 (1996).
26. Howes, B.D., Rodriguez-Lopez, J.N., Smith, A.T. & Smulevich, G. Mutation of distal residues of horseradish peroxidase: influence on substrate binding and cavity properties. *Biochemistry* **36**, 1532–1543 (1997).
27. Zhao, D., Gilfoyle, D. J., Smith A.T. & Loew, G. Refinement of 3D models of Horseradish peroxidase: Predictions of 2D NMR assignments and substrate binding sites. *Prot. Struc. Func. Gen.* **26**, 204–216 (1996).
28. Schuller, D.J., Ban, N., van Huysste, R.B., McPherson, A. & Poulos T.L. The crystal structure of peanut peroxidase. *Structure* **4**, 311–321 (1996).
29. Simon, P. et al. The peroxidase gene family of *Arabidopsis thaliana*. In *Plant Peroxidases: Biochemistry and Physiology* (eds Obinger, C. et al.) 179–183 (University of Geneva, Switzerland; 1996).
30. Poulos, T.L., et al., & Kraut J. The structure of cytochrome C at 2.5 Å resolution. *J. Biol. Chem.* **255**, 575–580 (1980).
31. Patterson, W.R. & Poulos, T.L. Crystal structure of recombinant pea cytosolic ascorbate peroxidase. *Biochemistry* **34**, 4331–4341 (1995).
32. Kunishima, N. et al., & Amachi, T. Crystal structure of the fungal peroxidase from *Arthromyces ramosus* at 1.9 Å resolution. *J. Mol. Biol.* **235**, 331–344 (1994).
33. Poulos, T.L., Edwards, S.L., Wariishi, H. & Gold M.H. Crystallographic refinement of lignin peroxidase at 2 Å. *J. Biol. Chem.* **268**, 4429–4440 (1993).
34. Piontek, K., Glumoff, T. & Winterhalter, K. Low pH crystal structure of glycosylated lignin peroxidase from *Phanerochaete chrysosporium* at 2.06 Å resolution. *FEBS Lett.* **315**, 119–124 (1993).
35. Smith, A.T., Du, P. & Loew, G.H. Homology modeling of horseradish peroxidase. In *Nuclear magnetic resonance of paramagnetic macromolecules* (ed. La Mar, G.N.) 75–93 (Kluwer Academic Publishers, Dordrecht, Netherlands; 1994).
36. Jones, A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these maps. *Acta Crystallogr.* **A47**, 110–119 (1991).
37. Welinder, K.G. Amino acid sequence studies of horseradish peroxidase. *Eur. J. Biochem.* **96**, 483–502 (1979).
38. Hiner, A.N. et al. comparative study of the inactivation of wild-type, recombinant and two mutant horseradish peroxidase isoenzymes C by hydrogen peroxide and m-chloroperoxybenzoic acid. *Eur. J. Biochem.* **234**, 506–512 (1995).
39. Fukuyama, K., Kunishima, N., Amada, F., Kubota, T. & Matsubara, H. Crystal structures of cyanide- and triiodide-bound forms of *Arthromyces ramosus* peroxidase at different pH values. *J. Biol. Chem.* **270**, 21884–21892 (1995).
40. Smulevich, G. et al. Characterization of recombinant horseradish peroxidase C and three site-directed mutants, F41V, F41W, and R38K, by resonance Raman spectroscopy. *Biochemistry* **33**, 7398–407 (1994).
41. Smulevich, G., English A.M., Martini, A.R. & Marzocchi M.P. Resonance Raman investigation of ferric iron in horseradish peroxidase and its aromatic donor complexes at room and low temperature. *Biochemistry* **30**, 772–779 (1991).
42. Pappa, H. S. & Cass, A.E. A step towards understanding the folding mechanism of horseradish peroxidase. Tryptophan fluorescence and circular dichroism equilibrium studies. *Eur. J. Biochem.* **212**, 227–235 (1995).
43. Shiro, Y., Kurono, M. & Morishima, I. Presence of endogenous calcium ion and its functional and structural regulation in horseradish peroxidase. *J. Biol. Chem.* **261**, 9382–9390 (1986).
44. Welinder, K.G. & Gajhed, M. Structure and Evolution of Peroxidases. Proceedings of : *Plant Peroxidases Biochemistry and Physiology*. III International Symposium July 10–14 (eds Welinder, K.G., Rasmussen, S.K., Penel, C. & Greppin, H.) 35–42 (University of Geneva, Elsinore, Denmark; 1993).
