Crystal structure of catalase HPII from *Escherichia coli*

Jerónimo Bravo[1], Núria Verdaguer[1], José Tormo[1], Christian Betzel[2], Jack Switala[3], Peter C Loewen[3]* and Ignacio Fita[1]*

[1]Departamento de Ingeniería Química, Universidad Politécnica de Cataluña, Diagonal 647, 08028 Barcelona, Spain, [2]EMBL c/o DESY, Notkestr. 85, 22603 Hamburg, Germany and [3]Department of Microbiology, University of Manitoba, Winnipeg, MB R3T 2N2, Canada

Background: Catalase is a ubiquitous enzyme present in both the prokaryotic and eukaryotic cells of aerobic organisms. It serves, in part, to protect the cell from the toxic effects of small peroxides. *Escherichia coli* produces two catalases, HPI and HPII, that are quite distinct from other catalases in physical structure and catalytic properties. HPII, studied in this work, is encoded by the *katE* gene, and has been characterized as an oligomeric, monofunctional catalase containing one *cis*-heme *d* prosthetic group per subunit of 753 residues.

Results: The crystal structure of catalase HPII from *E. coli* has been determined to 2.8 Å resolution. The asymmetric unit of the crystal contains a whole molecule, which is a tetramer with accurate 222 point group symmetry. In the model built, that includes residues 27–753 and one heme group per monomer, strict non-crystallographic symmetry has been maintained. The crystallographic agreement R-factor is 20.1% for 58477 reflections in the resolution shell 8.0–2.8 Å.

Conclusions: Despite differences in size and chemical properties, which were suggestive of a unique catalase, the deduced structure of HPII is related to the structure of catalase from *Penicillium vitale*, whose sequence is not yet known. In particular, both molecules have an additional C-terminal domain that is absent in the bovine catalase. This extra domain contains a Rossmann fold but no bound nucleotides have been detected, and its physiological role is unknown. In HPII, the heme group is modified to a heme *d* and inverted with respect to the orientation determined in all previously reported heme catalases. HPII is the largest catalase for which the structure has been determined to almost atomic resolution.

Structure 15 May 1995, **3**:491–502

Key words: bacterial catalase, heme *d*, heme proteins, molecular averaging

Introduction

Catalase (hydrogen peroxide:hydrogen peroxide oxidoreductase, EC 1.11.1.6) is an enzyme present in most aerobic prokaryotic and eukaryotic cells; it probably evolved when the earth developed an oxygenated atmosphere and organisms had to neutralize toxic oxygen-radical by-products. However, this protective role for the biological function of catalase remains unconfirmed.

The overall reaction catalyzed by the enzyme can be written as:

$$ROOH+HQOH \rightleftarrows QO+ROH+H_2O$$

where R is a hydrogen or an alkyl or acyl group and HQOH is a two-electron donor. This reaction proceeds by two steps: firstly, oxidation of the enzyme to a high-valent iron-oxo derivative, known as compound I, by a peroxide (ROOH); secondly; oxidation of the substrate HQOH to recover the free enzyme. Substrate oxidation is operationally divided into 'catalatic' activity, when Q is an oxygen atom, and 'peroxidatic' activity, when Q is some other chemical group. Catalases differ from most peroxidases in their ability to utilize hydrogen peroxide both as an electron acceptor and an electron donor (catalatic activity) yielding molecular oxygen and water in the dismutation reaction:

$$2H_2O_2 \rightleftarrows 2H_2O+O_2$$

Catalase was one of the first enzymes to be crystallized and crystals of catalases from a variety of sources have been described. Some of them have been used for high-resolution structural studies employing X-ray crystallography. The three-dimensional conformations of four heme-containing catalases have now been reported, including *Penicillium vitale* catalase (PVC) [1,2], beef liver catalase (BLC) [3,4], *Micrococcus lysodeikticus* catalase (MLC) [5] and *Proteus mirabilis* catalase (PMC) [6], although of these, only the sequences of BLC and PMC are available. All four catalases are tetrameric, with 222 molecular symmetry, and the main differences among them are found in the N- and C-terminal regions. In particular, PVC has an extra C-terminal domain of about 150 residues with a flavodoxin-like topology, but lacks the NADPH-binding site that is present in BLC [7] and PMC [6].

Escherichia coli produces two catalases, hydroperoxidase I (HPI) and HPII, which are quite distinct in physical structure and catalytic properties, both from each other and from other catalases. Catalase HPI, encoded by the *katG* gene, is a bifunctional catalase-peroxidase that contains two protoheme IX groups in a tetramer of identical subunits (MW 80000 Da) [8]. Catalase HPII, encoded by the *katE* gene, which in turn is regulated by the *katF* gene [9], has been characterized as a monofunctional catalase, with a high K_m for H_2O_2 [10], and one *cis*-heme *d* isomer per subunit (MW 84200 Da) associated in an apparently hexameric structure [10]. The larger subunit size, apparent hexameric structure, and the unusual heme *d* component, suggested that HPII is a unique catalase. However, when its primary structure was determined, striking

*Corresponding authors.

similarities between the sequences of catalases from several plant, mammalian, and fungal sources were noted [11].

In this work we report the crystallographic structure determination of the bacterial catalase HPII from *E. coli*, at 2.8 Å resolution.

Results and discussion

Overall structure and quality of the model

The HPII molecular structure determined in this work has been found to be a tetramer with accurate 222 point group symmetry. The four subunits of the molecule, comprising more than 3000 amino acids in total, are contained in the asymmetric unit of the crystal (Fig. 1). The HPII model includes residues 27–753 and one heme group per monomer. Only four well defined solvent molecules in the vicinity of the heme group have been explicitly added. Strict non-crystallographic symmetry has been maintained and, as a consequence, deviations from the local symmetry cannot be evaluated. External molecular regions involved in crystal contacts are expected to show discrepancies among the four subunits, which may contribute to the disorder in the N-terminal residues not included in the present model. For this model, the crystallographic agreement R-factor is 20.1% for 58477 reflections in the resolution shell between 8.0 Å and 2.8 Å.

The quality of the final averaged ($2F_o$–F_c) electron-density maps allows the unambiguous recognition of most amino acid side chains (Fig. 2). No extra density with the size of a nucleotide molecule, as observed for BLC [12], has been detected in the final (F_o–F_c) maps. All residues, except Ile274 and His739, for which the electron density is well defined (Fig. 2), are within, or very near, the energetically allowed regions of the Ramachandran plot (Fig. 3). Residue Ile274, with main-chain conformational angles of ϕ=53 and ψ=–42 (Fig. 2a), is structurally equivalent to Ser216 in BLC, which also showed unusual values and played a part in closing the β-barrel [4] between strands β4 and β5. Ile274 might also contribute to determining the orientation of the heme group (see below); His739 (ϕ=69, ψ=–69) has its imidazole ring forming hydrogen bonds with both the guanidinium group of Arg601 and the carbonyl oxygen of residue Ser301 (Fig. 2b).

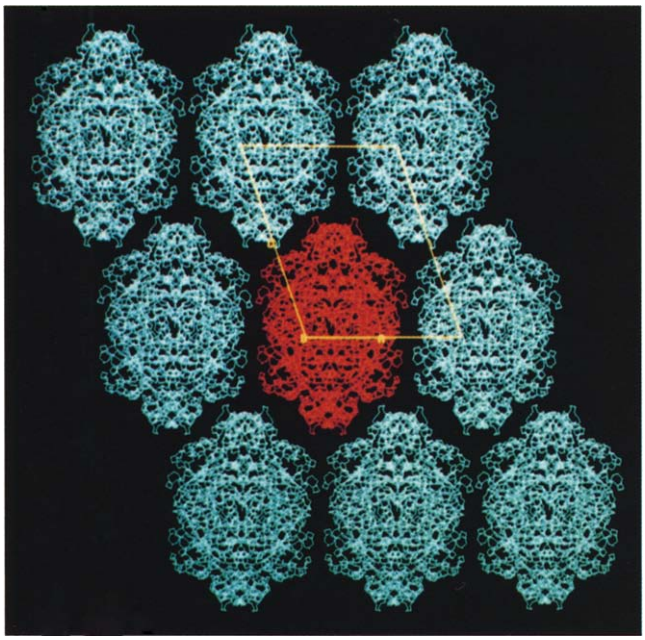

Fig. 1. View, down the b axis, of the HPII packing. Each asymmetric unit contains a whole molecule (the reference tetramer is shown in red). The unit cell contains two molecules placed in contiguous layers. The crystal solvent content is about 44% (V_m=2.22 Å^3 Da^{-1}).

