Structure of the Clade 1 Catalase, CatF of *Pseudomonas syringae*, at 1.8 Å Resolution

Xavi Carpena,[1] Manuel Soriano,[2] Martin G. Klotz,[3] Harry W. Duckworth,[4] Lynda J. Donald,[4] William Melik-Adamyan,[5] Ignacio Fita,[1] and Peter C. Loewen[6]*

[1]*CID-CSIC, Jordi-Girona 18-26, Barcelona, Spain*
[2]*Departamento de Bioquimica, Instituto de Quimica, UNAM, Mexico DF, Mexico*
[3]*Department of Biology, University of Louisville, Louisville, Kentucky*
[4]*Department of Chemistry, University of Manitoba, Winnipeg, Manitoba, Canada*
[5]*Institute of Crystallography, Russian Academy of Sciences, Moscow, Russia*
[6]*Department of Microbiology, University of Manitoba, Winnipeg, Manitoba, Canada*

ABSTRACT **Catalase CatF of *Pseudomonas syringae* has been identified phylogenetically as a clade 1 catalase, closely related to plant catalases, a group from which no structure has been determined. The structure of CatF has been refined at 1.8 Å resolution by using X-ray synchrotron data collected from a crystal flash-cooled with liquid nitrogen. The crystallographic agreement factors R and R_{free} are, respectively, 18.3% and 24.0%. The asymmetric unit of the crystal contains a whole molecule that shows accurate 222-point group symmetry. The crystallized enzyme is a homotetramer of subunits with 484 residues, some 26 residues shorter than predicted from the DNA sequence. Mass spectrometry analysis confirmed the absence of 26 N-terminal residues, possibly removed by a periplasmic transport system. The core structure of the CatF subunit was closely related to seven other catalases with root-mean-square deviations (RMSDs) of 368 core Cα atoms of 0.99–1.30 Å. The heme component of CatF is heme b in the same orientation that is found in *Escherichia coli* hydroperoxidase II, an orientation that is flipped 180° with respect the orientation of the heme in bovine liver catalase. NADPH is not found in the structure of CatF because key residues required for nucleotide binding are missing; 2129 water molecules were refined into the model. Water occupancy in the main or perpendicular channel of CatF varied among the four subunits from two to five in the region between the heme and the conserved Asp150. A comparison of the water occupancy in this region with the same region in other catalases reveals significant differences among the catalases. Proteins 2003;50:423–436.**

Key words: catalase; crystal structure; NADPH binding; heme orientation; water occupancy

INTRODUCTION

The enzyme catalase is a key enzyme in protecting organisms against reactive oxygen species through its degradation of hydrogen peroxide to water and oxygen (reaction 1).[1]

$$2H_2O_2 \rightarrow 2H_2O + O_2 \quad (1)$$

This deceptively simple reaction takes place in two steps with the first hydrogen peroxide molecule oxidizing the heme to an oxyferryl species in which one oxidation equivalent is removed from the iron and one from the porphyrin ring to generate a porphyrin cation radical (reaction 2). The second hydrogen peroxide is then used as a reductant of compound I to regenerate the resting state enzyme, water and oxygen (reaction 3).

$$\text{Enz (Por–Fe}^{III}\text{)} + H_2O_2 \rightarrow \text{Cpd I (Por}^{+\cdot}\text{–Fe}^{IV} = O) + H_2O \quad (2)$$

$$\text{Cpd I (Por}^{+\cdot}\text{–Fe}^{IV} = O) + H_2O_2 \rightarrow \text{Enz (Por–Fe}^{III}\text{)} + H_2O + O_2 \quad (3)$$

This catalatic reaction has evolved in at least three phylogenetically unrelated protein forms including the monofunctional or "classical" catalase, the bifunctional catalase-peroxidase, and the non-heme or Mn-containing catalase.[1] The most extensively characterized class is the monofunctional catalases, which were first reported in 1900, and have subsequently been found in both small subunit (55–60 kDa) and large subunit (78–84 kDa) forms. Seven such monofunctional catalases have had their crystal structures solved including the enzymes from bovine liver (BLC),[2,3] *Penicillium vitale* (PVC),[4,5] *Micrococcus lysodeikticus* (MLC),[6] *Proteus mirabilis* (PMC),[7] *Escherichia coli* (HPII),[8,9] *Saccharomyces cerevisiae* (CATA),[10,11] and human erythrocytes (HEC).[12,13]

A phylogenetic analysis of 72 separate monofunctional catalases from different organisms suggested that the enzyme had evolved through a minimum of two gene duplication events producing three main clades or groups.[14] Clade 1 contains small subunit catalases from plants and a subset of bacteria; clade 2 contains large subunit catalases

Grant sponsor: DGICYT; Grant sponsor: Natural Sciences and Engineering Research Council of Canada; Grant number: OGP9600.

*Correspondence to: Peter C. Loewen, Department of Microbiology, University of Manitoba, Winnipeg, MB R3T 2N2.
E-mail: peter_loewen@umanitoba.ca

Received 2 July 2002; Accepted 23 August 2002

TABLE I. Data Collection and Structural Refinement Statistics

A. Data collection statistics				
Space group	$P2_1$			
Cell *a, b, c* (Å) β°	60.63	153.87	109.19	102.8
Resolution range (Å)	21.9–1.8		(1.8–1.87)	
Unique reflections (F > 0)	175,279		(17,810)	
Completeness (%)	97.8		(95.1)	
$<I/\sigma(I)>$[a]	15.5		(5.9)	
R_{sym} (%)[b]	8.4		(21.0)	
B. Refinement statistics				
Working set	166,506		(16,910)	
Free reflections	8773		(900)	
R_{cryst} (%)[c]	18.7		(23.4)	
R_{free} (%)	24.1		(29.4)	
No. of nonhydrogen atoms				
Protein	15,232			
Water	2,136			
Heme	172			
RMSD from ideality				
Bond lengths (Å)	0.011			
Bond angles (°)	2.1			
Planarity (Å) (peptide)	0.022			
(aromatic)	0.009			
Estimated coordinate error (Å)				
Luzzati	0.18			
Averaged B factor (Å^2)				
(Subunit)	A	B	C	D
Main-chain	11.9	19.3	19.3	11.0
Side-chain	14.3	21.1	21.1	13.6
Water	29.0			

[a]Values in parentheses correspond to the highest resolution shell.
[b]$R_{sym} = {}_{hkl}\Sigma_j|I_{hkl,j} - <I_{hkl}>|/\Sigma_{hkl}<I_{hkl}>$
[c]$R_{cryst} = \Sigma||F_{obs}| - |F_{calc}||/\Sigma|F_{obs}|$. R_{free} is as for R_{cryst} but calculated for a test set comprising reflections not used in the refinement (5%).

from fungi and a second subset of bacteria; and clade 3 contains small subunit catalases from bacteria, fungi, protists, animals, and plants. Structures of clade 2 (PVC and HPII) and clade 3 (BLC, HEC, PMC, MLC, and CATA) enzymes have been determined, but a structure of a representative from clade 1 has not yet appeared. Significant conservation of structure is evident among the five clade 3 catalases and even between the core regions of the two large subunit clade 2 enzymes and the small subunit clade 3 enzymes. The report suggesting that KatA of *P. aeruginosa*, which seemed biochemically related to CatF of *P. syringae*, might be a heterotrimer, composed of two large and one small subunit,[15] presented the possibility that some clade 1 and 3 catalases may diverge significantly from the clade 2 catalases in structure. A preliminary report describing the crystallization of two clade 1 catalases, which included CatF from *P. syringae*, has appeared.[16] This article reports the structure of CatF refined at 1.8 Å resolution, revealing that it is a homotetramer with a very strong structural resemblance to other monofunctional catalases.

MATERIALS AND METHODS

Protein Purification and Size Analysis

The CatF protein analyzed in this work was purified from the catalase-deficient *E. coli* UM255 as previously described.[16] For mass spectrometry analysis, protein was dialyzed into 5 mM ammonium acetate. For the denatured spectrum, the protein was diluted to about 10 μM subunit in 2% acetic acid, 50% methanol. Spectra were acquired by using conventional electrospray ionization in an orthogonal time-of-flight mass spectrometer[17,18] with 150 V declustering voltage and nitrogen as the curtain gas. For nondenatured spectra, the protein was diluted to 5 μM subunit in 5 mM ammonium acetate and subjected to nanospray ionization[19] from a New Objective Pico Tip capillary, also on an orthogonal time-of-flight mass spectrometer with a declustering voltage of 250 V and SF_6 as the curtain gas.

Crystallization and Data Treatment

Crystals were obtained at room temperature by the vapor diffusion hanging drop method at a protein concentration of about 6 mg/mL over a reservoir containing 12% PEG 4000 in 0.1 M sodium cacodylate pH 6.[16] Crystals were monoclinic, space group $P2_1$ with one tetrameric molecule in the crystal asymmetric unit. Diffraction data were obtained from crystals cooled with a nitrogen cryostream giving unit cell parameters of a = 60.6 Å, b = 153.9 Å, c = 109.2 Å, and β =102.8°. The diffraction data set, obtained by using synchrotron radiation at ESRF in Grenoble, was processed with use of the program DENZO and scaled with program SCALEPACK[20]; 5% of the mea-

sured reflections in every data set were reserved for R_{free} monitoring during automatic refinement (Table I).