45. Veitch, N.C. et al. Structural studies by proton-NMR spectroscopy of plant peroxidase C, the wild-type recombinant protein from *Escherichia coli* and two protein variants, Phe41 Val and Arg38 Lys. *Eur. J. Biochem.* **207**, 521–531 (1992).
46. Choudhury, K. et al. Role of the proximal ligand in peroxidase catalysis. Crystallographic, kinetic, and spectral studies of cytochrome c peroxidase proximal ligand mutants. *J. Biol. Chem.* **269**, 20239–20249 (1994).
47. Mauro, J.M. et al. Tryptophan 191→phenylalanine, a proximal-site mutation in yeast cytochrome c peroxidase that strongly affects the kinetics of ferrocyanochrome c oxidation. *Biochemistry* **27**, 6243–6256 (1988).
48. Saravaj, M., Goodin, D.B., Smith, M. & Hoffman, B.M. Identification of Trp-191 as free radical site in CCP compound ES. *Science* **245**, 738–740 (1989).
49. Patterson, W.R., Poulos, T.L. & Goodin D.B. Identification of a porphyrin-cation radical in ascorbate peroxidase compound I. *Biochemistry* **34**, 4342–4345 (1995).
50. Veitch, N.C., Gao, Y., Smith, A.T. & White C.G. Identification of a critical phenylalanine residue in the horseradish peroxidase, Phe 179, by site directed mutagenesis and H-NMR: Implications for complex formation with aromatic donor molecules. *Biochemistry* (in press) (1997).
51. Kjalke, M. et al. Comparison of structure and activities of peroxidases from *Coprinus cinereus*, *Coprinus macrorhizus* and *Arthromyces ramosus*. *Biochim. Biophys. Acta.* **1120**, 248–256 (1992).
52. Braithwaite, A. Unit cell dimensions of crystalline horseradish peroxidase. *J. Mol. Biol.* **106**, 229–230 (1976).
53. Green, B.N. & Olivier, R.W.A. The study of intact proteins and glycoproteins by electrospray m.s. *Biochem. Soc. Trans.* **19**, 929–935 (1991).
54. Henriksen A. et al. Preliminary X-ray Diffraction Studies of Recombinant Peroxidase Isoenzyme C. *Acta. Crystallogr.* **D51**, 121–123 (1995).
55. Otwinowski, Z. Oscillation data reduction program. In *Proceedings of the CCP4 study weekend: Data collection and processing 29–30 January 1993* (eds Sawyer, L., Isaacs, N. & Bailey, S.) 56–62 (SERC Daresbury Laboratory, England; 1993).
56. Navaza, J. AMoRe: an automated package for molecular replacement. *Acta Crystallogr.* **A50**, 157–163 (1994).
57. Matthews, B.W. Solvent content of proteins. *J. Mol. Biol.* **33**, 491–497 (1968).
58. CCP4, 1994 Collaborative Computing Project, number 4 The CCP4 suite: programs for protein crystallography. *Acta Crystallogr.* **D50**, 760–763 (1994).
59. Schuller, D.J. MAGICSQUASH: more versatile non-crystallographic averaging with multiple constraints. *Acta Crystallogr.* **D52**, 425–434 (1996).
60. Kleywegt, G.J. & Jones, T.A. Efficient rebuilding of protein structures. *Acta Crystallogr.* **D52**, 829–841 (1996).
61. Bringer, A.T. X-PLOR Version 3.1. A system for X-ray Crystallography and NMR (Yale University Press, New Haven, Connecticut, USA; 1993).
62. Engh, R.A. & Huber, R. Accurate bond and angle parameters for X-ray protein structure refinement. *Acta. Crystallogr.* **A47**, 392–400 (1991).
63. Barton, G.J. Protein multiple sequence alignment and flexible pattern matching. *Meth. Enz.* **183**, 403–428 (1990).
64. Evans, S.V. SETOR: Hardware-lighted three dimensional solid model representations of macromolecules. *J. Mol. Graphics* **11**, 134–138 (1993).
65. Russell, R.B. & Barton, G.J. Multiple protein sequence alignment from tertiary structure comparison. *Proteins* **14**, 309–323 (1992).
66. Barton, G.J. ALSCRIPT: a tool to format multiple sequence alignments. *Protein Eng.* **6**, 37–40 (1993).
67. Humphrey, W., Dalke, A. & Schulten, K. VMD - Visual Molecular Dynamics. *J. Molec. Graphics* **14**, 33–38 (1996).