Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Impact of the physical and chemical environment on the molecular structure of *Coprinus cinereus* peroxidase

Karen Houborg,[a] Pernille Harris,[a] Jens Petersen,[a] Paul Rowland,[a] Jens-Christian Navarro Poulsen,[a] Palle Schneider,[b] Jesper Vind[b] and Sine Larsen[a]*

[a]Centre for Crystallographic Studies, Department of Chemistry, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen, Denmark, and [b]Novozymes A/S, Smørmosevej 25, DK-2880 Bagsvaerd, Denmark

Correspondence e-mail: sine@ccs.ki.ku.dk

Received 27 October 2002
Accepted 24 March 2003

PDB References: CiP-RT-I, 1h3j, r1h3jsf; CiP-RT-II, 1lyk, r1lyksf; CiPON-RT, 1ly9, r1ly9sf; CiPON-LT, 1lyc, r1lycsf.

The structure of the peroxidase from *Coprinus cinereus* (CiP) has been determined in three different space groups and crystalline environments. Two of these are of the recombinant glycosylated form (rCiP), which crystallized in space groups $P2_12_12_1$ and $C2$. The third crystal form was obtained from a variant of CiP in which the glycosylation sites have been removed (rCiPON). It crystallizes in space group $P2_1$ with $\beta \simeq 90^\circ$; the structure was determined from room-temperature data and low-temperature data obtained from twinned crystals. Two independent molecules of CiP related by non-crystallographic symmetry are contained in the three crystal forms. The packing in the two structures of the glycosylated form of rCiP is closely related, but differs from the packing in the unglycosylated rCiPON. A database search based on small-molecule porphinato iron (III) complexes has been performed and related to observations of the spin states and coordination numbers of the iron ion. The room-temperature structures of CiP and one structure of the almost identical peroxidase from *Arthromyces ramosus* (ARP) have been used to identify 66 conserved water molecules and to assign a structural role to most of them.

1. Introduction

Peroxidases (EC 1.11.1.7) constitute an important group of enzymes that catalyse the oxidation of a variety of substrates by hydrogen peroxide. The exact biological role of the enzymes is not fully understood, but it has been suggested that they protect against the accumulation of toxic peroxides (Everse *et al.*, 1991). The substrate does not react with the native enzyme but with its oxidized intermediates through a three-step cyclic process. In this way, the two-electron oxidation process is converted into two one-electron processes.

Peroxidases from different organisms show a great variation with respect to sequence and the nature, location and function of prosthetic groups (Everse *et al.*, 1991). Vanadium-containing peroxidases, which have been found in algae and certain fungi, and myelo peroxidases, found in mammalian tissues, are involved in the oxidation of halides. Plant peroxidases, which play an integral role in cell-wall biosynthesis, are characterized by having protoporphyrin IX as the prosthetic group like the myelo peroxidases (Everse *et al.*, 1991). The plant peroxidases have been found to be able to oxidize a variety of different aromatic substrates, which makes them potential candidates for applications in industrial processes.

Based on sequence alignment, the superfamily of plant peroxidases has been divided into three classes (Welinder, 1992*b*). Most of the class I peroxidases are of prokaryotic origin. The extracellular fungal peroxidases belong to class II

and the classical secretory plant peroxidases to class III. Within the classes, there is a sequence identity of up to 40–45%; the sequence identity is less than 20% between peroxidases from different classes.

The fungal peroxidase from *Coprinus cinereus* (CiP), belonging to class II, has been shown (Kjalke *et al.*, 1992) to be almost identical to two peroxidases from the fungi *C. macrorhizus* (CMP) and *Arthromyces ramosus* (ARP) (Welinder, 1992*a*,*b*). ARP contains one more amino acid in the N-terminal part of the sequence than CiP and CMP (Kunishima *et al.*, 1994); otherwise, the three enzymes appear to differ only in their degree of glycosylation (Baunsgaard *et al.*, 1993). The structure of recombinant CiP has previously been determined in the space group $P2_12_12_1$ from room-temperature diffraction data to 2.6 Å resolution (Petersen *et al.*, 1994). Here, we present two 'room-temperature' X-ray crystal structures of CiP (rCiP) at higher resolution, CiP-RT-I (PDB code 1h3j) and CiP-RT-II (PDB code 1lyk), the latter crystallizing in a different space group, and two X-ray crystal structures of the unglycosylated mutant enzyme (N142S-T331A-S338A; rCiPON), one determined at 'room temperature', CiPON-RT (PDB code 1ly9), and the other determined at 100 K, CiPON-LT (PDB code 1lyc).

We have used the structural data in an analysis of the variation of the structures of CiP with temperature and crystal packing. The availability of structures of CiP determined under very different experimental conditions has furthermore enabled us to identify and analyse the role of the structurally conserved water molecules.

Table 1
Crystallization conditions and data collection.

(*a*) Crystallization conditions and data collection for CiP-RT-I, CiP-RT-II, CiPON-RT and CiPON-LT.