The HPII tetramer is a compact molecule with extreme PQR molecular axes coordinate values of 49 Å, 39 Å and 71 Å (Fig. 4). The relative disposition of subunits in the tetramer and the domain organization found for HPII (Fig. 5) resembles those in PVC [1]. In spite of this structural homology, several attempts to use the polyglycine model of PVC (see Material and methods section) did not yield additional information with respect to BLC during the initial HPII structure determination. The most relevant point of similarity between the PVC and HPII structures is the presence in both molecules of a C-terminal domain of about 150 amino acids, that is absent in BLC, MLC and PMC. In turn, the most conspicuous difference between the two structures is the absence of a segment of about 70 amino acids from the N-terminal region of PVC. As the primary structure of PVC is not yet known, a complete comparison is precluded by the lack of certainty on the length and sequence of PVC. The superimpositions of Cα atoms from HPII with BLC and PVC give 467 and 632 equivalences for root mean square (rms) values of 1.2 Å and 1.9 Å, respectively.

The N-terminal portion of HPII (until about residue Ala79), which has no homology with the known catalase structures, contains two short α-helices, residues 52–57 and 62–71, but lacks any other discernible secondary structure besides turns. This extension of the N-terminal arm increases the contact area between subunits and closes, at least in part, the lateral entrance to the main channel described for BLC [13].

From residue 80–434, HPII presents very high levels of sequence homology (Fig. 6) and of structural homology with residues 26–375 of BLC. From HPII residue 434 (at the beginning of the 'wrapping' domain, according to the classification used for BLC [13]) to residue 561 (at the end of the α-helical domain) the amount of sequence homology with BLC is lower, but the corresponding three-dimensional organization is still very similar.

The HPII region spanning residues 562–600 is extended and exposed. It joins the α-helical domain to the C-terminal domain. As indicated above, the HPII C-terminal domain has a general shape and a relative disposition in the molecule that is similar to the 'flavodoxin-like' domain reported for PVC, although there are many differences in detail between the two structures, and even

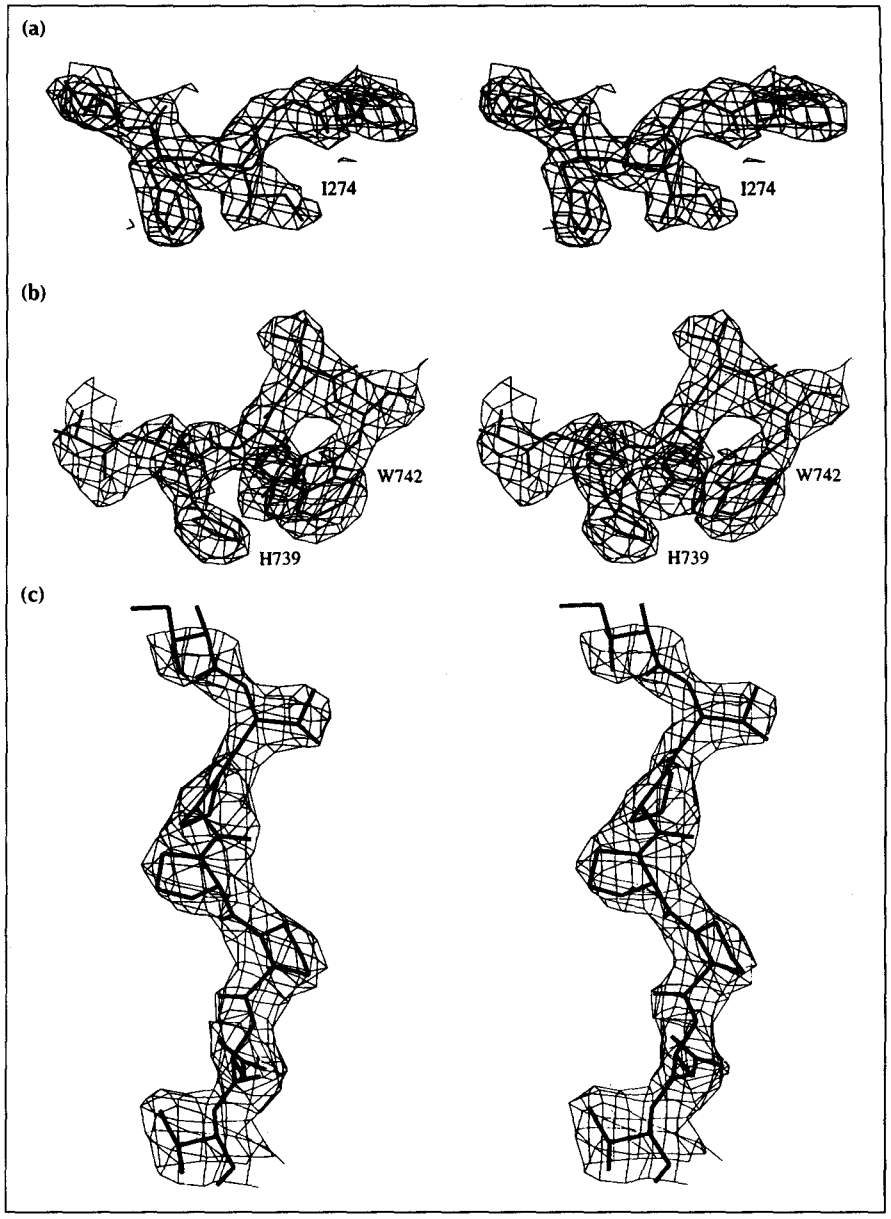

Fig. 2. Representative stereoviews of the final averaged ($2F_o$–F_c) electron-density map. Residues **(a)** Ile274 and **(b)** His739 are outside energetically favorable regions in the Ramachandran diagram (see Fig. 3). The identification of the bulky residue Trp742 (b) facilitated the tracing of the C-terminal domain. **(c)** Exposed segment in the hinge region, including residues Pro575-Pro576-Pro577.

the number and size of the secondary-structure elements differ (see below and Fig. 7). The superimposition of the HPII and PVC C-terminal domains has an rms of 2.4 Å for 125 equivalences. Comparison with the flavodoxin conformation (with reference entry 1FLV in the Brookhaven Protein Data Bank) gives an rms of 3.0 Å for 107 equivalences.

Atomic mobility, as derived from the isotropic temperature factors (Fig. 8), indicates that the residues with the lowest B-values are located, as expected, in the internal part of the tetramer. The flexibility is comparatively high in the exposed (accessible) regions and, in particular, in protein segments that participate in crystal contacts (around amino acids 60, 69, 155, 283, 345, 360 and contacts involving the C-terminal domain). Only after independent refinement will it be possible to quantify how differences among the four subunits contribute to the B-values obtained in the present averaged model.

The temperature factors increase continuously towards the N terminus. The initial residues are not visible in the averaged electron-density maps. Temperature factors in the C-terminal domain show strong fluctuations along the sequence. Non-accessible residues have values similar to those in the central parts of the molecule. Instead, exposed turns, in particular around residue 712, have higher B-values. As already suggested, it may be significant that these turns are always involved in intermolecular interactions.

The heme group and its environment

The four heme groups are well buried inside the HPII tetramer (about 20 Å from the molecular surface) in equivalent locations to those of the heme groups in BLC and PVC. The coordinates of the heme iron atom in the HPII reference subunit, referred to the PQR molecular axes are 17.0 Å, 3.5 Å and 15.2 Å, respectively. In the BLC structure, the PQR iron coordinates are coincident (within experimental error).