Structure determination was performed with the program AMoRe[21] using HPII with 90 residues and 160 residues removed from the N- and C-terminal regions, respectively, and with the appropriate residue changes to reflect the CatF sequence as the initial searching model. Refinements were completed by using the program REFMAC[22] with solvent molecules modeled with the program WATPEAK[23] and manually with the graphics program O.[24] Solvent molecules were only introduced when they corresponded to the strongest peaks in the difference Fourier maps that could make at least one hydrogen bond with atoms already in the model. In the final rounds of refinement, the four subunits were treated independently with the bulk solvent correction applied and the whole resolution range available used for each variant. The analysis of solvent accessibility and molecular cavities was performed with program VOIDOO[25] using a reduced atomic radius for polar atoms in accounting for possible hydrogen bonds.[26] All figures were prepared by using SETOR.[27]

Structure factors and coordinates are available from the Protein Data Bank under accession number 1M7S.

RESULTS AND DISCUSSION

Size Confirmation

CatF is a clade 1 catalase that serves as the housekeeping enzyme with comparable substrate affinity to the clade 3 catalase KatA in *P. aeruginosa*.[15,28] Kat A had been characterized as a heterotrimer with two 60-kDa subunits and one 44-kDa subunit,[18] suggesting a structure that would be novel among catalases. A heteromultimer was also proposed for CatF,[28] but the proportion of the smaller subunit was much diminished when the enzyme was expressed in the heterologous host, catalase-deficient *E. coli*. Furthermore, the preliminary report of the crystallization of CatF revealed a single subunit of 57 kDa[16] as determined by SDS polyacrylamide gel electrophoresis. To confirm more accurately the subunit composition of CatF, the size of denatured and nondenatured CatF was analyzed by electrospray mass spectrometry. A denatured mass of 53,853 Da and a nondenatured mass of 218,369 Da were determined (Fig. 1) consistent with CatF being a homotetramer with four associated heme groups. Unexpectedly, the 53,853-Da subunit size is smaller than the 56,730 Da predicted from the corrected open reading frame of *catF*[28] (Fig. 2) but is consistent with the 53,855 Da predicted size of the truncated, periplasmic form of CatF with 26 N-terminal residues removed.[28,29] Finding the catalase to be entirely in the truncated form was unexpected and speaks to a very efficient periplasmic transport system and to the likelihood that the untruncated form of the protein does not fold sufficiently well to avoid proteolysis, which is a common fate of poorly folding proteins in *E. coli*.[30] A slightly larger cytoplasmic CatF species observed in cytoplasmic fluids of expression hosts was likely the product of regulation of translation from an overlapping open reading frame.[28]

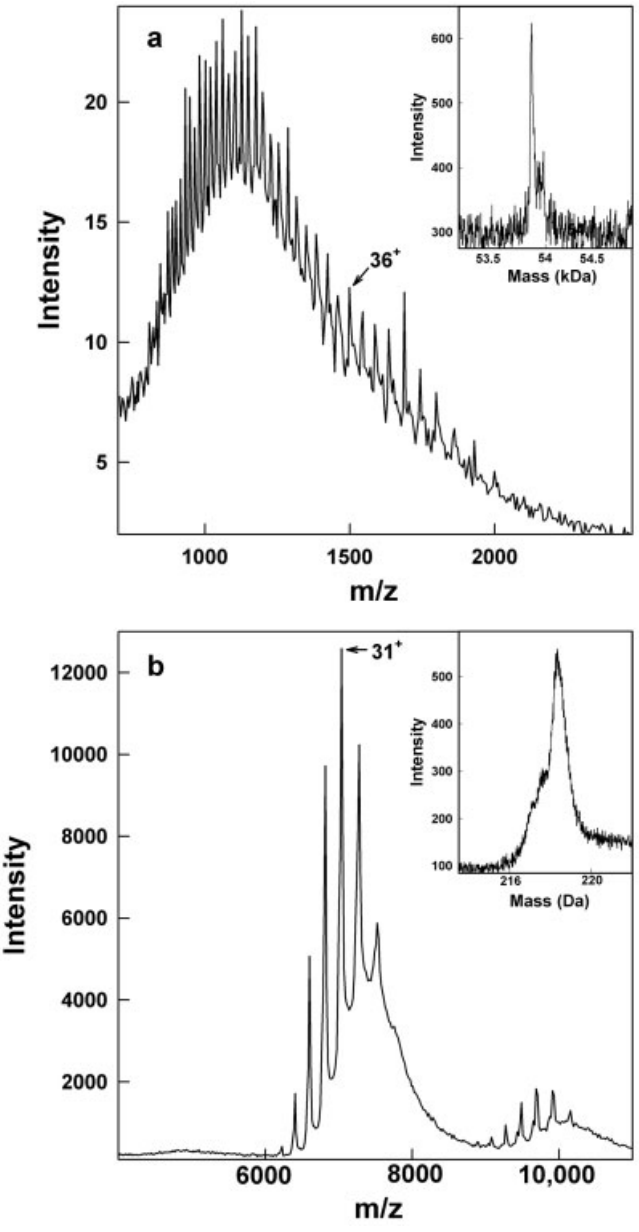

Fig. 1. Mass analysis by electrospray mass spectrometry of denatured (**a**) and nondenatured (**b**) samples of CatF.

Quality of the CatF Model

The electron density map defines main-chain and side-chain atoms of 1934 amino acids and four heme groups in four subunits and 2129 water molecules. As predicted by the mass spectrometry data, the 26 N-terminal residues of all four subunits do not appear in the structure. The maps show clear continuity in subunits A and D over the complete length from Thr27 to Asp510 and in subunits B and C over the length from Ala28 to Asp510. Residue 28 is clearly refined as the expected Asp in subunits A and D but has been refined as Ala in subunits B and C where the side-chain density is unclear beyond the β-carbon. The model has crystallographic agreement R and R_{free} factors

*mpllnwsrhmvcltaaglitvptvya*TDTLTRDNGAVVGDNQNSQTAGAQGPVLLQDVQL 60

LQKLQRFDRERIPERVVHARGTGVKGEFTASADISDLSKATVFKSGEKTPVFVRFSSVVH 120

GNHSPETLRDPHGFATKFYTADGNWDLVGNNFPTFFIRDAIKFPDMVHAFKPDPRTNLDN 180

DSRRFDFFSHVPEATRTLTLLYSNEGTPAGYRFMDGNGVHAYKLVNAKGEVHYVKFHWKS 240

LQGIKNLDPKEVAQVQSKDYSHLTNDLVGAIKKGDFPKWDLYVQVLKPEELAKFDFDPLD 300

ATKIWPDVPEKKIGQMVLNKNVDNFFQETEQVAMAPANLVPGIEPSEDRLLQGRVFSYAD 360

TQMYRLGANGLSLPVNQPKVAVNNGNQDGALNTGHTTSGVNYEPSRLEPRPADDKARYSE 420

LPLSGTTQQAKITREQNFKQAGDLYRSYSAKEKTDLVQKFGESLADTLTESKNIMLSYLY 480

KEDRNYGTRVAEVAKGDLSKVKSLAASLKD 510

Fig. 2. Sequence of CatF. The N-terminal 26 residues that are missing are shown in lower case. Specific residues that are referred to in the text are indicated with a residue number below.

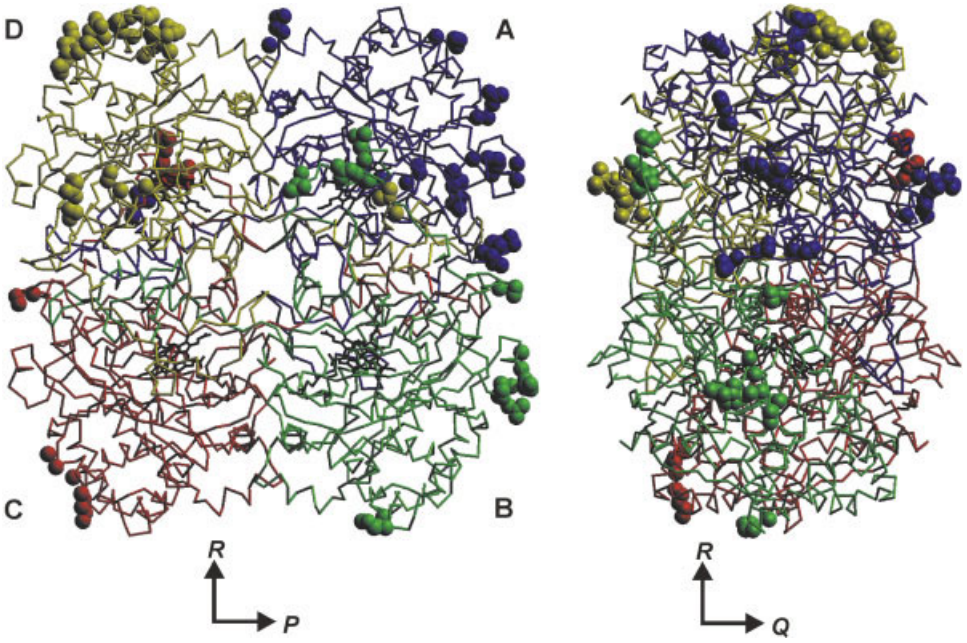

Fig. 3. Two views of the tetramer of CatF rotated by 90° around the R axis. The four subunits are labeled, A–D in the left panel and are colored blue (**A**), green (**B**), red (**C**), and yellow (**D**). The residues involved in crystal contacts are shown as spheres.

of 18.3 and 24.0% for 175,279 reflections in the resolution shell between 1.8 and 21.9 Å.