Crystal form	CiP-RT-I	CiP-RT-II	CiPON-RT	CiPON-LT
Temperature (K)	274	288	290	100
X-ray source	R-AXIS II	R-AXIS II	R-AXIS II	Merged data
Crystallization conditions	18%(*w*/*v*) PEG 6000, 0.35 *M* $MgCl_2$	18%(*w*/*v*) PEG 6000, 0.35 *M* $MgCl_2$, 10 m*M* KCN	1.8 *M* $NH_4H_2PO_4$	1.5–1.6 *M* $NH_4H_2PO_4$
Protein concentration (mg ml^{-1})	20–25	20–25	17	17–21
pH	7.0	?	4.0–4.5	4.0–4.5
Space group	$P2_12_12_1$	$C2$	$P2_1$	$P2_1$ (twinned)
Resolution† (Å)	8.0–2.0 (2.11–2.0)	25.0–2.0 (2.03–2.00)	29.0–2.0 (2.03–2.00)	14.89–1.57 (1.60–1.57)
No. measured reflections	185626	104109	73087	155437
No. unique reflections	50198	41462	40199	46580
R_{merge}†	0.075	0.085	0.071	0.139
Completeness (%)	100 (100)	83.3 (72.9)	90.9 (74.2)	99.0 (96.0)
Unit-cell parameters				
a (Å)	127.4	77.7	77.60 (5)	
b (Å)	75.4	128.41	116.9 (2)	
c (Å)	76.7	75.48	36.80 (5)	
β (°)		93.2	90.7 (2)	
Mosaicity (°)	0.35	0.25	0.5	
$I/\sigma(I)$†	7.2 (3.3)	12.0 (1.3)	14.2 (3.3)	18.7 (3.3)

(*b*) Individual data collections for CiPON-LT.

X-ray source	BW7B EMBL	R-AXIS II	MAR 345
Cryoprotectant	24% glycerol	24% glycerol	30% sucrose
Resolution (Å)	35.0–1.57	35.0–2.05	30.0–1.8
No. measured reflections	192375	87173	155261
No. unique reflections	76875	37625	52771
R_{merge}	0.084	0.085	0.073
Completeness (%)	85.4	96.2	88.2
Unit-cell parameters			
a (Å)	76.70 (17)	76.91	77.29
b (Å)	113.59 (7)	114.96	115.057
c (Å)	36.36 (3)	36.24	36.97
Mosaicity (°)	0.9	0.6–1.6	0.9

† Values in parentheses are for the highest resolution shell.

2. Experimental

2.1. Crystallization

2.1.1. Glycosylated protein. rCiP was expressed in *Aspergillus oryzae*, purified and crystallized as described previously (Petersen *et al.*, 1993). The crystal used for the data collection reported here was grown with 18%(*w*/*v*) PEG 6K and 0.35 *M* $MgCl_2$ buffered at pH 7.0 with 0.1 *M* HEPES. This structure, determined at 274 K, will be referred to as CiP-RT-I (see Table 1).

In an attempt to obtain a cyanide complex of rCiP, hanging drops were set up with a reservoir solution of 18%(*w*/*v*) PEG 6K, 0.35 *M* $MgCl_2$ buffered at pH 7.0 with 0.1 *M* HEPES and 10 m*M* KCN. The structure determined at 288 K from crystals which appeared after eight months is referred to as CiP-RT-II (see Table 1). The subsequent structure determination showed no evidence for cyanide in the electron density close to the iron ion, which indicates that the pH in the drop had changed during the long crystallization period.

2.1.2. Unglycosylated protein. The mutant enzyme (N142S-T331A-S338A; rCiPON) with all O- and N-glycosylation sites removed was cloned and expressed in *A. oryzae* and purified as previously described (Tams, 1994). Crystals appeared in condition 48 (2.0 *M* $NH_4H_2PO_4$, 0.1 *M* Tris–HCl pH 8.5) from Crystal Screen I from Hampton Research (Cudney *et al.*, 1994; Jancarik & Kim, 1991) within 48 h. To obtain large single crystals reproducably, the drops were kept at 277 K for the first 24 h and then moved to room temperature. The structures of rCiPON from crystals grown this way are denoted CiPON-RT (from data collected at 290 K) and CiPON-LT (from data collected at 100 K) (see Table 1).

2.2. Data collection, structure solution and refinement of the data of the three 'room-temperature' structures

The crystals were mounted in glass capillaries and data were collected on a Rigaku R-AXIS II imaging-plate system equipped with a Rigaku RU-200 rotating-anode generator. Cu $K\alpha$ radiation was obtained from a graphite monochromator. A Molecular Structure Corporation cryosystem was employed to keep the temperature of the crystals at the values shown in Table 1. Reflections were indexed and integrated using *DENZO* (Otwinowski & Minor, 1997). For CiP-RT-I and CiP-RT-II, scaling and averaging were performed using the programs *AGROVATA* and *ROTAVATA* (Collaborative Computational Project, Number 4, 1994). For CiPON-RT, the *HKL* suite was employed (Otwinowski & Minor, 1997). A summary of the results of the data reductions is presented in Table 1.

The structure determined from the 2.6 Å data (Petersen *et al.*, 1994) was used as a starting model for the refinement of CiP-RT-I with *X-PLOR* (Brünger, 1992). The two molecules were treated as independent entities during refinement. The results of the refinement of CiP-RT-I are listed in Table 2. Though the data extends to higher resolution than that of the earlier structure determination (Petersen *et al.*, 1994), it was still impossible to locate the first seven residues in the electron density; this disorder could arise from the presence of four glycines (residues 4–7).

The structures of CiP-RT-II and CiPON-RT were determined by molecular replacement using *AMoRe* (Navaza & Saludjian, 1997). One molecule from CiP-RT-I was used as the search unit. For both crystal forms a solution was obtained with two molecules in the asymmetric unit, a correlation coefficient larger than 70% and an R value of around 30.0%. CiP-RT-II was refined as described above for CiP-RT-I. The refinement results are listed in Table 2.