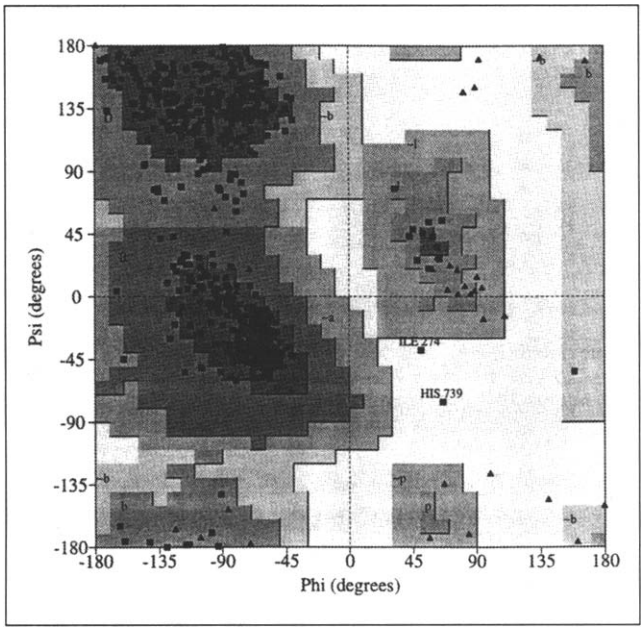

Fig. 3. Ramachandran plot for the HPII averaged model obtained with the program PROCHECK [29]. Most residues are well within the energetically stable regions which, for non-glycine residues (squares), are indicated as shaded regions in the plot. Glycine residues are represented with triangles.

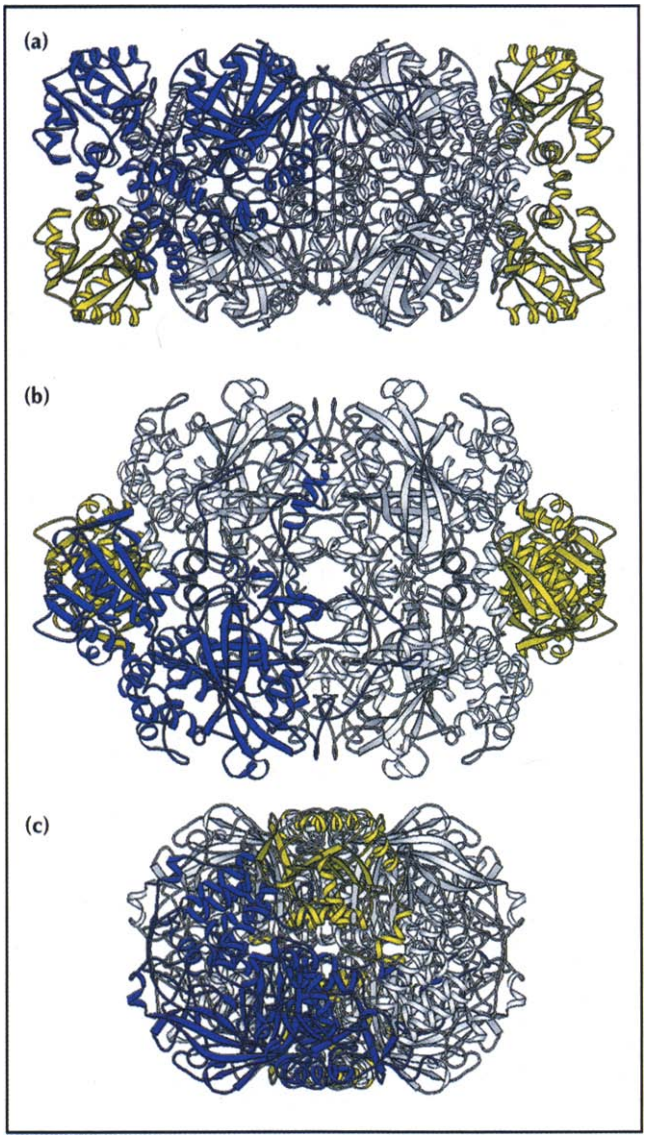

Fig. 4. Views of the HPII tetramer down the **(a)** P, **(b)** Q and **(c)** R molecular axes drawn with the program MOLSCRIPT [30]. The reference subunit is shown in blue. The C-terminal domains of the remaining subunits are depicted in yellow. The largest molecular size along the R axis is about 150 Å, see (a), (b). Even so, a channel around this R axis, with a diameter larger than a solvent molecule, crosses the whole length of the molecule (c).

The structure of the heme pocket, both on the distal and proximal sides, is well preserved. The essential residues His74, Ser113, Asn147, Arg353 and Tyr357 in BLC can be well superimposed on residues His128, Ser167, Asn201, Arg411 and Tyr415 in HPII, as already suggested by the sequence alignment of the two molecules (Fig. 6). Despite this high level of structural homology of the heme pocket, there are significant peculiarities in the heme group itself in HPII. The heme group lies in an inverted orientation with respect to the one reported for BLC. This results in the relative positions of pyrrole rings I and IV interchanging with rings II and III, respectively, and in a change in the relative positions of the vinyl and methyl substituents on rings I and II (Fig. 9). The differential contacts of these methyl and vinyl groups in one heme orientation or in the reversed orientation, are those that, in principle, should determine the orientation of the heme inside the pocket. Residues governing heme orientation appear to be Ile274 and Pro356. Ile274, structurally equivalent to Ser216 in BLC, interacts with the heme methyl I, as oriented in HPII. With the BLC heme orientation, these interactions would become a steric hindrance with the corresponding vinyl IV group. Pro356 is close to vinyl II of the heme as oriented in HPII. Replacing this proline residue with a bulkier one (equivalent residue is a leucine in BLC) should destabilize the vinyl group in this position. To a lesser extent differences between Ile114Met, Leu407Met and Ser414Ala could also contribute to the different orientation of the heme group in HPII with respect to BLC. Temperature factors for the terminal carbon atoms (CB) from the vinyl groups are very low (3.5 $Å^2$ and 2.0 $Å^2$ for rings I and II, respectively), suggesting that the orientation modeled for the heme group in HPII presents a high level of occupancy.

The heme component of HPII in solution has been spectroscopically and chromatographically characterized as *cis*-heme *d* [14], although as much as 10–20% could be proto-heme, depending on the oxidative environment (which affects the conversion of proto-heme to heme *d*) of the cells from which the enzyme was isolated [15]. The propionic acid chain on ring III and the pyrrole ring III itself appear distorted with respect to the geometry of a heme *b* structure, consistent with the heme *d* modification. Unfortunately, the current resolution (2.8 Å) and the simultaneous presence of at least two heme types [16], does not allow the certain determination of either the