The average root-mean-square deviation (RMSD) after superimposition of the C^{α} atoms of the four subunits is 0.12 Å (B to C), 0.14 Å (A to D), and 0.16 Å (A–B, A–C, B–D, and C–D). The average temperature factor values for the main- and side-chain atoms of subunits A and D were 11.5 and 14.0, respectively, and for the main- and side-chain atoms of subunits B and C were 19.3 and 21.1, respectively. The crystal contacts are principally on subunits A and D, involving similar residues on both subunits, which may account, at least in part, for the subtle structural differences between the A–D and B–C pairs, the latter pair having a different and smaller subset of contact residues (Fig. 3). The involvement of the N-terminal Thr27 of subunits A and D in crystal contacts may be the reason for the clear definition of these two residues in the electron density, compared to subunits B and C where Thr27 is disordered.

Also a testament to the quality of the data was the ease with which eight errors in the predicted protein sequence were identified. The errors were clearly evident in the electron density, corroborated by the calculation of omit maps and later corroborated in the DNA sequence. These included changing I369 to N, F445 to Y, R459 to K, H468 to L, V478 to Y, E482 to A, H485 to N, and D492 to A

Ramachandran plots confirm that only one residue, Val219, lies outside the energetically favorable regions in

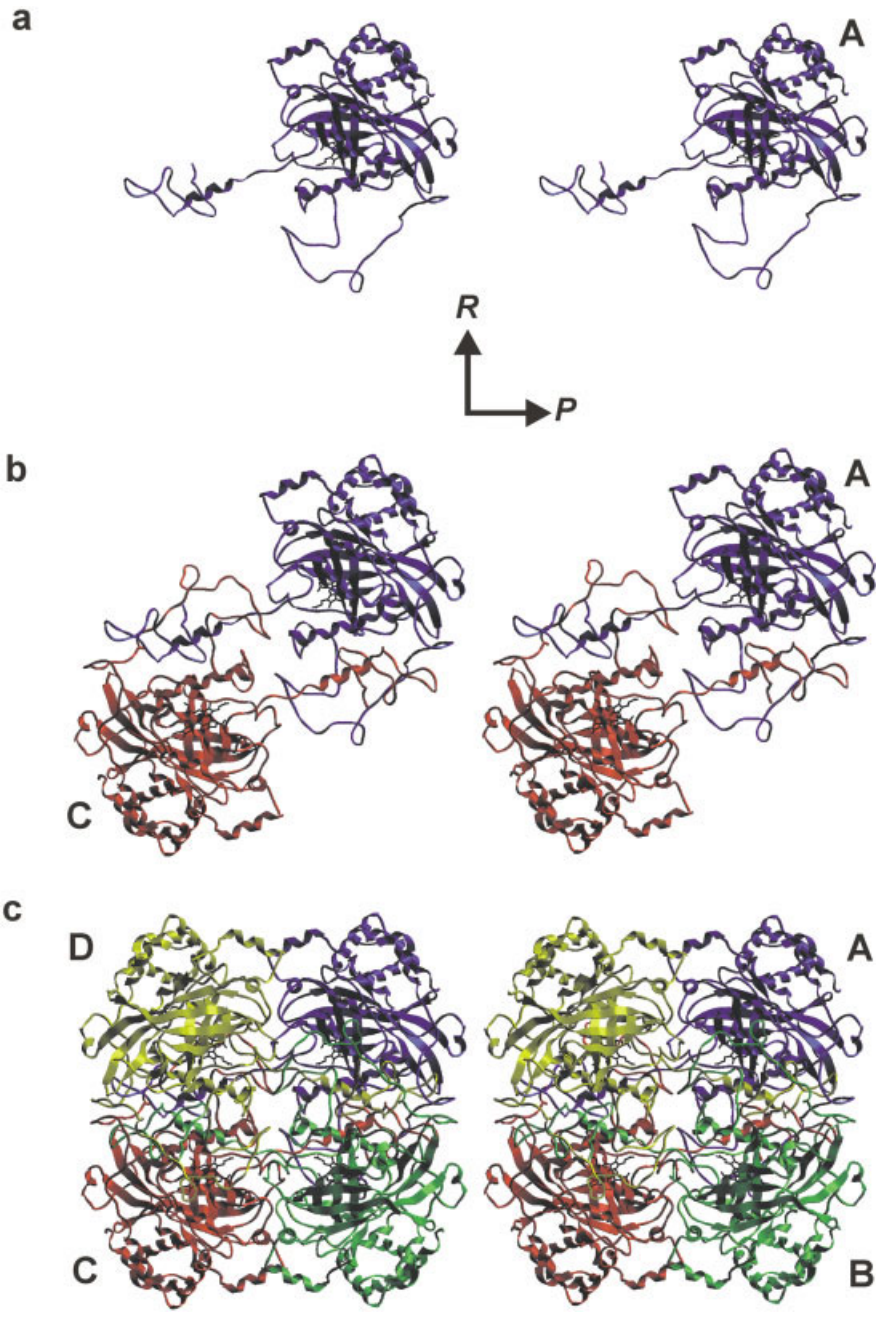

Fig. 4. Stereo views of CatF to illustrate the association of subunits. **a:** Subunit A is shown in blue. **b:** Subunits A and C (in red) associate so that their N-termini are overlapped by a loop on the opposite subunit. **c:** Subunits B (green) and D (yellow) have formed a similar intertwined dimmer, which is associated with the A–C dimer to form the tetramer.

all four subunits. This residue is equivalent to the residues in other catalases (Ser216 in BLC, Ser196 in PMC, Ser217 in HEC, Ser and Ile274 of HPII) that are also found in an energetically unfavorable conformation. In CatA, the equivalent residue is a Gly. Whether the location of these residues adjacent to the heme at the entrance to the lateral channel leading from the heme cavity to the surface of the protein gives them a significant role in catalysis remains to be determined.

Overall Structure

CatF, composed of four identical subunits, is a compact molecule with maximum dimensions of 98, 61, and 98 Å, respectively, along the P, Q, and R axes (Fig. 3). The four regions typical of small subunit catalases, including the N-terminal arm (to Val77), an antiparallel eight-stranded β-barrel (residues His78 to Ala333), an extended wrapping loop (Phe334 to Gln440) and a helical domain at the

TABLE II. RMSD of Cα Atoms of Catalases After Superimposition of the Core 368 Residues

	HPII	PVC	MLC	PMC	CatA	BLC	HEC
CatF	1.30	1.24	1.04	0.99	1.22	1.09	1.09
HPII		0.70	1.29	1.38	1.33	1.28	1.26
PVC			1.25	1.41	1.39	1.33	1.31
MLC				0.85	1.08	0.95	0.95
PMC					1.02	0.83	0.86
CatA						0.77	0.79
BLC							0.33

Sequences were aligned by using ClustalW and any intervening residues in longer sequences were removed to eliminate gaps and noncorresponding residues. The sequence of 368 residues from L28 to T396 were superimposed. The RMSD values were determined by using COMPAR.[23]

TABLE III. Subunit Reconciliation of CatF, HEC, CatA, and HPII

CatF 1M7S	HPII 1GGE	CatA 1A4E	HEC(2.2) 1DGB	HEC(PAA) 1DGF	HEC(AT) 1DGH	HEC(CN) 1DGG
A	A	A	A	A	A	A
B	B	C	B	D	D	D
C	C	B	D	B	B	B
D	D	D	C	C	C	C

carboxy terminus (Ala441 to Asp510) are evident in the monomer structure (Fig. 4).

The general folding of the CatF model is very similar to the folding of the five other small subunit catalases for which structures are available and of the amino terminal 550 residues of the large subunit catalases. The superimposition of the core regions of subunits of HEC, HPII, MLC, PVC, BLC, PMC, and CATA gave an average RMSD for the Cα atoms that ranged from 0.33 to 1.38 Å (Table II). As expected, the smallest RMSD values were for the phylogenetically most closely related sequences, BLC and HEC from clade 3, and the largest values were for the phylogenetically most distant sequences such as CatA from clade 3 and HPII from clade 2, even though the latter was used as the initial model.

The intertwining or "arm exchange" of subunits, another feature unique to catalases, is evident in CatF with a short section of 5 N-terminal residues extending through a loop in the wrapping domain on an adjacent subunit. In this way, the N-termini of subunits A and C are overlapped by C and A, respectively (see the dimer in Fig. 4), whereas the N-termini of B and D are overlapped by D and B, respectively. The intertwined structure of catalases may be one factor contributing to their heat stability, in particular of HPII which is stable to over 80°C,[31] and the very short segment overlapped in CatF is consistent with it being one of the most heat sensitive of catalases.[32]

Subunit naming relative to the conventional P, R, and Q axes has varied among catalase structures with HEC and CatA having different assignments from the group of BLC, HPII, and CatF. To simplify the comparison of subunits in this article, the subunit assignments of HEC and CatA have been rationalized to the system in BLC, HPII, and CatF (Table III). This involves the subunit designations of B and C being reversed for CATA, of C and D being reversed in HEC and a more complex reassignment in HEC-PAA, HEC-CN, and HEC-AT of B–C, C–D, and D–B. With these changes, the arm-exchanged dimers are A with C and B with D in all reported multimeric catalase structures (Fig. 4). This reassignment also reveals that the asymmetry between A and D and B and C in CatF parallels the observed asymmetry among the subunits of HEC whereby subunits A and D (as revised in Table III) have NADPH bound and appear to be less active because they are not modified by 3-aminotriazole.