The initial refinement of CiPON-RT was performed with *X-PLOR* (Brünger, 1992); in the final steps, *CNS* (Brünger *et al.*, 1998) was employed. During refinement, non-crystallographic symmetry restraints were imposed on the two molecules in the asymmetric unit. The refinement statistics and the contents of the models are shown in Table 2.

2.3. Data collection, structure solution and refinement of CiPON-LT from the twinned crystal

Many attempts were made to find suitable cryoprotectants for the crystals; as a result of these trials, we obtained several data sets measured on crystals with different cryoprotectants which all showed very high mosaicity. The data sets were collected on the local R-AXIS II system, on a local MAR 345 system and on beamline BW7B at the EMBL Outstation, Hamburg (see Table 1). All data sets were processed using the *HKL* suite (Otwinowski & Minor, 1997) and had unit-cell parameters similar to those of CiPON-RT (see Table 1), but a monoclinic β angle equal to 90.0°. It was possible to merge them in the orthorhombic crystal system, but the solution obtained in space group $P2_12_12_1$ gave unclear electron-density maps and high R factors. The solution obtained in space group $P2_1$ (similar to the CiPON-RT structure) was, however, not satisfying. We realised that twinning could be the origin of these difficulties and tested the data for twinning with the crystal twinning server (Yeates, 1988).

Table 2
Refinement statistics and final models.

Structure	CiP-RT-I	CiP-RT-II	CiPON-RT	CiPON-LT
Resolution range (Å)	8.0–2.0	25.0–2.0	30.0–2.0	8.0–1.57
No. reflections	47491	41511	40191	86540
No. peptide chains	2	2	2	2
Residue range	8–343	8–343	8–343	8–343
No. non-H protein atoms	4916	4928	4918	4918
No. haem groups	2	2	2	2
No. calcium ions	4	4	4	4
No. magnesium ions	1	2	0	0
No. waters	498	475	344	342
No. *N*-acetylglucosamines	4	2	0	0
No. mannoses	2	2	0	0
R	0.186	0.179	0.178	0.223
$R_{\rm free}$	0.256	0.235	0.215	0.265
Solvent content (%)	49	52	44	44
Average B factor (Å^2)	19.5	22.6	21.2	34.0
Ramachandran plot†				
Most favoured regions (%)	91.1	91.3	90.2	89.3
Allowed regions (%)	8.9	8.7	9.8	10.7

† Ramachandran plot parameters were calculated using the program *PROCHECK* (Collaborative Computational Project, Number 4, 1994).

2.3.1. Resolving the twinning and merging the data sets. Crystals crystallizing in monoclinic space groups with $\beta \simeq 90^\circ$ may show pseudo-merohedral twinning caused by an overlap between reflections hkl of one twin domain and $\bar{h}kl$ of the other twin domain. In the case of equal amounts of the two twin components, the apparent Laue symmetry becomes *mmm*.

The test for partial twinning (Yeates, 1988) showed twin fractions between 0.41 and 0.45 for the different data sets. None of the cryocooled data sets were of very good quality, mainly because of the apparent high mosaicity caused by the twinning (β is not exactly 90°). The individual data sets suffered from incompleteness owing to overlap of reflections and saturation of the detectors at low scattering angles. We therefore decided to merge all the cryocooled data sets in the orthorhombic crystal system and used these merged data to generate a pseudo-monoclinic data set corresponding to a twin component of $\alpha = 0.5$. The merging statistics from this process are shown in Table 1.

The final refinement of this merged data set gave $R = 0.223$ and $R_{\rm free} = 0.265$ based on data in the resolution range 8.0–1.57 Å. The final refinement statistics and the contents of the model are shown in Table 2. Despite the lower temperature and higher diffraction limit of the data, this structure has high B factors and relatively few water molecules could be introduced which could be assigned to the twinned nature of the data. We have observed similar characteristics in the refinement from other twinned crystals (Harris *et al.*, 2002).

3. Results and discussion

The overall structures of CiP and ARP have previously been described thoroughly (Kunishima *et al.*, 1994; Petersen *et al.*,

1994). The haem group divides the structure into two domains: the proximal domain containing His183, which coordinates to the iron in the haem group, and the distal domain (Kunishima *et al.*, 1994; Petersen *et al.*, 1994). Each of these domains contains a structural Ca^{2+} ion. The Ca^{2+} ion in the distal domain is heptacoordinated, displaying the coordination geometry of a pentagonal bipyramid. In the proximal domain the Ca^{2+} ion is octacoordinated, but the geometry may be described as a pentagonal bipyramid in which one of the ligands (Asp201) acts as a bidentate ligand. 11 salt bridges are found in the structure in addition to one short hydrogen bond between two glutamic acids (Glu214 and Glu239) which indicates that one of the carboxylate groups is protonated. All three crystal forms contain two monomers per asymmetric unit, with the largest (0.84 Å) r.m.s. difference in main-chain atomic positions being between monomer *A* in CiP-RT-I and monomer *A* in CiPON-RT.

3.1. Crystal packing

The three 'room-temperature' structures were determined from crystals which diffracted to similar resolution (2.0 Å). Analysis of the intermolecular interactions revealed some differences in the crystal packing. The crystals of the unglycosylated enzyme rCiPON has twice as many enzyme–enzyme contacts as the crystal forms of the glycosylated enzyme, CiP-RT-I and CiP-RT-II. This leads to a more dense packing of the unglycosylated enzyme. However, the similar translation period (~77 Å) in the different crystal forms may indicate that similar features influence the crystal packing.