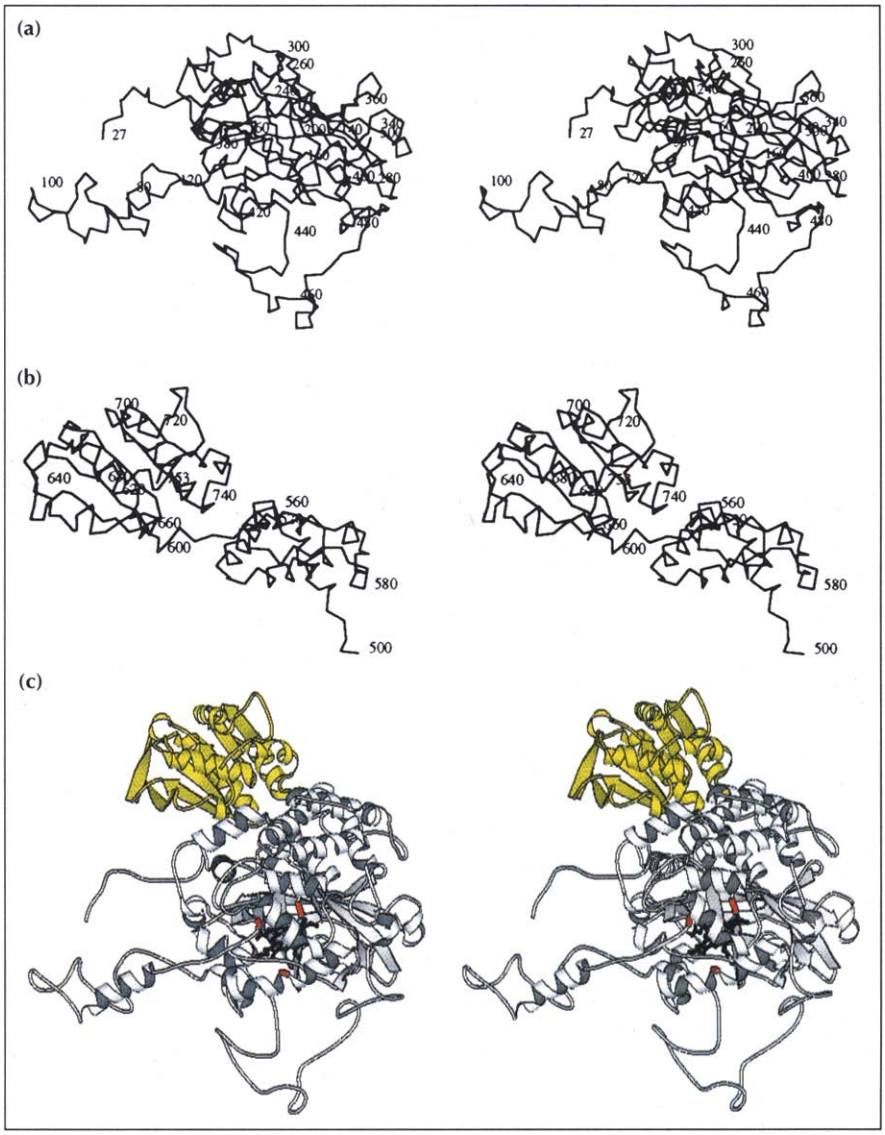

Fig. 5. The HPII subunit. For clarity, given its large size, the subunit was divided into two fragments to display the Cα trace. **(a)** The N-terminal fragment (residues 27–499). **(b)** The C-terminal fragment (residues 500–753) including a cluster of helices and the C-terminal domain. **(c)** Ribbon diagram of a single subunit. The positions of the essential residues His128, Asn201 and Tyr415 are indicated in red. The C-terminal domain is colored yellow. The heme group is represented in detail. The orientation is the same in the three drawings.

conformation of the hydroxylation on ring III, whether distal or proximal, or the actual nature of the modification, whether dihydroxyl, epoxide or lactone (Fig. 9).

When the protein model was completed, omit maps (F_o–F_c) and ($2F_o$–F_c) electron-density maps showed some extra density, in the vicinity of the heme group, probably corresponding to solvent molecules (Fig. 9c). At 2.8 Å resolution, however, solvent is often not well defined and only four water molecules, with clear electron density and very low temperature factors, have been explicitly introduced. The first water molecule, W1, appears strongly bound (continuous density) to His128 and too distant (>3 Å) from the iron atom to be directly coordinated. A solvent molecule in a similar position had been reported for PVC and MLC [2,5]. The second molécule, W2, bridges Ser167 with the propionate group from pyrrole ring IV. A solvent molecule with similar interactions and a possible role in the catalytic process had also been reported for BLC [13]. The third water, W3, is located between the two carboxylates of the propionate groups and is also hydrogen bonded to the main-chain nitrogen atom of Phe391 (distance 2.9 Å). A water molecule in an equivalent position was postulated in BLC [13] and is now clearly defined in MLC and PVC (WR Melik-Adamyan, personal communication). The last water molecule, W4, is hydrogen bonded to Asn201 and completes almost perfect tetrahedral geometry to Ile205, Asn252 and Thr203. The coincidence of these solvent molecules with the position of solvent molecules in other catalases suggests that not only the essential residues but also the solvent structure is well preserved in the heme pocket of HPII.

It has been proposed [17] that in biological systems two major factors control the properties of metal ions: firstly, the structure of the metal, including the geometry of the complex and the nature of the ligands attached to the metal, and secondly, the environment of the metal complex, including the polarity of the immediate surroundings and the steric constraints on the accessibility of substrates to the metal, and of the metal to solvent. The large degree of homology between the heme environments of BLC and HPII suggests that the formation of compound I and its reduction to free enzyme by

```
HPII   1   msqhneknph qhqsplhdss eakpgmdsla pedgshrpaa eptppgaqpt
                                          80                91
BLC   51    apgslkapdt rneklnsled vrkgsenyaL TTNQGVRIAD DQNSLRAGSR
                                         -L TTGGGNPVGD KLNSLTVGPR
                                          26

HPII  101   GPTLLEDFIL REKITHFDHE RIPERIVHAR GSAAHGYFQP YKSLSDITKA
BLC    47   GPLLVQDVVF TDEMAHFDRE RIPERVVHAK GAGAFGYFEV THDITRYSKA

HPII  151   DFLSDPNKIT PVFVRFSTVQ GGAGSADTVR DIRGFATKFY TEEGIFDLVG
BLC    97   KVFEHIGKRT PIAVRFSTVA GESGSADTVR DPRGFAVKFY TEDGNWDLVG

HPII  201   NNTPIFFIQD AHKFPDFVHA VKPEPHWAIP qgqsaHDTFW DYVSLQPETL
BLC   147   NNTPIFFIRD ALLFPSFIHS QKRNPQTHLK d     PDMVW DFWSLRPESL
                    ..                              178

HPII  251   HNVMWAMSDR GIPRSYRTME GFGIHTFRLI NAEGKATFVR FHWKPLAGKA
BLC   193   HQVSFLFSDR GIPDGHRHMD GYGSHTFKLV NADGEAVYCK FHYKTDQGIK

HPII  301   SLVWDEAQKL TGRDPDFHRR ELWEAIEAGD FPEYELGFQL IPEEDEFKFD
BLC   243   NLSVEDAARL AHEDPDYGLR DLFNAIATGN YPSWTLYIQV MTFSEAEIFP

HPII  351   FDLLDPTKLI PEELVPVQRV GKMVLNRNPD NFFAENEQAA FHPGHIVPGL
BLC   293   FNPFDLTKVW PHGDYPLIPV GKLVLNRNPV NYFAEVEQLA FDPSNMPPGI

HPII  401   DFTNDPLLQG RLFSYTDTQI SRLGGPNFHE IPINRPTCPY HNFQRDGMHR
BLC   343   EPSPDKMLQG RLFAYPDTHR HRLGPNYLQI PVNCPYRARV ANYQRDGPMC

                454
HPII  451   MGI DTNPAN YEPNSiNDnW PRETPPgpkr ggFESYQERV EGNKVRERsp
BLC   393   MMDnQGGAPN YYPNS FS A PEHQPS       ALEHRTHF SGDVQRFN
                              408               417
                 502            514
HPII  500   sF GEYYS H PRLFWL SQT PFEQRHIVDG FSFELSKVvR PYIRERVVDQ
BLC   433    SaNDDNVtQ VRTFYLkVLN EEQRKRLCEN IAGHLKDA Q LFIQKKAVKN
                                                    470

HPII  547   LAHIDLTLAQ AVAKNLGiel tddqlnitpp pdvnglkkdp slslyaipdg
BLC   481   FSDVHPEYGS RIQALLD-

HPII  597   dvkgrvvail lndevrsadl lailkalkak gvhakllysr mgevtaddgt

HPII  647   vlpiaatfag apsltvdavi vpcgniadia dngdanyylm eaykhlkpia

HPII  697   lagdarkfka tikiadqgee giveadsadg sfmdelltlm aahrvwsrip

HPII  747   kidkipa
                753
```

Fig. 6. Sequence alignment of BLC and HPII. Equivalent residues with rms deviations greater than 3.0 Å are in italics. Lower-case letters indicate non-equivalent residues (rms >5.0 Å). BLC non-equivalent residues in the N- and C-terminal regions have been omitted.

two-electron donors can follow similar steps [13] involving equivalent residues in both enzymes. On the other hand, catalytic differences [16], in particular reduction of compound I by one-electron donors giving compound II (which has never been reported for HPII), should mainly be due to the different nature of the heme groups in BLC and HPII. However, the observed reduction potentials for *d*-type hemes fall within the range observed for *b*-type hemes in different proteins [18] and thus it is not clear how the heme *d* structure can fine-tune the redox potential of the iron, or its ligand binding ability, in HPII.