The region involving residues 121–123 is of interest because it breaks the exact 222 molecular symmetry and introduces minor heterogeneity into the homotetramer. The density clearly shows two alternative conformations for main-chain atoms of Gly121-Asn122-His123 and for the side-chain of Asn122 (Fig. 5). The two conformations of the main-chain involve a 180° flip about the peptide bond joining Asn122 and His123, and the two conformations are complementary in the two subunits, A and D or B and C. The conformation with the carbonyl oxygen pointing toward the opposing subunit cannot simultaneously coexist in the symmetry-related subunit without giving rise to serious steric conflicts. This feature of alternative conformations and breakage of symmetry is detected in most catalases refined with high-resolution data, including PVC, HPII, MLC, and HEC. This element of asymmetry may have functional importance in view of the strong hydrogen bond between the Asn122 side-chain, in one conformation, and the Glu74 side-chain, which is conserved in most catalases as part of the RIPER signature sequence, and of the proximity to the active site His (8.0 Å), the heme (12.4 Å) and the conserved Asp130 (9.2 Å) in the main channel.

Heme Orientation

The heme of catalases can exist in two orientations relative to the active site residues. In the small subunit, clade 3 catalases, including HEC, BLC, PMC, CATA, and MLC, the heme is oriented with the active site His located above ring III [Fig. 6(a)], whereas in the large subunit, clade 2 catalases, including HPII and PVC, the heme is rotated 180° resulting in the active site His being located over ring IV [Fig. 6(b)]. For the subsequent discussion, the clade 2 orientation is named the His-IV orientation, and the clade 3 orientation is named the His-III orientation referring to the active site His and heme ring closest to it. To confirm the orientation of the heme in CatF, an omit map was calculated (with the heme omitted from the model during refinement) and the final $F_o - F_c$ difference Fourier map for the heme region [Fig. 6(c)] revealed that the orientation of the heme in CatF was the same as in HPII, the His-IV orientation.

The only structural difference between the hemes in the two orientations lies in the locations of the two methyl and two vinyl groups on rings I and II, and it is the interactions of these four groups with the protein that must determine the orientation. The residues of CatF in the vicinity of the vinyl and methyl groups of the heme create a matrix of

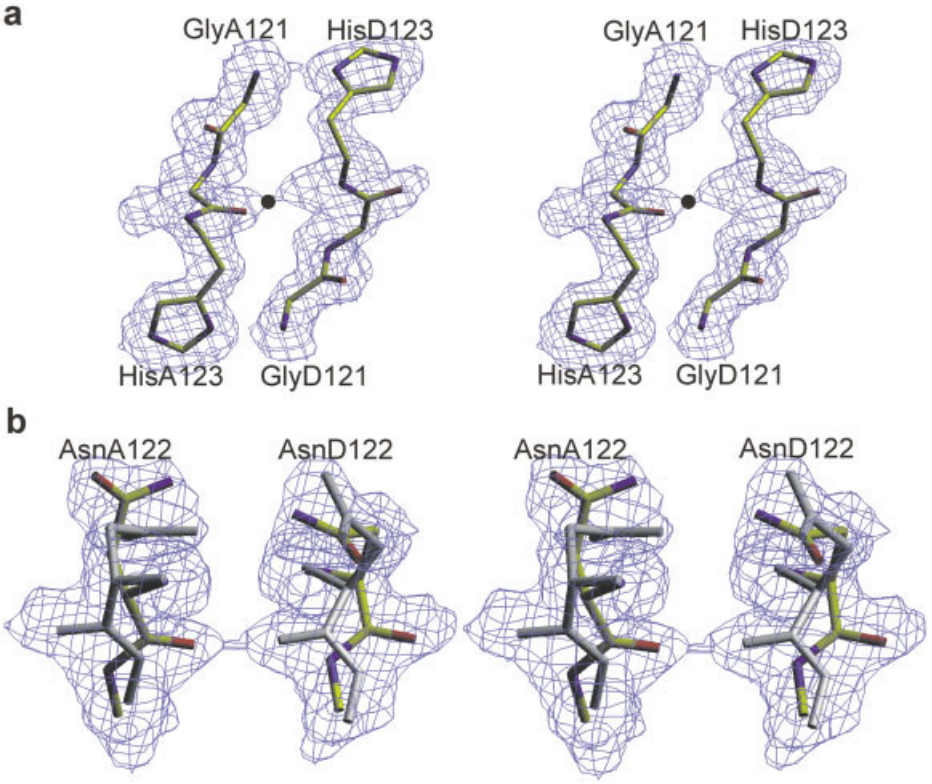

Fig. 5. Stereo views of a portion of the $2F_o - F_c$ electron density map for the region in which there is a break in the 222 symmetry of the whole tetramer. **a:** The map corresponding to the backbone of Gly121-Asn122-His123 in subunits A and D is viewed looking down the twofold axis, indicated by the filled circle. Only one of the two possible conformations of the Asn122 main-chain carbonyl is superimposed on the map, but the location of the carbonyl in the flipped conformation is clear. The side-chain of Asn122 is not shown to enhance the clarity of the map of the main-chain atoms. The side-chain of His123 is included to help visualize the breakage in symmetry caused by the carbonyl. **b:** The map corresponding to the entire Asn122 in subunits A and D is shown. The two alternate conformations are shown, one in elemental colors as in (a) and the second colored in gray. The view is rotated slightly from that in (a) and is not oriented on the axis of symmetry.

favorable van der Waals interactions that would stabilize the heme in the His-IV orientation. Attempting to place the heme in the His-III orientation in CatF results in Leu350 interfering sterically with the vinyl group on ring I of the heme. The equivalent residue is also Leu in HPII and PVC (also with heme in the His-IV orientation), but it is replaced by Met (MLC, PMC, and BLC) or Val (CATA) in the enzymes with heme in the His-III orientation. The alternate exercise of inserting the heme in the His-IV orientation into the HEC structure results in Leu299 (of HEC) interfering with the vinyl group of ring I. The equivalent residue in PMC, MLC, BLC, and CATA is also Leu and all have heme in the His-III orientation. In CatF, Ala301 is the residue equivalent to Leu299 of HEC, and in HPII and PVC, Pro is the equivalent residue, neither of which interfere with the vinyl group.

Alignment of 228 catalase sequences suggests that residues at the equivalent of 301 and 350 in the CatF sequence may be the key determinants in heme orientation. At position 350, 69% (54 of 78) of clade 1, 100% (52 of 52) of clade 2 and 0% (0 of 98) of clade 3 enzymes have Leu, and, at position 301, 0% of clade 1 and clade 2 enzymes, but 80% (78 of 98) of clade 3 enzymes have Leu or Ile. These data suggest that most clade 1 and all clade 2 enzymes will have heme in the His-IV orientation and most clade 3 enzymes will have heme in the His-III orientation. Whether the exceptions can harbor a mixture of heme orientations or if there is a small subset of clade 1 enzymes with heme in the His-III orientation and a small subset of clade 3 enzymes with heme in the His-IV orientation remains to be tested.

NADPH-Binding Site

A number of catalases including the small subunit clade 3 enzymes, BLC, HEC, PMC, and CATA bind NADPH, whereas the large subunit clade 2 enzymes, HPII and PVC, do not. CatF is the first clade 1 enzyme to be characterized, and the question of whether NADPH binds to CatF was of interest. The final $F_o - F_c$ difference Fourier map contains no significant positive density throughout the model. This was of particular interest in the region corresponding to the NADPH-binding site in BLC, PMC, and CATA. The potential NADPH-binding site of subunit A is typical of the sites on the other three subunits in containing clearly distinguishable waters, virtually no positive $F_o - F_c$ density, and no density that could be correlated to a bound NADPH (Fig. 7), making it clear that NADPH is not bound to CatF.