In CiP-RT-I and CiP-RT-II, the two independent CiP molecules are related by a non-crystallographic twofold axis parallel to the crystallographic *c* and *a* axis, respectively; in CiPON-RT they are related by a non-crystallographic twofold screw axis parallel to the crystallographic *c* axis. The non-crystallographic twofold symmetry in CiP-RT-I and CiP-RT-II generates a virtually identical dimeric unit, which is shown in Fig. 1. The interactions connecting the two subunits are hydrogen bonds from Ser150 OG to Gly156 O and Asn157 ND2. The packing viewed down the non-crystallographic twofold axes (the *c* axis in CiP-RT-I or the *a* axis in CiP-RT-II) shows that these 'dimers' are organized in structurally similar layers (Fig. 2). The dimers in the layers are connected by hydrogen bonds between Thr10 N(O) and Thr331 O(N), between Ser8 O and Gly333 N and between Ala305 and Ser20 OG. A hexaaquamagnesium ion on the non-crystallographic twofold axis functions as an additional link through its interactions with Asp316 from both independent molecules, as illustrated in Fig. 2. The only significant difference between the packing of CiP-RT-I and CiP-RT-II is seen in the stacking of these layers of 'dimers', which leads to the

Figure 1
The packing of molecules *A* (red) and *B* (blue) in CiP-RT-I and CiP-RT-II. The axes correspond to the CIP-RT-II structure. Residues involved in intermolecular hydrogen bonds are shown in green. The figure was prepared using the program *MOLSCRIPT* (Kraulis, 1991).

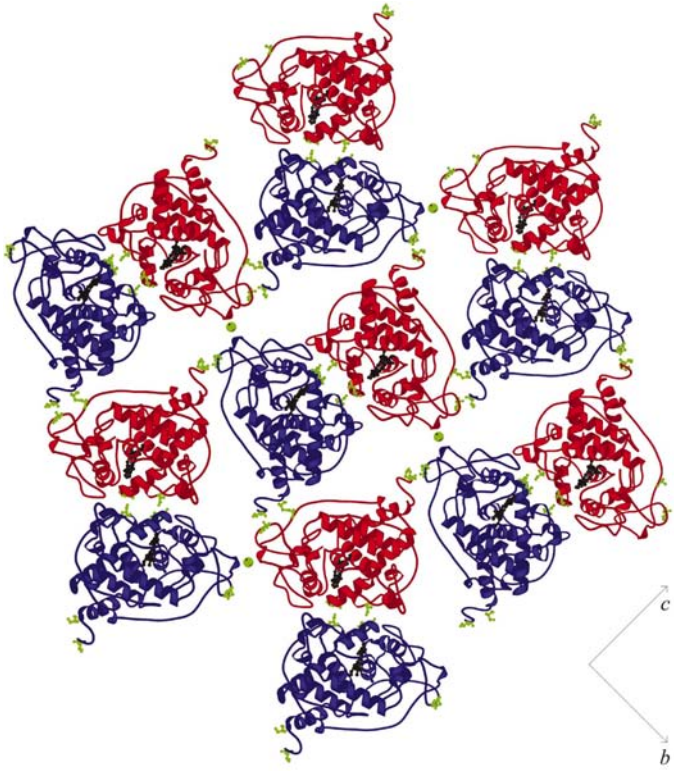

Figure 2
The layer of molecules in CiP-RT-I and CiP-RT-II. The axes correspond to the CIP-RT-II structure. Residues and the magnesium ion involved in intermolecular hydrogen bonds are shown in green. The figure was prepared using the program *MOLSCRIPT* (Kraulis, 1991).

different overall symmetry. The inter-layer connections are created by a few hydrogen bonds. In CiP-RT-I the inter-layer connection is between Asp97 OD2 (*A*) and Glu324 OE2 (*B*). The interlayer connection in CiP-RT-II is constituted of hydrogen bonds between Asn287 ND2 (*B*) and Asn35 OD1 (*A*) and the interaction of one additional hexaaquamagnesium ion with Asp174 (*B*) from two molecules related by crystallographic symmetry.

The two independent molecules in rCiPON are connected into a chain by the non-crystallographic twofold screw axis, as shown in Fig. 3. The hydrogen bonds involved in this arrangement are from Gly156 N (*A*/*B*) to Ala66 O (*B*/*A*) and from Gly190 O (*A*/*B*) to Gln69 NE2 (*B*/*A*). The crystal packing is also influenced by hydrogen bonds connecting the molecules related by translational symmetry along the short *c* axis: Asp174 OD1 (*A*/*B*) to Ser193 O (*A*/*B*), Ser323 OG (*A*/*B*) to Pro223 O (*A*/*B*) and Glu324 OE1 (*A*/*B*) to Pro223 O (*A*/*B*) and Gly238 N (*A*/*B*). It should be noted that these latter hydrogen bonds would not be possible if a mannose was attached to Ser338 as in the glycosylated protein, as the mannose would clash with the neighbouring molecule. The chains of molecules viewed down the *c* axis (Fig. 4) create a 'checkerboard' pattern similar to that in Fig. 2. The chains are connected by hydrogen bonds from Ser8 O (*A*/*B*) to Gly333 N (*B*/*A*), from Thr10 N(O) (*A*/*B*) to Ala331 O(N) (*B*/*A*), from Pro12 O (*A*/*B*) to Glu328 OE2 (*B*/*A*) and from Asp315 OD1 (*A*/*B*) and OD2 (*A*/*B*) to Thr221 OG1 (*B*/*A*).