C-terminal domain

The C-terminal domain spans from Gly600 to the C-terminal residue Ala753. The conformation of this domain is characterized by a large amount of secondary structure (Fig. 7). When using the DSSP classification [19], four α-helices, varying in length from 12–16 residues, and eight β-strands may be distinguished. Both the first and second helices (α_I and α_{II}) are regular with amphiphilic character and are formed by 16 residues. α_I extends from Val611–Lys626 and α_{II} from Asn681–Lys693. The central part of α_{II} is formed by bulky residues (from Tyr683–His691, Fig. 7) and the corresponding electron density was visible even in the initial maps. α_{III} starts and ends with short segments of 3_{10}-helices. The first two residues of α_{III} (Ala698 and Gly699) also belong to strand β_{VII}. The terminal helix α_{IV} is continued with several i+3 hydrogen bonds for about 10 more residues. Six β-strands (β_I, β_{II}, β_V, β_{VI}, β_{VII} and β_{VIII}) form a parallel β-sheet. The longest strand, β_{II}, is hydrogen bonded with β_I and β_V, and includes residues Gly627–Ser635. Two residues along, at Gly638, strand β_{III} starts and this forms an antiparallel β-sheet with β_{IV}.

When the C-terminal domain of HPII was compared with that observed in PVC, some structural changes were apparent. These included a deletion (about 20 additional amino acids are present in PVC between the β_{VI} and α_{II} of HPII), at least one insertion (the antiparallel β-sheet containing β_{III} and β_{IV} is absent in PVC), and some α-helix displacements or modifications (the i+3 hydrogen bond continuation of the terminal helix is not present in PVC). The role of this domain, absent in most other smaller heme catalases, remains unknown.

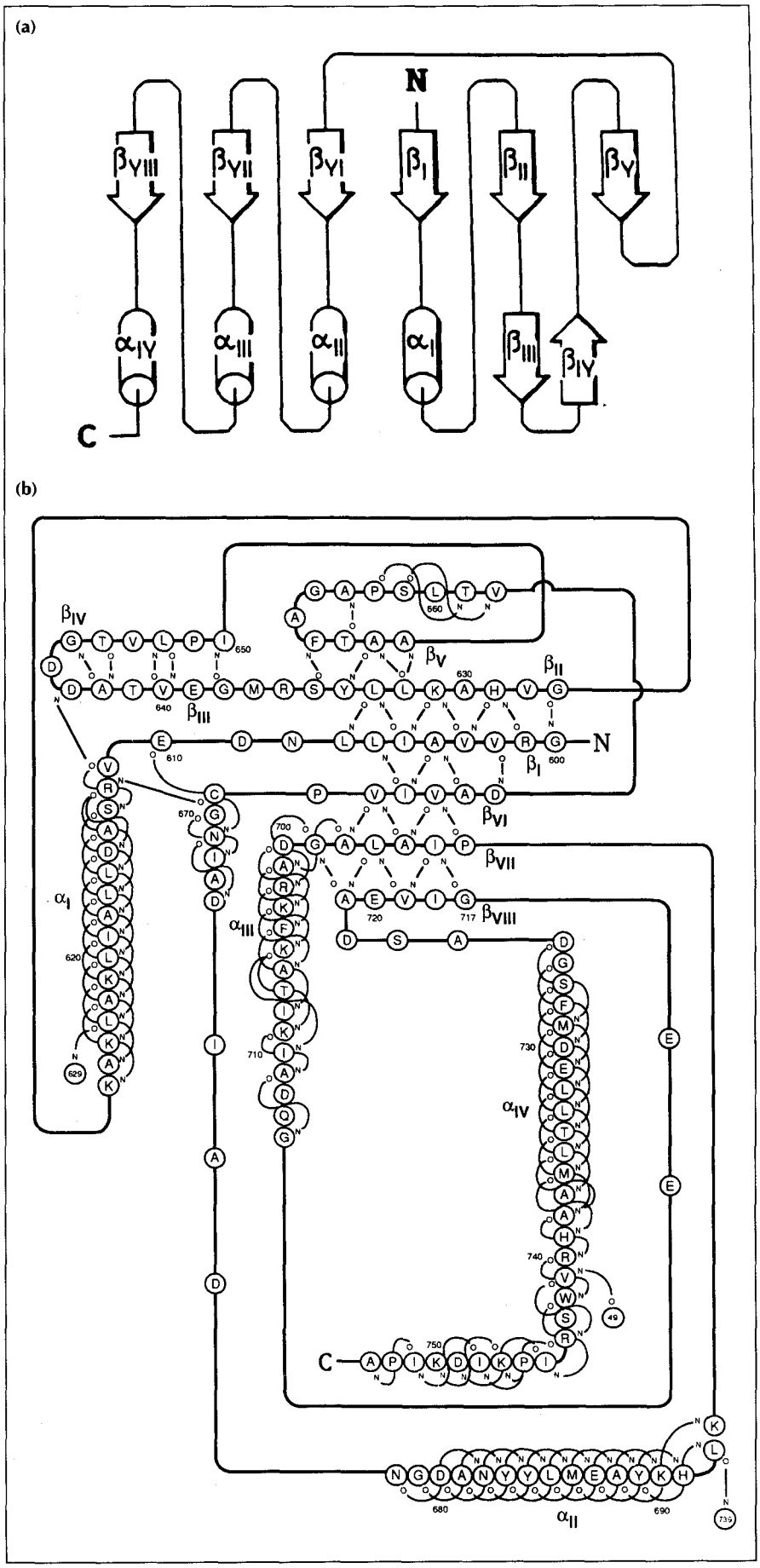

Fig. 7. Schemes of **(a)** the secondary structure topology and **(b)** the main-chain hydrogen bond organization in the C-terminal domain (see text).

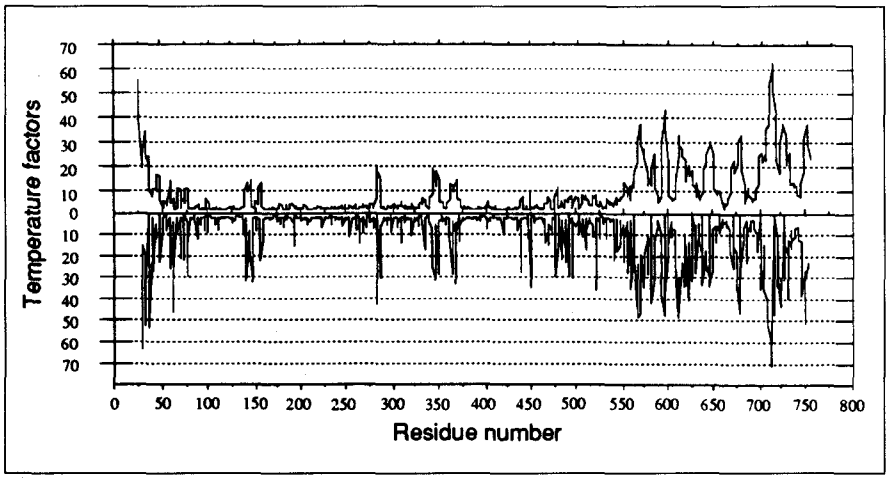

Fig. 8. Temperature factors (in Å^2) for main-chain (top line) and side-chain (bottom line) atoms. Deviations from the strict non-crystallographic symmetry used may contribute to increases in the averaged temperature factors (see text). In particular, this may contribute to the high mobility of the N-terminal region and to the strong B-factor fluctuations along the sequence in the C-terminal domain. Residues with the highest temperature factors appear always to be involved in crystal interactions.