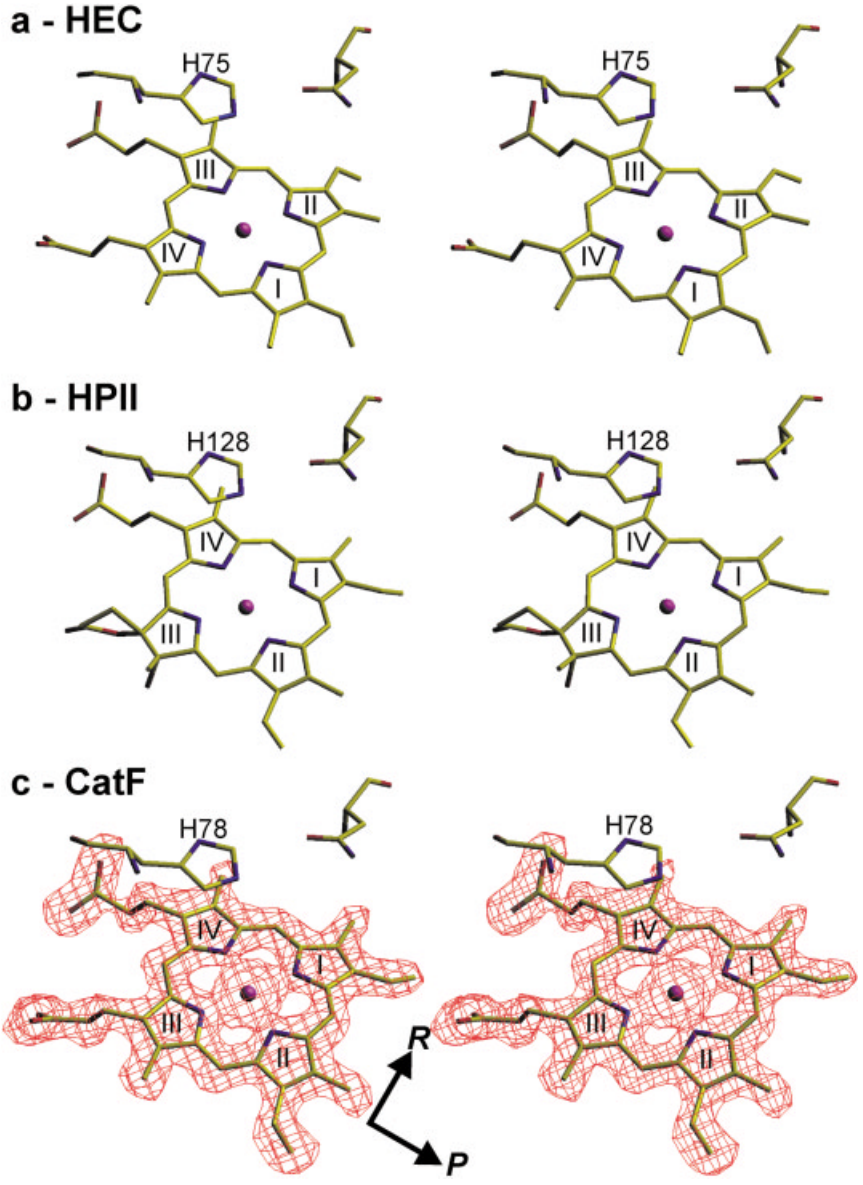

Fig. 6. Stereo view of the heme orientation in HEC (**a**) and HPII (**b**). A stereo view of the $F_o - F_c$ omit map of the heme pocket in subunit A of CatF generated without heme in the model during refinement is shown in red in (**c**). The heme is superimposed into the density to illustrate clearly the similarity to the orientation in HPII. The active site His and Asn are indicated. In HEC, the active site His is located above ring III of the heme (His-III orientation), and in HPII the active site His is located above ring IV of the heme (His-IV orientation). The axis of rotation, indicated in (c), is rotated slightly relative to the view in other figures.

Analysis of the NADPH-binding pocket of HEC, BLC, and CatA suggests that His193, Arg202, Val301, and His304 (BLC numbering) are important for NADPH binding and may even serve as a signature for NADPH-binding ability. In CatF, the equivalent four residues are Arg196, Glu205, Ile304, and Asp307, and each change would interfere with NADPH binding either through electrostatic repulsion (Arg196 and Glu205), direct steric interference (Arg196 and Ile304), or loss of favorable contacts (Asp307). The combination of these changes clearly precludes NADPH binding to CatF. Analysis of the equivalent residues in other clade I catalase sequences strongly suggests that clade 1 catalases generally do not bind NADPH. Most clade 1 enzymes lack the equivalent of His193 and Val301, and all lack the equivalent of Arg202 and His304. Large subunit catalases such as HPII also do not bind NADPH, although they retain two of the signature residues (His251 and Arg260). However, HPII has the remaining two signature residues changed (Val to Leu359 and His to Glu362), and probably even more importantly, has a segment of the extended C-terminal domain projecting into the region.

Prevention of inactive compound II formation is the role attributed to NADPH in catalases.[1,33] By extension, it has

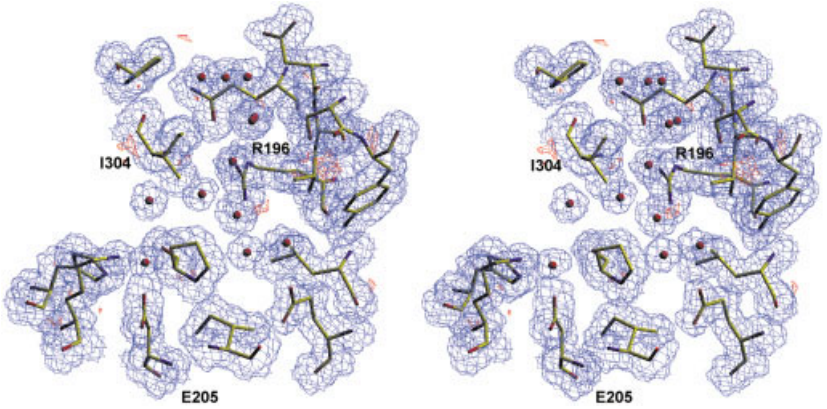

Fig. 7. Stereo view $2F_o - F_c$ (blue) and $F_o - F_c$ (red) electron density map of the potential NADPH-binding site on CatF. The $2F_o - F_c$ density was modeled at $\sigma = 1.0$ and the $F_o - F_c$ density at $\sigma = 3.0$. The model is included showing the locations of the waters as red spheres.

been argued that NADPH does not bind to HPII because the enzyme does not form compound II.[34] A similar argument may hold for CatF, although a study of compound II formation in CatF has not been attempted. It is also intriguing to speculate whether there is a mechanistic link between the His-III heme orientation and the binding of NADPH in clade 3 enzymes.

The observations of NADPH binding and the His-III heme orientation only in clade 3 enzymes is consistent with a recent phylogenetic analysis[35] of >225 catalase sequences that concluded that clade III enzymes evolved considerably later than the clade 1 and 2 enzymes. The model arising from that analysis is that clade 1 enzymes evolved from clade 2 enzymes soon after the emergence of the latter. A gene duplication event and subsequent loss of sequence at the 5′ and 3′ ends of the gene were likely involved. The clade 3 enzymes evolved after a second gene duplication event much later, resulting in clade 1 and 2 enzymes being more closely related than they are to clade 3 enzymes.

Channel Structure

CatF presents a channel structure typical of other small subunit catalases including the funnel-shaped main or perpendicular channel, the bifurcated lateral channel, and the channel leading to the central cavity in each subunit (Fig. 8). The fact that none of the channels appear continuous in the diagram is a result of the probe constraints used in the program VOIDOO, which were set to the approximate size of a water molecule, thereby accentuating the narrowness of the channels at the hydrophobic constrictions at the entrances to the heme cavity created by Val118 in the perpendicular channel, Val219 in the lateral channel, and Val76 in the cavity channel. A portion of subunit A in Figure 8 is rotated slightly to bring the three channels clearly into view and to show the orientation of the view used in Figures 9–11.

Solvent Location in the Perpendicular Channel of Catalases

The location and occupancy of solvent molecules were largely identical among all four subunits of CatF except in the perpendicular channel. For example, the waters labeled I, II, and III in Figure 9 are present in all subunits of all catalases and are included for reference. Similarly, the water labeled 6 at the top of the channel is present in all four subunits of CatF and most other catalases. The differences in water occupancy lie in the lower (as viewed with the heme at the bottom) 12 Å portion of the channel extending from the heme to just above the conserved Asp130. In this region, solvent is found at five locations in the D subunit (waters 1–5), four in subunit A, three in B, and two in C (Fig. 9 and Table IV). The asymmetry of water occupancy among the subunits of CatF (more waters in subunits A and D, fewer in subunits B and C) parallels the greater relatedness of A–D and B–C noted earlier (Table I). Additional asymmetry is evident in the distances between the heme iron and water 1, which average 2.61 Å for subunits A and D and 2.15 Å or subunits B and C (also evident in Fig. 9).

Given the importance of solute flow in the catalatic mechanism, a comparison of the occupancy and locations of water in the perpendicular channels of catalases was undertaken. The solvent distribution has been refined in all four subunits of three catalase structures, including HPII, HEC, and CatF, making an intersubunit comparison possible. In addition, the structures of CatA complexed with azide (CatA-azide) and HEC after treatment with peroxyacetic acid (HEC-PAA), cyanide (HEC-CN), and aminotriazole (HEC-AT) also have the solvent distribution refined in all four subunits. A proportionally similar number of waters have been refined in the single subunit of MLC, and a smaller number of waters have been refined into BLC and PMC. As with CatF, the solvent distribution is similar throughout subunits of the same enzyme, including both the lateral and cavity channels, and within the constraints of sequence differences, is similar even among the various enzymes. By contrast, the differences in solvent distribution observed in the heme pockets and lower portions of the main or perpendicular channel (Table IV) of different subunits of CatF are also evident among subunits of other enzymes (Fig. 10).