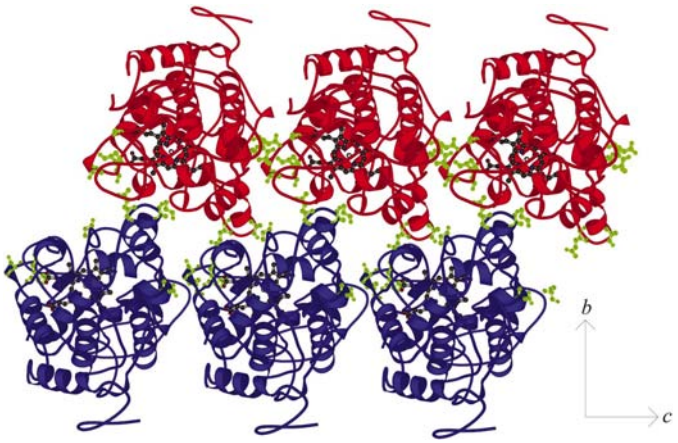

Figure 3
The chains of molecules *A* (red) and *B* (blue) in CiPON-RT. Residues involved in intermolecular hydrogen bonds are shown in green. The figure was prepared using the program *MOLSCRIPT* (Kraulis, 1991).

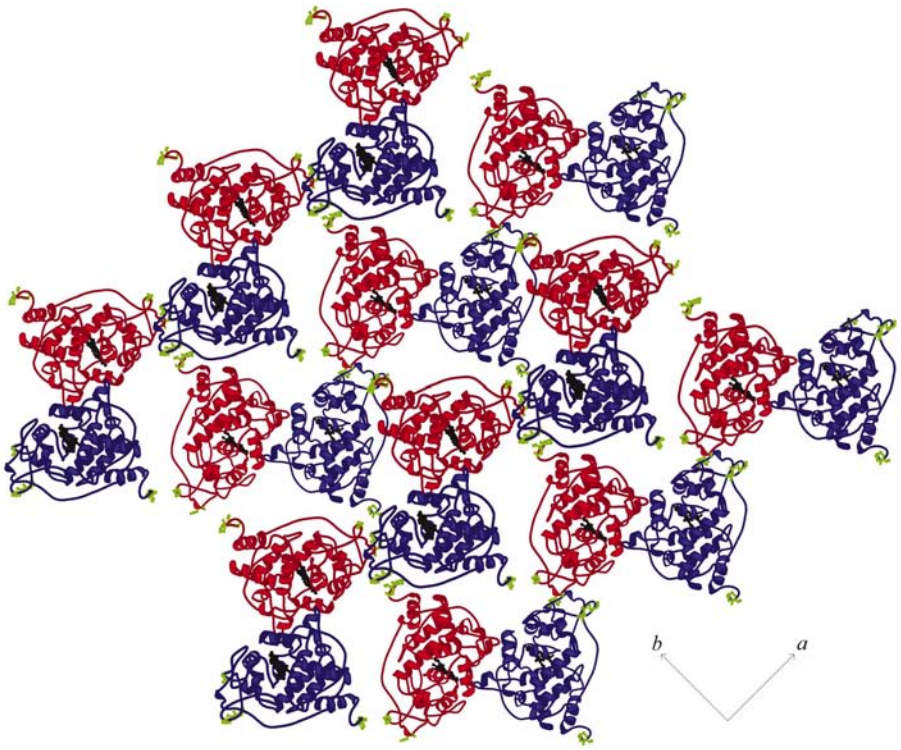

Figure 4
A view down the chains in CiPON-RT. Residues involved in intermolecular hydrogen bonds are shown in green. The figure was prepared using the program *MOLSCRIPT* (Kraulis, 1991).

3.2. Haem environment

3.2.1. The effect of cryocooling. Cytochrome *c* peroxidase has a pentacoordinated high-spin haem iron at room temperature, which has been shown by spectroscopic measurements to change to hexacoordinated low spin at low temperature (Smulevich *et al.*, 1989). In the presence of glycerol, it was found to also be pentacoordinated high spin at low temperatures (Yonetani & Anni, 1987).

The haem irons at 100 K observed in CiPON-LT and in a thermostable mutant enzyme of CiP (Houborg *et al.*, 2003) are found to be pentacoordinated, whereas the haem irons at 293 K (CiP-RT-I, CiP-RT-II and CiPON-RT) are loosely hexacoordinated, with an average Fe—O distance of 2.59 Å for the sixth ligand (water molecules 1011 or 2011). The crystallization conditions are identical for CiPON-RT and CiPON-LT, but it is not possible to conclude whether the change in coordination geometry is a consequence of the difference in temperature or of the addition of cryoprotectant (2% glycerol or 30% sucrose) (see Table 1).

3.2.2. The stereochemical relationship in iron porphyrins. X-ray studies of the structures of several haem-containing proteins made by Perutz (1970) suggested that the position of the iron relative to the haem plane depended on its spin state and coordination number. This result was later elaborated based on the X-ray crystal structures of iron porphyrins (Scheidt & Reed, 1981; Scheidt *et al.*, 1987).

We have conducted a search of the Cambridge Structural Database (Allen & Kennard, 1993) for structures of iron

Table 3
The haem environment for CiP.