The terminal amino acid (Ala753), clearly visible in the electron-density maps, is close to the molecular R dyad axis (Fig. 10). Its carboxylate group forms a salt bridge with the terminal amino group of Lys690 and a hydrogen bond with the hydroxyl group of Tyr683 from the same subunit.

Biological implications

Heme-containing catalases are ubiquitous enzymes present in the cells of most aerobic organisms. These enzymes serve, in part, to protect the cell from the toxic effects of small peroxides. However, the entire range of biological functions of catalases remains unclear.

***Escherichia coli* produces two catalases, HPI and HPII, which appear quite distinct in physical structure and catalytic properties, both from each other and from other catalases. HPII, studied in this work, had been characterized as a monofunctional catalase with a large polypeptide chain of 753 residues, containing one *cis*-heme *d* isomer per subunit. Despite the differences in size, oligomerization and chemical properties, suggesting a unique catalase, the organization of HPII resembles that of tetrameric *Penicillium vitale* catalase (PVC), whose sequence is as yet chemically unknown. The most striking differences of HPII from the conformations of beef liver catalase (BLC), *Micrococcus lysodeikticus* catalase (MLC) and *Proteus mirabilis* (PMC) are firstly, the presence in HPII of a C-terminal domain of about 150 residues that is structurally related to the 'flavodoxin-like' domain described for PVC, and secondly, inversion of the orientation of the heme group and its modification into a heme *d*, which could be related events. An additional difference of HPII with respect to the other catalases whose structures are known, is the presence of about 60 amino acids at the N-terminal end that increase the contact area between subunits and partially close the lateral entrance to the main channel, described for BLC. In contrast, the central part of HPII, about 475 residues, is structurally very similar to BLC.**

Assuming catalases diverged from a common ancestor antedating the prokaryote/eukaryote divergence, the additional C-terminal domain may have been deleted from the primordial catalase gene or added by fusion. The simultaneous presence and absence of the C-terminal domain in both prokaryote and eukaryote catalases makes a simple explanation difficult. Another apparent paradox is that despite the fact that neither PVC nor HPII has been shown to bind nucleotides, they both contain a domain with nucleotide-binding topology, whereas BLC, which does not contain this domain, has been shown to bind NADPH.

The structure of HPII, reported in this work, represents the largest catalase structure determined to date, to near atomic resolution. Higher resolution studies, now in progress, are required to accurately define the structural differences in the active sites of HPII and other catalases that may account for their specific chemical properties.

Materials and methods

Enzyme purification

Cultures of *E. coli* strain UM255 *pro leu rpsL hsdM hsdR endI lacY katG2 katE12::Tn10 recA* [20] transformed with the plasmid pAMkatE72 [11] containing the *katE* gene encoding HPII, were grown in Luria broth (LB) medium containing 10 g L^{-1} tryptone, 5 g L^{-1} yeast extract, and 5 g L^{-1} NaCl for 16 h at 37°C with shaking. Cells were harvested and HPII was isolated as described by Loewen and Switala [10] with the modification that DEAE-cellulose DE52 (Whatman) was used in place of DEAE-Sephadex A-25.

Crystallization and data collection

Bulky HPII crystals reported previously [21] presented high anisotropic mosaicity, and diffracted to 2.8 Å resolution. More satisfactory crystals (0.8 mm×0.6 mm×0.2 mm) with plate-like morphology, that diffracted beyond 2.0 Å resolution, were

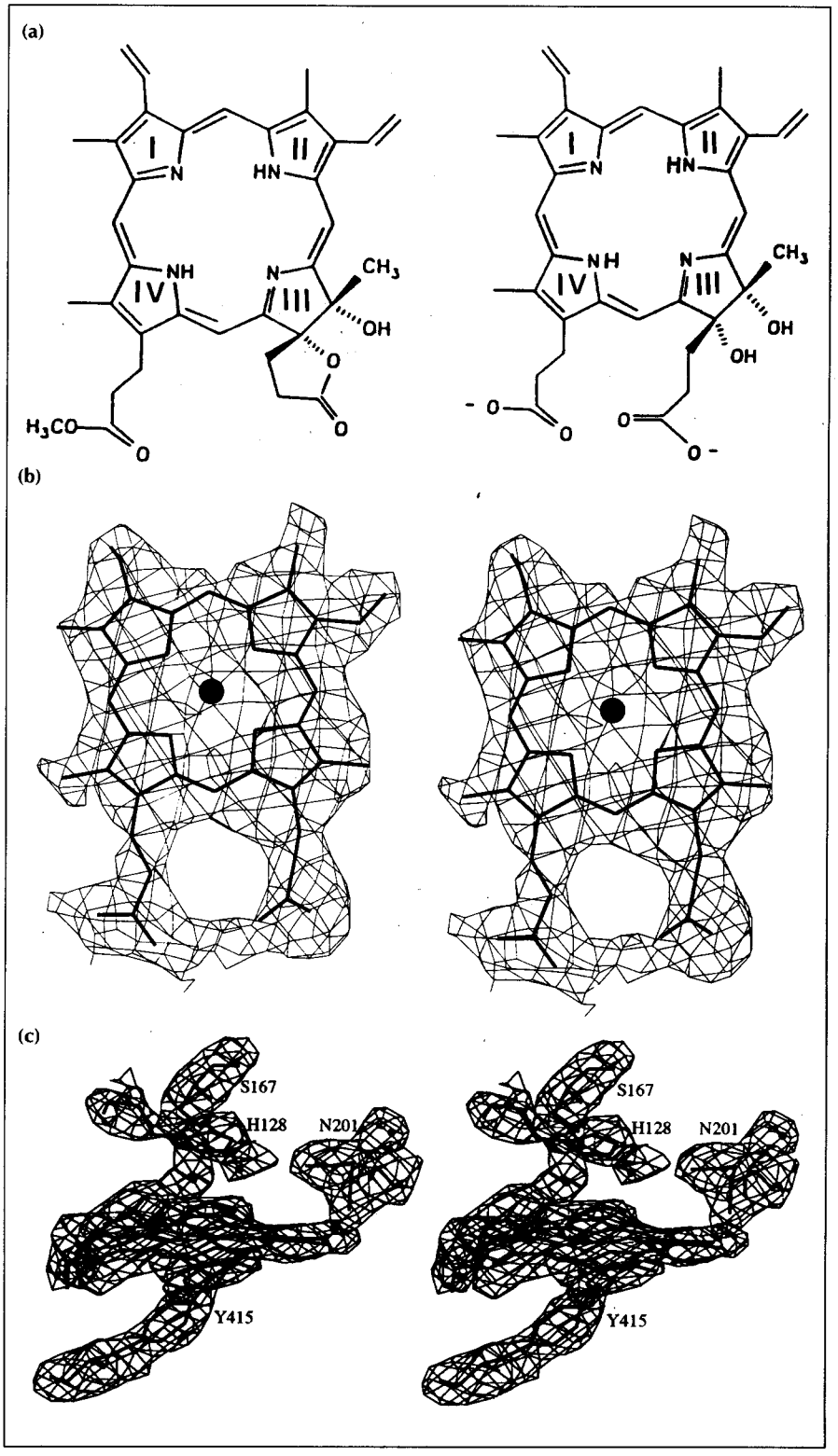

Fig. 9. **(a)** Structures of the *cis*-diol heme *d* (right) and of the related spirolactone fused to ring III (left). **(b),(c)** Stereoviews of the heme environment. In (b), the averaged omit (F_o–F_c) map computed with the heme group omitted from the model is shown, viewed from the distal side. As modifications in the heme ring III could not be well defined at the present resolution (see text), the protoporphyrin IX model is used for the heme representation. The disposition of the vinyl and methyl groups, in pyrrole rings I and II, appears inverted with respect to the orientation determined in BLC. In (c), the unaveraged ($2F_o$–F_c) electron density of the heme pocket is shown with the essential residues His128, Ser167, Asn201 and Tyr415. Four well defined solvent molecules are also shown.