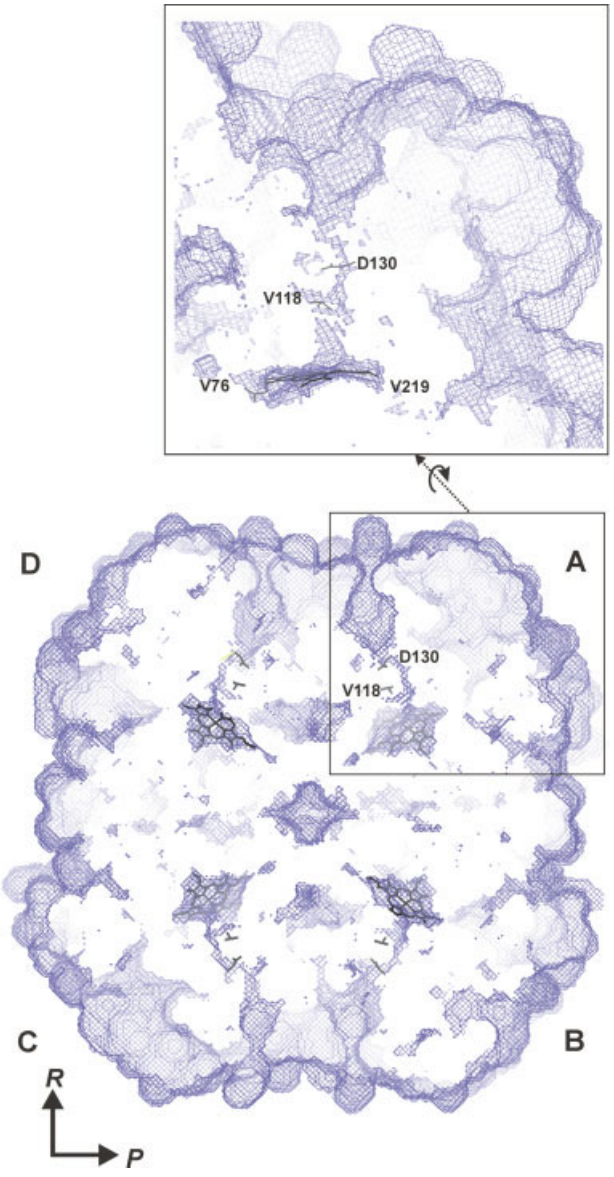

Fig. 8. View of channels leading to the four active sites in the CatF tetramer. The accessibility surfaces, determined by using VOIDOO,[25] are shown as a blue lattice and are a reflection of what parts of the protein are accessible to a molecule about the size of water. The cross section or slab is chosen to illustrate most clearly the perpendicular or main channels. The portion of subunit A in the box is rotated slightly and enlarged to illustrate clearly all three channels in subunit A and to provide the orientation that is maintained in Figures 9, 10, and 11. The locations of the conserved Asp130 and Val118 in the perpendicular channel are indicated as are the locations of Val76 and Val219 at the entrances to the central cavity and lateral channels, respectively.

The highly conserved protein structure in the heme pocket seems to preclude a steric explanation for CatF, HPII, and MLC having relatively high occupancy at water 1 about 2.5 Å from the heme iron and HEC having relatively low occupancy at water 1b, 5.6 Å from the heme iron. The inductive effect of residues that interact with the heme through its propionate groups and that differ among catalases may provide at least a partial explanation. In HEC, Tyr370 has a relatively high, calculated electronic coupling with the heme iron[36] and can directly transfer electrons for reduction of compound I via Asp335, His362, and the heme propionate.[13] The inductive effect of Tyr370 and Asp335 may also decrease the effective positive charge on the heme iron and lower the affinity for water (Fig. 11). In CatF, this inductive effect would be absent or much reduced because the His-Asp-Tyr triad is absent, replaced with Gln362, Gly370, and Ala335, and affinity for water is higher (Fig. 11). HPII and MLC, like CatF, lack the His-Asp-Tyr triad and have glutamine alone interacting with the propionate of the heme. PMC represents an intermediate state with water 1 located 3.5 Å from the heme iron. This may be a result of a slightly larger inductive effect arising from His341, associated with His349, interacting with the heme propionates (Fig. 11).

Water 2 is the only solvent position that is fully occupied in all subunits of all unmodified catalases. The only exceptions are HEC-AT and CATA-azide where the inhibitor complex overlaps the site of water 2. Water 2 is hydrogen bonded to the catalytic His (His 78 in CatF at an average distance of 2.73 Å) and is weakly associated with the catalytic Asn (Asn160 in CatF at an average distance 3.41 Å).

Water 3, adjacent to the conserved Val in the hydrophobic portion of the channel, is a low-occupancy water present in only one subunit of CatF and HPII. Only HEC-PAA has it present in all four subunits. The reason for the low occupancy at water 3 may be that it can form only one hydrogen bond interaction with water 4, which, in turn, is present in only two subunits of CatF. HEC-CN and HEC-AT present a different solvent pattern in this portion of the channel, with waters 3 and 4 replaced by waters 3b and 4b,which create a hydrogen-bonded matrix of waters 1b, 2, 3b, and 4b, which link the conserved residues, His75 and Asp128. This matrix is broken in native HEC, which lacks water 3b in all subunits. The shift of water from 4 to 4b in HEC and MLC is a result of Ala (117 and 102, respectively) replacing the bulkier Val119 of CatF and Gln170 of HPII.

Water 5 is present in only three subunits of both CatF and HEC and is absent in MLC, a surprising situation given its close interaction with the highly conserved Asp130/128. In HPII, the situation in the upper part of the channel differs considerably because the C-terminal domain constricts the channel and increases its effective length. In the vicinity of the conserved Asp181, the constriction shifts water 5 to 5b where it is in closer contact with the conserved Asp181 and present in all four subunits.

These data might be interpreted to suggest a correlation between less water in the channel and higher activity. For example, unmodified HEC has less water in the perpendicular channels of its subunits than CatF, which in turn has less water than HPII, and the maximum turnover rates decrease with increasing water (548,000/s for HEC, 248,000/s for CatF, and 151,000/s for HPII[32]). Furthermore, the inactivated forms of HEC contain more waters in the channel than the active form, although this conclusion must be considered in light of the poorer refinement of the untreated HEC. Whether this correlation is real or simply

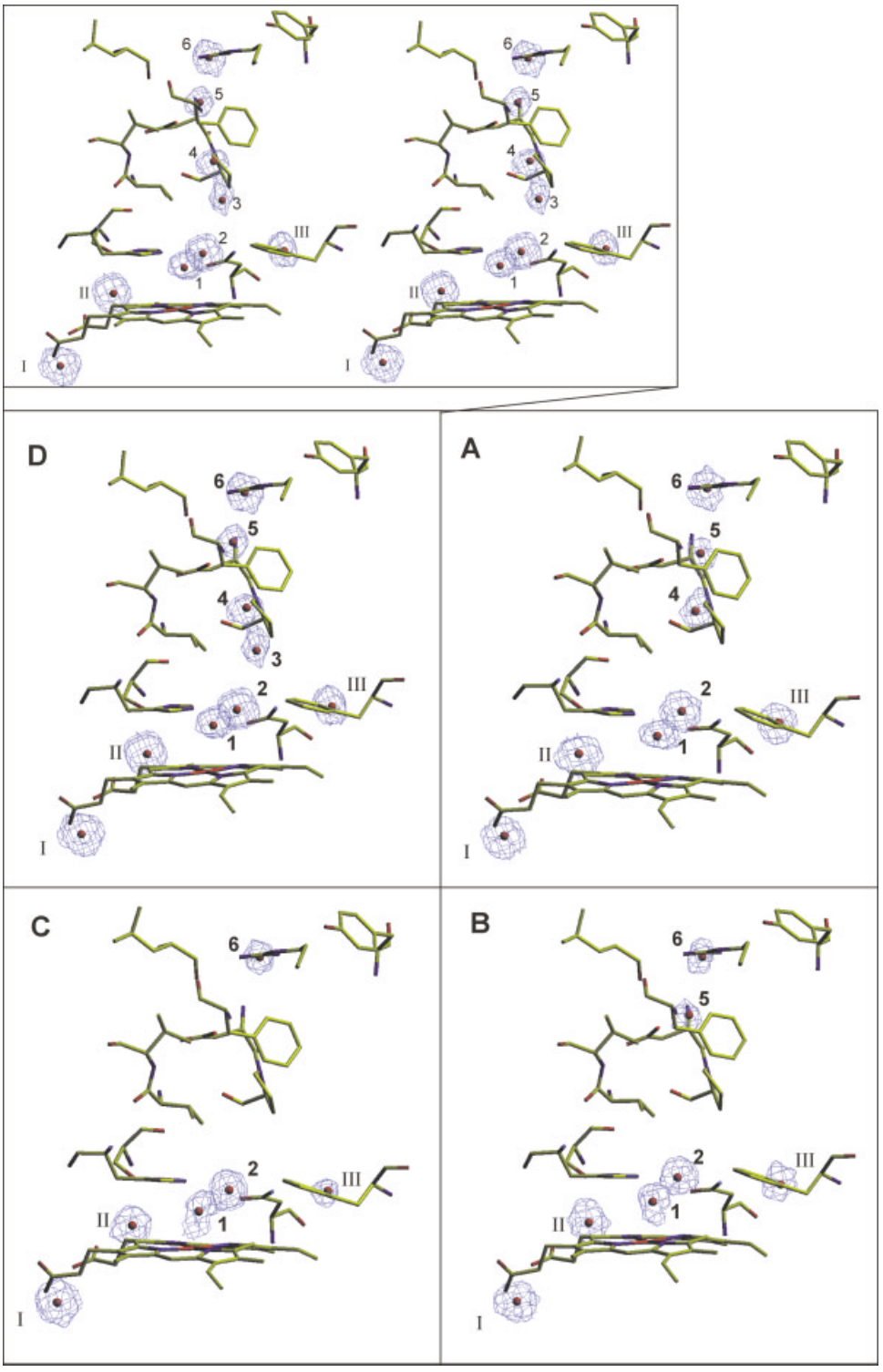

Fig. 9. View of solvent molecules in the main or perpendicular channel of CatF. In all panels, the $2F_o - F_c$ electron density corresponding to solvent is shown in blue, modeled waters as red spheres, and surrounding residues with standard atomic colors. The electron density was modeled at $\sigma = 1.0$. The orientation shown in Figure 8 is maintained for all four subunits with the heme group at the bottom and the channel extending upward to the surface of the protein near water 6. **A–D:** The channels for each of the subunits A–D. **E:** A stereo view of the channel from subunit D.