Structure	PDB code	Temperature	Resolution (Å)	pH	Coordination No.†	Spin state†	Fe—ligand, distal (Å)	Fe—N_p‡ (Å)	Fe—Ct_N§ (Å)
CiP-RT-I	1h3j	274	2.0	7.0	5¶	High¶	2.45, wat	2.02 (5)	0.10
							2.33, wat	1.99 (4)	0.06
CiP-RT-II	1lyk	288	2.0	?			2.77, wat	2.03 (7)	0.01
							2.66, wat	2.03 (4)	0.04
CiPON-RT	1ly9	293	2.0	4.5	5††	High††	2.65, wat	2.02 (6)	0.05
							2.53, wat	2.05 (4)	0.03
CiPON-LT	1lyc	100	1.57	4.5				2.12 (4)	0.14
								2.13 (5)	0.21

† Determined by spectroscopic methods ‡ N_p is the porphinato N atom. § Ct_N is the centre of the best plane of the four N_ps. ¶ Smulevich *et al.* (1994). †† Smulevich *et al.* (1996).

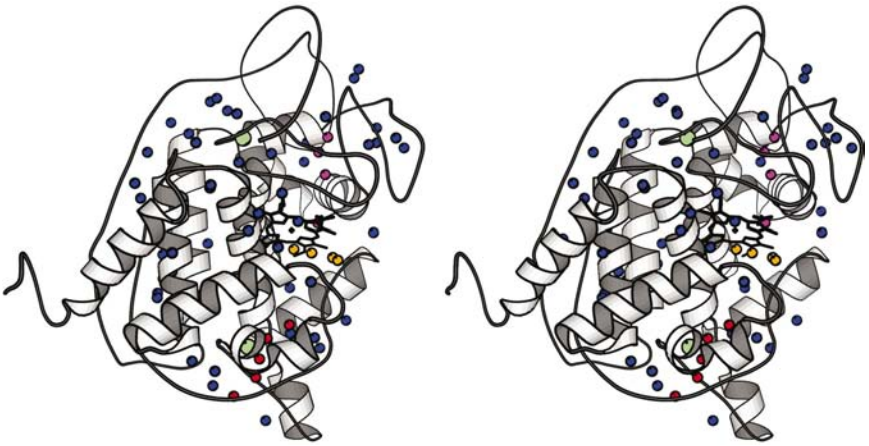

Figure 5
A structural representation of CiPON-RT with the conserved water molecules in red (1028, 1027, 1009, 1023, 1034), pink (1001, 1006, 1025, 1017), orange (1011, 1029, 1066, 1056) and blue, the two calcium ions in green and the haem group in black. The figure was prepared using the program *MOLSCRIPT* (Kraulis, 1991).

porphyrins for which the spin states have been determined by other experimental methods. This search[1] showed that the distances from the iron to the centre of the plane of the four N atoms (the Fe—Ct_N distance) are larger in the pentacoordinated systems than in the hexacoordinated system, being between 0.13 and 0.33 Å for the low-, intermediate-and mixed intermediate-spin states and between 0.43 and 0.46 Å for the high-spin state. The hexacoordinated iron (III) porphinato complexes have the iron ion virtually in the plane (0–0.18 Å) of the porphyrine ring system for all the spin states.

In Table 3, we have listed the haem geometries of the structures of CiP and the associated spin state where it has been measured independently. The Fe—Ct_N distances in CiP are much smaller in the 'room-temperature' structures, 0.03 and 0.05 Å in CiPON-RT, 0.06 and 0.10 Å in CiP-RT-I and 0.01 and 0.04 Å in CiP-RT-II, compared with CiPON-LT, where the Fe—Ct_N distances are 0.14 and 0.21 Å. This indicates that the loosely bound water molecule affects the position of the iron ion in the haem group, so from a geometrical criterion we would consider CiP-RT-I and CiPON-RT as hexacoordinated. It should be noted that this conclusion is at variance with spectroscopic measurements, which indicate that CiP-RT-I and CiPON-RT have a high-spin pentacoordinated haem group (Smulevich *et al.*, 1994, 1996).

3.3. Structurally conserved water molecules

The water molecules located in an X-ray crystal structure of a protein have various origins. Some of the bound water molecules may reflect the crystallization conditions, *e.g.* the pH and precipitant, whereas other water molecules can be considered as an integrated part of the protein and can be regarded as structural water. Normally, it is difficult to identify structural water molecules from the bulk water structure found in and around the protein. The structures of CiP determined at 'room temperature', where crystals grew under three different conditions with two molecules in the asymmetric unit, and the single structure of ARP (PDB code 1arp; Kunishima *et al.*, 1994) have enabled us to identify the conserved water molecules. For this analysis, we have used the program *WATNCS* (Collaborative Computational Project, Number 4, 1994). Water molecules were considered to be invariant if their positions in the different molecules were within 1 Å of each other[2]. The positions of the 66 structurally invariant water molecules in the CiP structure identified this way are shown in Fig. 5. Though they are distributed over the whole molecule, there is a significant number near the surface and a high concentration around the haem group and the distal calcium ion.

A more detailed view of the water molecules near the haem group is shown in Fig. 6. Water molecule 1011 has been reported to have distances to the iron in the range 2.33–2.96 Å and is hydrogen bonded to the catalytically important His55 and to another water molecule (1029). Water molecule 1029 makes hydrogen bonds to Arg51 (NH and NE) and to another water molecule (1066) which is connected to water molecule 1056. This arrangement constitutes a chain of conserved water molecules (1011, 1029, 1066 and 1056) shown in orange in Figs. 5 and 6. It stretches from the active site to the surface of the molecule. Water molecule 1014, which serves as a

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: HV0003). Services for accessing this material are described at the back of the journal.

[2] *WATNCS* is constructed so that only waters from NCS-related molecules of the same structure can be compared, but we 'cheated' the program by translating the coordinates of one structure by 1000 Å of the next structure by 2000 Å *etc*.