obtained by changing the crystallization conditions and were used in this work. The crystals appeared in a few days using the hanging-drop vapor-diffusion method at room temperature, with 15% (w/v) PEG 3350 and 1.5 M LiCl in 0.2 M Tris-HCl, pH 9.0 as precipitants. Crystals are monoclinic, space group $P2_1$ (see below), and unit cell dimensions a=94.13 Å, b=133.66 Å, c=123.05 Å and β=109.4°. A data set (Table 1), with about 90% of the unique reflections at 2.8 Å, and a total internal agreement factor (R_{sym}) of 10.8%, was obtained by merging two partial data sets. The first data set was collected from one crystal using synchrotron radiation (Hamburg, EMBL DESY, λ=0.92 Å). The second data set was collected using a conventional source (a GX21 generator running at 2.4 kW with monochromator, λ=1.54 Å) and four different crystals. Both data sets were measured with an Image Plate (MAR RESEARCH) area detector and processed with the MOSFLM package [22]. The ratio $\langle I \rangle / \langle \sigma(I) \rangle$ for the highest resolution shell is 3.5.

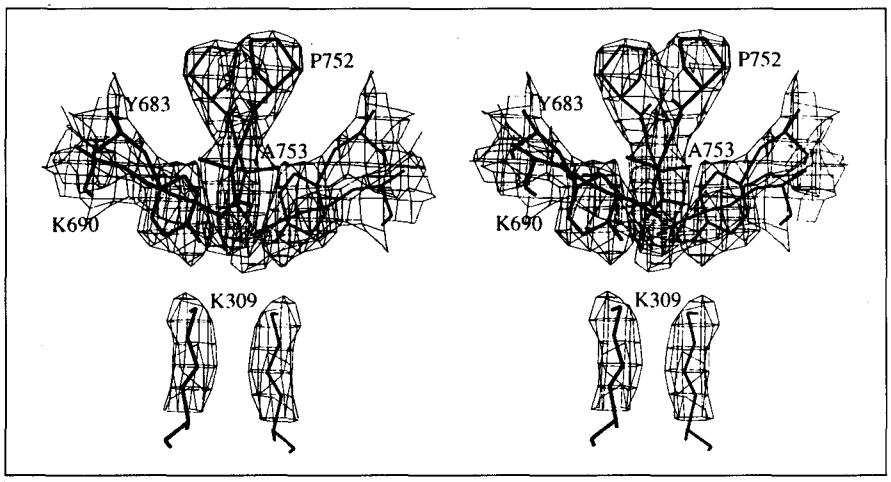

Fig. 10. Stereoview of the electron density in the terminal carboxylate environment (residue Ala753). The molecular dyad R-axis-related residues are shown with thinner bonds. The terminal carboxylate charged group appears to be neutralized by Lys309.

Table 1. Data collection statistics.

	No. of crystals	Resolution (Å)	Completeness (%)	R_{merge} (%)§	Total reflections	Unique reflections
Data set 1*	1	2.80	43.5	9.7	107 702	30 807
Data set 2†	4	2.80	50.2	10.5	117 543	35 516
Merged	5	2.80	90.2	10.8	225 245	65 329

*Data set 1 was collected using synchrotron radiation at the EMBL, DESY. †Data set 2 was collected using a rotating anode (see text). It was obtained attempting to minimize the number of additional images required to complete data set 1. §$R_{merge} = \Sigma|I - \langle I \rangle|/\Sigma I$.

Structure determination

Space group and molecular symmetry: To examine the molecular symmetry two possibilities were considered. Firstly, a hexameric molecule (as suggested by molecular weight analysis [10]) and secondly a tetrameric molecule (with 222 molecular symmetry as found in all previously determined heme catalase structures) with four subunits in the asymmetric unit (specific volume, V_m=2.2 Å^3 Da^{-1}). For the hexameric model the asymmetric unit had to contain three subunits (V_m=2.9 Å^3 Da^{-1}), a two-fold molecular axis coinciding with a binary crystal axis and the space group had to be P2. For tetrameric molecules either the whole tetramer or two dimers (again with a molecular axis coinciding with a binary crystal axis) could be placed inside the asymmetric unit.

The self rotation was used to discriminate between hexameric and tetrameric models. No clear peaks were observed when searching for a three-fold axis. Instead one well defined peak, perpendicular to the b axis, was observed for κ=180°. The locked self-rotation [23] gave a strong correlation when the 222 point group symmetry was used. Four subunits were assumed to be placed in the asymmetric unit of the crystal with space group $P2_1$. Crystal structures of catalase previously reported had at least one molecular axis coincident with a crystallographic axis and contained two subunits in the asymmetric unit for PVC [1] and BLC [2] and only one for MLC [5] and PMC [6].

Initial structure determination: Given the molecular symmetry found for HPII and the sequence similarity between BLC (containing about 505 residues per subunit) with the central part of HPII (containing 753 residues per subunit) [11] molecular replacement was attempted. The search model used was the BLC structure including residues 25–490 (PDB reference 7CAT). When only a single BLC subunit was used (the model corresponded to about 1/8 of the asymmetric unit content) the cross-rotation (applied with the program GLRF [23]) did not give an interpretable result. Instead the locked cross-rotation [23], with the 222 point group symmetry observed in BLC, gave a clean solution. The orientation determined initially was further refined both using smaller grid steps and explicitly the whole tetramer.

With the refined orientation, the translation in the XZ plane was found using the BRUTE program [24] and the BLC tetramer as model. Systematic searches were done in the contact asymmetric unit of the $P2_1$ space group, starting with grid steps of 0.5 Å. The steps were progressively reduced and the searches repeated around the positions with the highest correlation coefficients. The best correlation obtained was 0.30 for 2429 reflections in the resolution shell from 4.75 Å to 5.5 Å. Simultaneously, a steric analysis, done with the PACKFUN option from the MERLOT package [25] indicated that the position found with BRUTE was located in the narrow region between total absence of contacts and steric hindrance for both the BLC and the PVC tetramers.

Finally a rigid-body refinement was done with the X-PLOR program [26] giving an agreement R-factor of 48.0% for 21 698 reflections in the resolution shell from 10.0 Å to 4.0 Å. The skew matrix and transformation vector origin of the skew frame are, respectively:

$$\mathbf{P} = \begin{pmatrix} -0.9938 & -0.10913 & 0.02217 \\ 0.1091 & -0.99403 & -0.00148 \\ 0.0222 & 0.00103 & 0.99975 \end{pmatrix}$$

$$\mathbf{O} = (15.75,\ 0.0,\ 6.08)$$

From coordinates in the skew reference frame X_2 (PQR coordinates) the coordinates X_1 in the reference orthogonal Å

frame (the PDB standard reference frame) are obtained, according to the notation used in X-PLOR, as:

$$X_1 = PX_2 + O$$

In the skew (molecular) frame the reference subunit was taken, as in BLC [4], with all the iron PQR coordinates positive. The orthogonal molecular reference frame PQR is defined with its origin coincident with the center of the molecular point group and with every axis parallel to a molecular dyad axis. The molecular axis Q is only 6.3° apart from the crystal axis b.

Attempts to use the available polyglycine PVC coordinates (PDB reference 4CAT) or to use a hybrid formed by the BLC model plus the C-terminal domain of PVC, did not show improvement at any level during the initial structure determination.