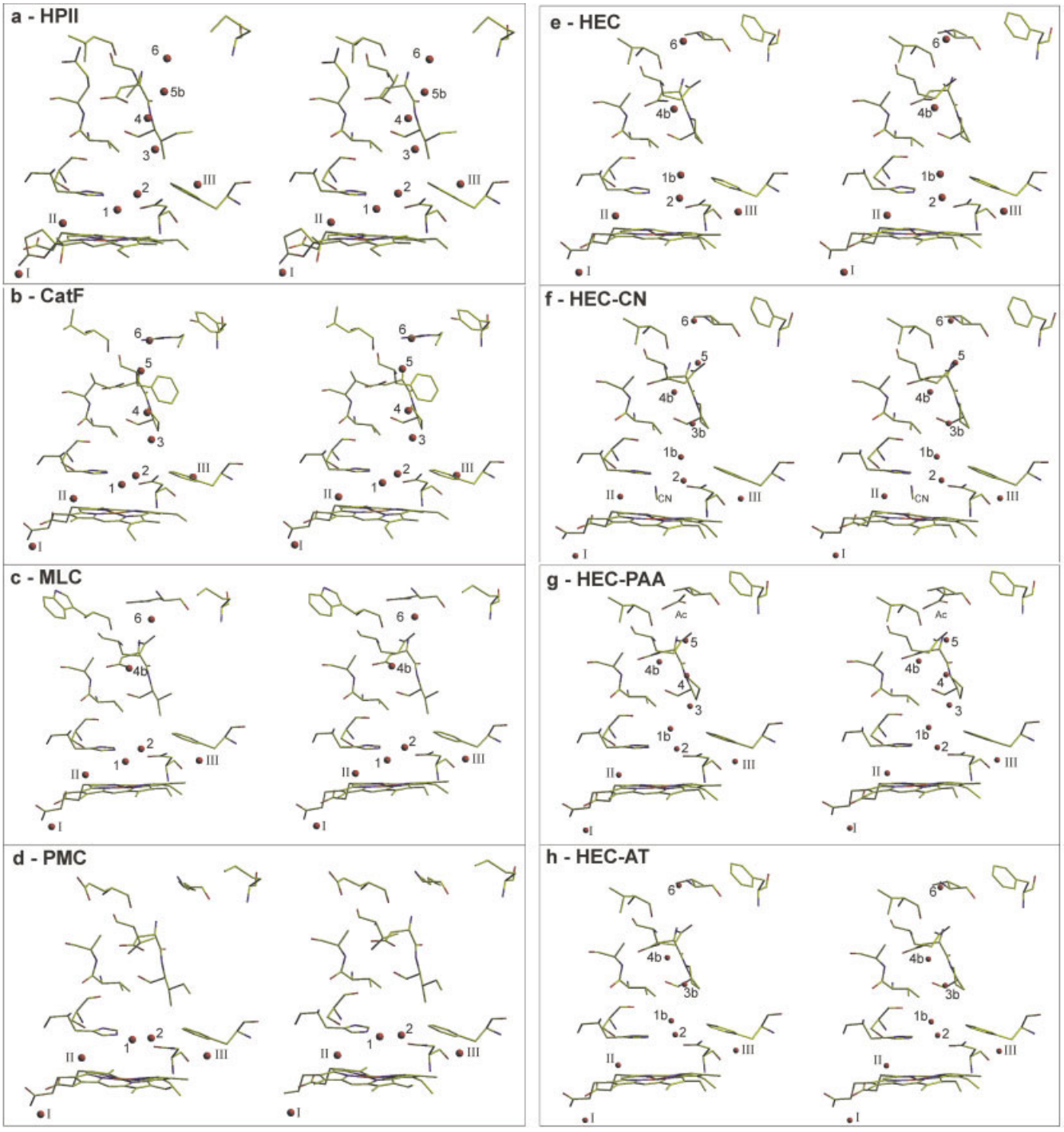

Fig. 10. Stereo views of water molecules in the main or perpendicular channel of various catalases. The same orientation as in Figures 8 and 9 is maintained. The identity of the catalase is indicated in each panel. Where there is a variation in the number of waters in different subunits (Table IV), the subunit with the largest number was chosen. Waters are shown as red spheres and surrounding residues with standard atomic colors. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

a result of differing data collection and refinement techniques will have to await the appearance of further catalase structures.

Despite a century of study, which has presented detailed information about the structure of eight different catalases, a full understanding of the catalatic mechanism continues to be elusive. This work has revealed significant differences among enzymes in solvent occupancy in the perpendicular access channel to the active site but has failed to explain substantial differences in turnover rates that exist. Associated with this question is one raised by the observation that catalases are capable of attaining

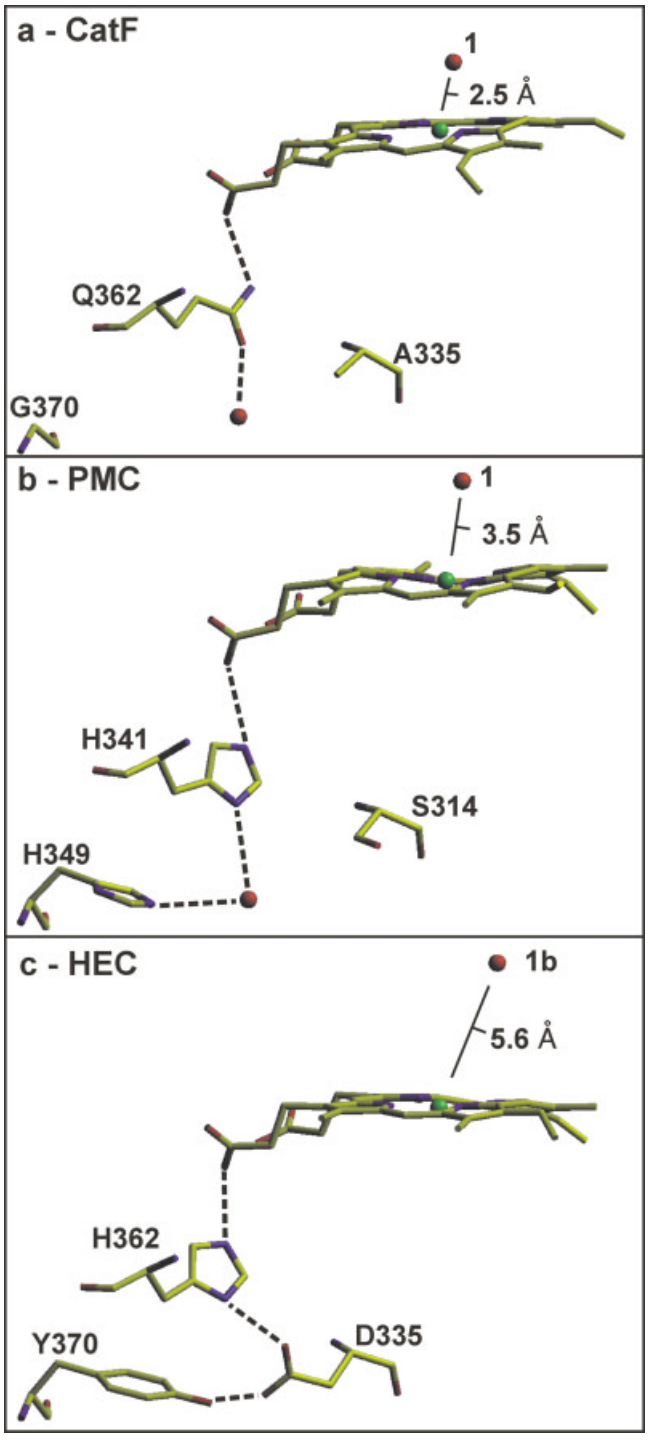

Fig. 11. Organization of the residues on the proximal side of the heme in CatF (**a**), PMC (**b**), and HEC (**c**) that correspond to the His362-Asp335-Tyr370 triad that forms a possible route for electron transfer from the Tyr to the heme through the propionate group of the heme. The locations of water 1 (or 1b in HEC) and their distance from the heme iron are indicated. Hydrogen bonds are indicated by dashed lines. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

TABLE IV. Water Occupancy as B-factor (Å^2) in the Perpendicular Channel of Catalase Subunits

No.	A	B	C	D
A CatF (1.8Å) Av. B-factor: 16.4 (protein), 29.0 (water)				
1	24.7	42.6	25.2	29.0
2	22.0	33.7	29.5	22.2
3	—	—	—	36.3
4	24.5	—	—	38.7
5	50.6	44.2	—	36.8
6	30.1	44.3	—	26.1
I	10.7	13.2	15.6	9.2
II	4.5	24.8	14.8	5.0
III	14.8	25.8	30.9	16.1
B HEC (2.2Å; 1DGB) Av. B factor: 62.9 (protein), 67.7 (water).				
1b	90.6	—	—	—
2	85.9	56.2	60.9	59.9
4b	81.0	73.0	65.6	66.4
6	65.5	51.1	58.8	56.2
I	58.1	54.7	57.3	57.3
II	58.9	55.9	56.5	53.7
III	55.4	68.4	57.4	60.1
E HEC-AT (2.0Å; 1DGH) Av. B factor: 34.7 (protein), 45.6 (water)				
1b	58.8	—(AT)	—(AT)	60.3
2	44.4	—(AT)	—(AT)	43.9
3b	56.3	—	63.0	62.2
4b	38.9	44.5	48.3	57.6
5	—	58.5	—	—
6	42.2	43.4	48.5	45.4
I	38.3	29.6	27.3	39.4
II	31.1	39.7	31.9	31.8
III	41.6	42.4	49.9	46.2
F HPII (1.9 Å; 1GGE) Av. B factor: 9.2 (protein), 17.7 (water)				
1	40.0	28.1	42.7	—
2	15.3	17.4	16.3	12.9
3	—	—	32.3	—
4	24.0	20.5	22.5	29.9
5b	30.7	41.2	29.0	27.9
6	10.7	11.6	14.8	10.4
I	3.7	7.1	4.8	6.2
II	5.1	6.1	5.0	7.1
III	4.0	8.9	6.4	8.1
G CATA-azide (2.4 Å; 1A4E) Av. B factor: 32.1 (protein), 39.6 (water)				
4b	—	81.1	—	83.8
5	47.0	43.2	—	52.2
6	—	—	47.8	53.2
I	22.3	22.6	15.3	17.0
II	15.8	25.9	20.3	20.8
III	36.1	21.3	19.9	26.3
H MLC (1.5 Å; 1HBZ) Av. B factor: 11.1 (protein), 27.0 (water)				
1	14.2			
2	17.3			
4b	15.6			
6	35.2			
I	9.0			
II	7.3			
III	9.9			
I PMC (2.2 Å; lcae) Av. B factor: 25.8 (protein), 31.6 (water)				
1	36.6			
2	49.7			
I	21.0			
II	27.7			
III	39.4			