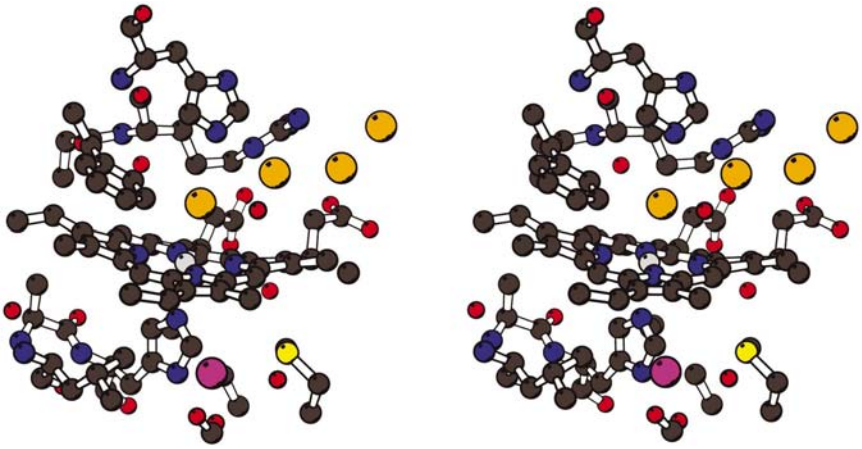

Figure 6
The water molecules found in the region around the haem group in the CiP and ARP structures. The figure was prepared using the program *MOLSCRIPT* (Kraulis, 1991).

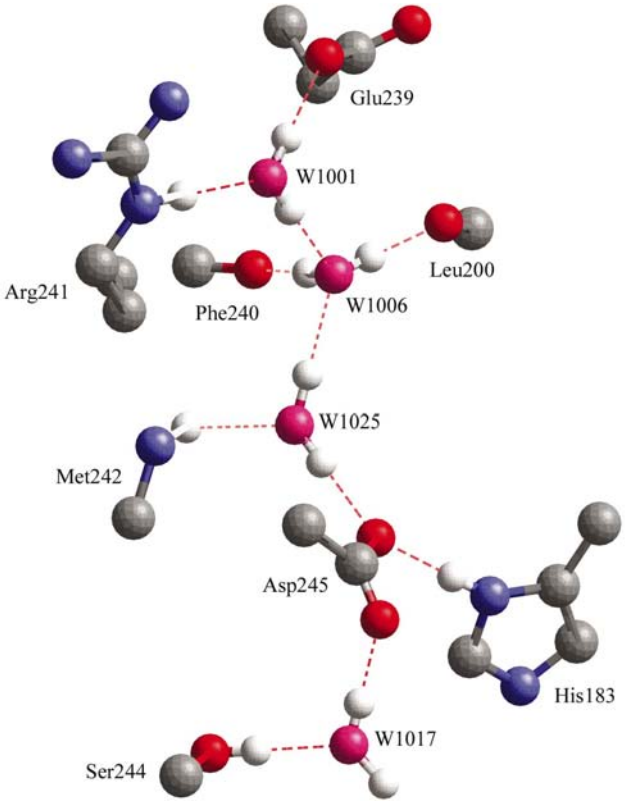

Figure 7
The hydrogen-bonding system that involves the four conserved water molecules close to the proximal His183. The figure was prepared using the program *CERIUS*^{2} (Molecular Simulations Inc., 1997).

hydrogen-bond bridge between the two side chains of Arg51 and Lys48, is also present in cytochrome *c* peroxidase (Kunishima *et al.*, 1994). The water molecules at the active site have been compared with complexes of ARP with benzhydroxamic acid (PDB code 1hsr; Itakura *et al.*, 1997), with salicylhydroxamic acid (PDB code 1ck6; Tsukamoto *et al.*, 1999) and with hydroxylamine (PDB code 1c8i; Wariishi *et al.*, 2000). Water 1011, which is coordinated to the iron ion, is missing in the hydroxylamine complex and water molecules 1029 and 1066 are replaced with the aromatic inhibitors in the other two complexes. In the cyanide (PDB codes 1aru, 1arv and 1arw; Fukuyama *et al.*, 1995) and the iodide ARP complexes (PDB codes 1arx and 1ary; Fukuyama *et al.*, 1995) water 1011 is missing and waters 1029 and 1066 have been replaced by iodine in the iodide complexes.

On the proximal side of the haem group, water molecule 1017 is hydrogen bonded to Ser244 O and Asp245 OD. This water molecule represents one end of a structurally essential chain of water molecules close to the proximal His and involving three other water molecules, 1025, 1006 and 1001; it is shown in pink in Fig. 5. These water molecules are so well integrated into the structure that it is possible to assign positions to the protons as shown in Fig. 7.

The Ca^{2+} ion in the distal domain is heptacoordinated and two of the oxygen ligands are the conserved water molecules 1009 and 1027. Additional structurally conserved water molecules are found close to the distal Ca^{2+}-binding site. A chain of water molecules (1028-1027-1009-1023-1034; drawn in red in Fig. 5) with water 1028 hydrogen bonded to His55 O goes through the molecule from the active site to the distal Ca^{2+}-binding site and ends at the surface.

4. Discussion and conclusions

X-ray crystal structures provide a picture of the protein in an environment far from the physiological conditions. The above analysis for CiP shows that it is important to take the physical and chemical environment into account in the analysis of the crystal structure. It also illustrates how the presence of metal (magnesium) ions can influence the crystal packing.

Though previous spectroscopic measurements had suggested that CiPON-RT and CiP-RT-I contain a high-spin pentacoordinated haem iron, the present X-ray crystal structures indicate that the haem iron is loosely hexacoordinated. The haem iron is pentacoordinated in the low-temperature structure CiPON-LT, but it cannot be concluded whether this is a consequence of the addition of cryoprotectant or of the lowering of the temperature.