Averaging, model building and structure completion: The first $(2F_o-F_c)$ electron-density map already allowed us to introduce some of the HPII sequence into the model. However, this map was noisy and uninterpretable in many locations. In particular, no clear density could be distinguished for the (at least) 250 extra residues that HPII contains with respect to BLC. Cycles of four-fold non-crystallographic symmetry averaging (ENVELOPE package [27]) were then used alternating with manual model rebuilding in a graphics system and with automatic cycles of positional refinement using X-PLOR.

The averaging was started at 3.7 Å and the resolution was slowly extended to 2.9 Å. For each resolution shell, cycles of averaging were repeated until the average phase shift was smaller than 4°. The starting phases for each new resolution shell were obtained from the molecular model. The last cycle of the iterative averaging (at 2.9 Å resolution) had a correlation coefficient [28] of 89.7% and an R-factor of 17.0%. Skewed and averaged $(2F_o-F_c)$ and (F_o-F_c) maps were used for the manual rebuilding. Positional refinement was in general applied using simulated annealing and followed the slow cooling protocol [26]. Finally, with the agreement R-factor <25%, individual atomic isotropic B-factors (with restrained values between neighboring atoms) were also refined within X-PLOR. Only four, well defined, water molecules in the vicinity of the heme group were included in the final model. Strict non-crystallographic symmetry has been imposed during all the refinement steps. Some tests were performed in which the non-crystallographic symmetry was relaxed, to check if the electron density could be improved particularly in the N-terminal region. No new residues could be added in the resulting unaveraged maps. In these tests the R-factor decreased easily below 18% mostly due to the important increment in the number of parameters used.

The coordinates for *E. coli* HPII are being deposited with the Brookhaven Protein Data Bank.

Acknowledgements: Programs GLRF and ENVELOPE were kindly provided by MG Rossmann and implemented in a CRAY YMP computer with the help of Jin-Bi Dai and Hao Wu. We thank WR Melik-Adamyan and JC Ferrer for inspiring comments. Many thanks are given to the Computing Center (CESCA) for efficient assistance. We thank J Ripol for technical help. This Research was supported by grants BIO90-0756 and PB92-0707 from the DGICYT of Spain to IF and grant OGP9600 from the Natural Sciences and Engineering Research Council of Canada to PCL.

References

1. Vainshtein, B.K., Melik-Adamyan, W.R., Barynin, V.V., Vagin, A.A. & Grebenko, A.I. (1981). Three-dimensional structure of the enzyme catalase. *Nature* **293**, 411–412.
2. Vainshtein, B.K., *et al.*, & Rossmann, M.G. (1986). Three-dimensional structure of catalase from *Penicillium vitale* at 2.0 Å resolution. *J. Mol. Biol.* **188**, 49–61.
3. Murthy, M.R.N., Reid, T.J., III, Sicignano, A., Tanaka, N. & Rossmann, M.G. (1981). Structure of beef liver catalase. *J. Mol. Biol.* **152**, 465–499.
4. Fita, I., Silva, A.M., Murthy, M.R.N. & Rossmann, M.G. (1986). The refined structure of beef liver catalase at 2.5 Å resolution. *Acta Crystallogr. B* **42**, 497–515.
5. Murshudov, G.N., *et al.*, & Wilson, K.S. (1992). Three-dimensional structure of catalase from *Micrococcus lysodeikticus* at 1.5 Å resolution. *FEBS Lett.* **312**, 127–131.
6. Gouet, P. (1993). Determination de la structure 3-D de la catalase de la bacterie *Proteus mirabilis* par diffraction aux rayons X: structure avec et sans NADPH [PhD Thesis]. University of Paris, XI Orsay, France.
7. Kirkman, H.N. & Gaetani, G.F. (1984). Catalase: a tetrameric enzyme with four tightly bound molecules of NADPH. *Proc. Natl. Acad. Sci. USA* **81**, 4343–4347.
8. Claiborne, A. & Fridovich, I. (1979). Purification of the O-dianisidine peroxidase from *Escherichia coli B. J. Biol. Chem.* **254**, 4245–4252.
9. Sak, B.D., Eisenstrk, A. & Touati, D. (1989). Exonuclease III and the catalase hydroperoxidase II in *Escherichia coli* are both regulated by the *katF* gene product. *Proc. Natl. Acad. Sci. USA* **86**, 3271–3275.
10. Loewen, P.C. & Switala, J. (1986). Purification and characterization of catalase HPII from *Escherichia coli* K12. *Biochem. Cell Biol.* **64**, 638–646.
11. von Ossowski, I., Mulvey, M.R., Leco, P.A., Borys, A. & Loewen, P.C. (1991). Nucleotide sequence of *Escherichia coli katE*, which encodes catalase HPII. *J. Bacteriol.* **173**, 514–520.
12. Fita, I. & Rossmann, M.G. (1985). The NADPH binding site on beef liver catalase. *Proc. Natl. Acad. Sci. USA* **82**, 1604–1608.
13. Fita, I. & Rossmann, M.G. (1985). The active center of catalase. *J. Mol. Biol.* **185**, 21–37.
14. Chiu, J.T, Loewen, P.C., Switala J., Gennis R.B. & Timkovich R. (1989). Proposed structure for the prosthetic group of the catalase HPII from *Escherichia coli. J. Am. Chem. Soc.* **111**, 7046–7050.
15. von Ossowski, I., Hausner, G. & Loewen, P.C. (1993). Molecular evolutionary analysis based on the amino acid sequence of catalase. *J. Mol. Evol.* **37**, 71–76.
16. Loewen, P.C., *et al.*, & Nicholls, P. (1993). Catalase HPII of *Escherichia coli* catalyzes the conversion of protoheme to *cis*-heme *d. Biochemistry* **32**, 10159–10164.
17. Dawson, J.H. (1988). Probing structure-function relations in heme-containing oxygenases and peroxidases. *Science* **240**, 433–439.
18. Timkovich, R. & Bondoc, L.L. (1990). Diversity in the structure of hemes. *Adv. Biophys. Chem.* **1**, 203–247.
19. Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637.
20. Mulvey, M.R., Sorby, P.A., Triggs-Raine, B.L., & Loewen, P.C. (1988). Cloning and physical characterization of *katE* and *katF* required for catalase HPII expression in *Escherichia coli. Gene* **73**, 337–345.
21. Tormo, J., Fita, I., Switala, J. & Loewen, P.C. (1990). Crystallization and preliminary X-ray diffraction analysis of catalase HPII from *Escherichia coli. J. Mol. Biol.* **213**, 219–220.
22. Leslie, A.G.W. (1991). *Recent Changes to the MOSFLM Package for Processing Film and Image Plate Data.* CCP4 and ESF-EACMB, Newsletter on Protein Crystallography, SERC Laboratory, Daresbury, Warrington, UK.
23. Tong, L. & Rossmann, M.G. (1990). The locked rotation function. *Acta Crystallogr. A* **46**, 783–792.
24. Fujinaga, M. & Read, R.J. (1987). Experiences with a new translation-function program. *J. Appl. Crystallogr.* **20**, 517–521.
25. Fitzgerald, P.M.D. (1988). MERLOT, an integrated package of computer programs for the determination of crystal structures by molecular replacement. *J. Appl. Crystallogr.* **21**, 273–278.

26. Brünger, A.T. (1992). *X-PLOR, Version 3.1: A System for X-ray Crystallography and NMR.* Yale University Press, New Haven and London.
27. Rossmann, M.G., *et al.*, & Choi, H.-K. (1992). Molecular replacement real-space averaging. *J. Appl. Crystallogr.* **25**, 166–180.
28. Rossmann, M.G., *et al.*, & Vriend, G. (1985). Structure of a human common cold virus and functional relationship to other picornaviruses. *Nature* **317**, 145–153.
29. Laskowski, R.A., McArthur, M.W., Moss, D.S. & Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 282–291.
30. Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structure. *J. Appl. Crystallogr.* **24**, 946–950.

Received: **30 Jan 1995**; revisions requested: **16 Feb 1995**; revisions received: **2 Mar 1995**. Accepted: **22 Mar 1995**.