The positions associated with the water numbering are shown in Figures 9 and 10.

extremely high turnover rates, on the order of 10^5–10^6 per second but only at concentrations of substrate hydrogen peroxide far higher than are likely to be experienced in vivo. What advantages accrue in having an enzyme with maximum catalytic rates at apparently nonphysiological substrate concentrations? Finally, for this discussion, what is the relationship between solvent location in the access

channels and the ability of substrate to reach the active site? At least two molecular dynamics modeling have addressed this issue and agree that the perpendicular channel is the main route for substrate access.[37,38] Unfortunately, they present different conclusions about the need and identity for a unique channel to exhaust the products, water and oxygen. It remains to experimentally differentiate among the options presented.

ACKNOWLEDGMENTS

This work was supported by a grant from DGICYT to I.F. and OGP9600 from the Natural Sciences and Engineering Research Council of Canada (NSERC) to P.C.L. We thank R. Perez for help with data collection. The cooperation of Dr. K. Standing and W. Ens in arranging use of the mass spectrometer in the TOF Laboratory of the Department of Physics, University of Manitoba is also appreciated. M.S. was supported by a IBERDROLA Fellowship.

REFERENCES

1. Nicholls P, Fita I, Loewen PC. Enzymology and structure of catalases. Adv Inorg Chem 2001;51;51–106.
2. Murthy MRN, Reid TJ, Sicignano A, Tanaka N, Rossmann MG. Structure of beef liver catalase. J Mol Biol 1981;152:465–499.
3. Fita I, Silva AM, Murthy MRN, Rossmann MG. The refined structure of beef liver catalase at 2.5 A resolution Acta Crystallogr 1986;B 42:497–515.
4. Vainshtein BK, Melik-Adamyan WR, Barynin VV, Vagin AA, Grebenko AI. Three-dimensional structure of the enzyme catalase. Nature1981; 293:411–412.
5. Vainshtein BK, Melik-Adamyan WR, Barynin VV, Vagin AA, Grebenko AI, Borisov VV, Bartels KS, Fita I, Rossmann MG. Three-dimensional structure of catalase from *Penicillium vitale*, at 2.0 Å resolution. J Mol Biol 1986;188:49–61.
6. Murshudov GN, Melik-Adamyan WR, Grebenko AI, Barynin VV, Vagin AA, Vainshtein BK, Dauter Z, Wilson K. Three-dimensional structure of catalase from *Micrococcus lysodeikticus* at 1.5Å resolution. FEBS Lett 1982;312:127–131.
7. Gouet P, Jouve HM, Dideberg O. Crystal structure of *Proteus mirabilis* PR catalase with and without bound NADPH. J Mol Biol 1995;249:933–954.
8. Bravo J, Verdaguer N, Tormo J, Betzel C, Switala J, Loewen PC, Fita I. Crystal structure of catalase HPII from *Escherichia coli*. Structure 1995;3:491–502.
9. Bravo J, Maté MJ, Schneider T, Switala J, Wilson K, Loewen PC, Fita I. Structure of catalase HPII from *Escherichia coli* at 1.9 Å resolution. Proteins 1999;34:155–166.
10. Berthet S, Nykyri L, Bravo J, Maté MJ, Berthet-Colominas C, Alzari PM, Koller F, Fita I. Crystallization and preliminary structural analysis of catalase-A from *Saccharomyces cerevisiae*. Protein Sci1997; 6:481–483.
11. Maté MJ, Zamocky M, Nykyri LM, Herzog C, Alzari PM, Betzel C, Koller F, Fita I. Structure of catalase-A from *Saccharomyces cerevisiae*. J Mol Biol 1999;286:135–139.
12. Ko T-P, Safo MK, Musayev FN, Di Salvo ML, Wang C, Wu S-H, Abraham DJ. Structure of human erythrocyte catalase. Acta Crystallogr 1999;D 56:241–245.
13. Putnam CD, Arvai AS, Bourne Y, Tainer JA. Active and inhibited human catalase sturctures: ligand and NADPH binding and catalytic mechanism. J Mol Biol 1999;296:295–309.
14. Klotz MG, Klassen GR, Loewen PC. Phylogenetic relationships among prokaryotic and eukaryotic catalases. Mol Biol Evol 1997; 14:951–958.
15. Ma JF, Ochsner UA, Klotz MG, Nanayakkara VK, Howell ML, Johnson Z, Posey JE, Vasil ML, Monaco JJ, Hassett DJ. Bacterioferritin A modulates catalase A (KatA) activity and resistance to hydrogen peroxide in *Pseudomonas aeruginosa*. J Bacteriol 1999; 181:3730–3742.
16. Carpena X, Perez R, Ochoa WF, Verdaguer N, Klotz MG, Switala J, Melik-Adamyan W, Fita I, Loewen PC. Crystallization and preliminary X-ray analysis of clade I catalases from *Pseudomonas syringae* and *Listeria seeligeri*. Acta Crystallogr 2001;D57:1184–1186.
17. Krutchinsky AN, Chernushevich IV, Spicer VL, Ens W Standing KG. A collisional damping interface for an electrospray ionization time-of-flight mass spectrometer. J Am Soc Mass Spectrom 1998;9: 569–579.
18. Verentchikov AN, Ens W, Standing KG. Reflecting time-of-flight mass spectrometer with an electrospray ion source and orthogonal extraction. Anal Chem 1994;66:126–133.
19. Wilm M, Mann M. analytical properties of the nanelectrospray ion source. Anal Chem 1996;68:1–8.
20. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol 1996;276:307–326.
21. Navaza J. AMoRe: an automated package for molecular replacement. Acta Crystallogr 1994;A50:157–163.
22. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr 1997;D53:240–255.
23. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. Acta Crystallogr 1994;A50: 760–763.
24. Jones TA, Zou JY, Cowan SW, Kjeldgaard M. Improved methods for building protein models in electron density maps. Acta Crystallogr 1991;A47:110–119.
25. Kleywegt GJ, Jones TA. Detection, delineation, measurement and display of cavities in macromolecule structures. Acta Crystallogr 1994;D50:178–185.
26. Maté MJ, Sevinc MS, Hu B, Bujons J, Bravo J, Switala J, Ens W, Loewen PC, Fita I. Mutants that later the covalent structure of catalase hydroperoxidase II from *Escherichia coli*. J Biol Chem 1999;274:27717–27725.
27. Evans S. SETOR: hardware lighted three-dimensional solid model representations of macromolecules. J Mol Graphics 1993;11:134–138.
28. Klotz MG, Kim YC, Katsuwon J, Anderson AJ. Cloning, characterization and phenotypic expression in *Escherichia coli* of *catF*, which encodes the catalytic subunit of catalase isozymer CatF of *Pseudomonas syringae*. Appl Miocrobiol Biotechnol. 1995;43:656–666.
29. von Heijne G, Abrahmsen L. Species specific variation in signal peptide design. Implications for protein secretion in foreign hosts. FEBS Lett 1989;244:439–446.
30. Sevinc MS, Switala J, Bravo J, Fita I, Loewen PC. Truncation and heme pocket mutations reduce production of functional catalase HPII in *Escherichia coli*. Protein Eng 1998;11:549–555.
31. Switala J, O'Neil JO, Loewen PC. Catalase HPII from *Escherichia coli* exhibits enhanced resistance to denaturation. Biochemistry 1999;38:3895–3901.
32. Switala J. Loewen PC. Diversity of properties among catalases. Arch Biochem Biophys. 2002;401:145–154.
33. Kirkman HN, Rolfo M, Ferraris AM, Gaetani GF. Mechanisms of protection of catalase by NADPH. J Biol Chem 1999;274:13908–13914.
34. Hillar A, Nicholls P. A mechanism for NADPH inhibition of catalase compound II formation. FEBS Lett 1992;314:179–182.
35. Klotz MG, Loewen PC. The molecular evolution of catalatic hydroperoxidases: evidence for multiple lateral transfer of genes between bacteria and from bacteria into eukaryota. Eukaryotic Cell. In review.
36. Olson LP, Bruice TC. Electron tunneling and ab initio calculations related to the one-electron oxidation of NAD(P)H bound to catalase. Biochemistry 1995;34:7335–7347.
37. Kalko SG, Gelpi JL, Fita I, Orozco M. Theoretical study of the mechanisms of substrate recognition by catalase. J Am Chem Soc 2001;123:9665–9672.
38. Amara P, Andreoletti P, Jouve HM, Field MJ. Ligand diffusion in the catalase from Proteus mirabilis: a molecular dynamics study. Protein Sci 2001;10:1927–1935.