Identification of the structurally conserved water molecules in the CiP structure revealed four water molecules adjacent to the haem group in the active-site pocket at the distal site. These water molecules, which form a chain from the entrance of the pocket to the position of the iron, can be replaced by substrate or product. It is more difficult to assign the role of the two other chains of structurally conserved water molecules. It should be noted that introduction of water molecules into the protein may render some dynamic flexibility in the molecule which might be necessary for catalysis, as illustrated in a recent study of carboxypeptidase A (Jensen *et al.*, 2002).

We are grateful to the EMBL Outstation, Hamburg, Germany for providing beam time for the project, and to DANSYNC and the EU for contribution to travel expenses under the 'Access to Research Infrastructures' program. Flemming Hansen are thanked for technical support and Dr Leila Lo Leggio is thanked for discussions on the twinning problem. The project has been funded by the Danish Research Foundation and through a Novo Nordisk Scholarship to Karen Houborg.

References

Allen, F. H. & Kennard, O. (1993). *Chem. Des. Autom. News*, **8**, 31–37.

Baunsgaard, L., Dalbøge, H., Houen, G., Rasmussen, E. M. & Welinder, K. G. (1993). *Eur. J. Biochem.* **213**, 605–611.

Brünger, A. T. (1992). *X-PLOR Version 3.1. A System for X-ray Crystallography and NMR.* Yale University Press.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Cudney, R., Patel, S., Weisgraber, K., Newhouse, Y. & McPherson, A. (1994). *Acta Cryst.* D**50**, 414–423.

Everse, J., Everse, K. E. & Grisham, M. B. (1991). Editors. *Peroxidases in Chemistry and Biology*, Vols. 1 and 2. Boca Raton, FL, USA: CRC Press.

Fukuyama, K., Kunishima, N., Amada, F., Kobota, T. & Matsubara, H. (1995). *J. Biol. Chem.* **270**, 21884–21892.

Harris, P., Poulsen, J.-C. N., Jensen, K. F. & Larsen, S. (2002). *J. Mol. Biol.* **318**, 1019–1029.

Houborg, K., Harris, P., Poulsen, J.-C. N., Svendsen, P. S. A. & Larsen, S. (2003). *Acta Cryst.* D**59**, 997–1003.

Itakura, H., Oda, Y. & Fukuyama, K. (1997). *FEBS Lett.* **412**, 107–110.

Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.

Jensen, A. F., Bukrinsky, J. T., Bjerrum, M. J. & Larsen, S. (2002). *J. Biol. Inorg. Chem.* **7**, 490–499.

Kjalke, M., Andersen, M. B., Schneider, P., Christiansen, B., Schülein, M. & Welinder, K. G. (1992). *Biochim. Biophys. Acta*, **1120**, 248–256.

Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.

Kunishima, N., Fukuyama, K., Matsubara, H., Hatanaka, H., Shibano, Y. & Amachi, T. (1994). *J. Mol. Biol.* **235**, 331–344.

Molecular Simulations Inc. (1997). *CERIUS^2 User Guide.* San Diego: Molecular Simulations Inc.

Navaza, J. & Saludjian, P. (1997). *Methods Enzymol.* **276**, 581–594.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Perutz, M. F. (1970). *Nature (London)*, **228**, 726–729.

Petersen, J. F. W., Kadziola, A. & Larsen, S. (1994). *FEBS Lett.* **339**, 291–296.

Petersen, J. F. W., Tams, J. W., Vind, J., Svensson, A., Dalbøge, H., Welinder, K. G. & Larsen, S. (1993). *J. Mol. Biol.* **232**, 989–991.

Scheidt, W. R., Geiger, D. K., Lee, Y. J., Reed, C. A. & Lang, G. (1987). *Inorg. Chem.* **26**, 1039–1045.

Scheidt, W. R. & Reed, C. A. (1981). *Chem. Rev.* **81**, 543–555.

Smulevich, G., Feis, A., Focardi, C., Tams, J. & Welinder, K. G. (1994). *Biochemistry*, **33**, 15425–15432.

Smulevich, G., Mantini, A. R., English, A. M. & Mauro, J. M. (1989). *Biochemistry*, **28**, 5058–5064.

Smulevich, G., Neri, F., Marzocchi, M. P. & Welinder, K. G. (1996). *Biochemistry*, **35**, 10576–10585.

Tams, J. W. (1994). PhD thesis, University of Copenhagen, Denmark.

Tsukamoto, H., Itakura, H., Sato, K., Fukuyama, K., Miura, S., Takahashi, S., Izekawa, H. & Hosoya, T. (1999). *Biochemistry*, **38**, 12558–12568.

Wariishi, H., Nonaka, D., Johjima, T., Nakamura, N., Naruta, Y., Kubo, S. & Fukuyama, K. (2000). *J. Biol. Chem.* **275**, 32919–32924.

Welinder, K. G. (1992*a*). *Plant Peroxidases 1980–1990, Topics and Detailed Literature on Biochemical and Physiological Aspects*, edited by T. Gaspar, C. Penel & H. Greppin, pp. 1–24. University of Geneva.

Welinder, K. G. (1992*b*). *Curr. Opin. Struct. Biol.* **2**, 388–393.

Yeates, T. O. (1988). *Acta Cryst.* A**44**, 142–144.

Yonetani, T. & Anni, H. (1987). *J. Biol. Chem.* **262**, 9547–9554.