Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Crystallization and preliminary X-ray analysis of the hydroperoxidase I C-terminal domain from *Escherichia coli*

Xavier Carpena,[a] Alba Guarné,[a] Juan C. Ferrer,[b] Pedro M. Alzari,[c] Ignacio Fita[a] and Peter C. Loewen[d]*

[a]Institut de Biología Molecular de Barcelona CSIC, Jordi-Girona 18-26, 08034 Barcelona, Spain, [b]Department de Bioquímica i Biologia Molecular, Facultat de Química, Universitat de Barcelona, Martí i Franquès 1, 08028, Barcelona, Spain, [c]Unité de Biochimie Structurale, Institut Pasteur, 25 Rue du Dr Roux, 75724 Paris CEDEX 15, France, and [d]Department of Microbiology, University of Manitoba, Winnipeg, Manitoba R3T 2N2, Canada

Correspondence e-mail: peter_loewen@umanitoba.ca

Received 12 November 2001
Accepted 4 March 2002

Hydroperoxidases (HP) are normally large haem-containing bifunctional enzymes capable of acting as both catalases and peroxidases. The C-terminal domain of HPI from *Escherichia coli* (KatG), extending from residue Tyr422 to Leu726, was found to be resistant to trypsin proteolysis. The segment of *katG* encoding this domain was cloned and overexpressed to produce a haemless protein that is soluble even at concentrations above 30 mg ml^{-1}. This protein shows a 25% sequence identity with cytochrome *c* peroxidase (*CCP*) from *Saccharomyces cerevisae*, despite lacking the characteristic catalytic and iron-binding residues. Crystals from this protein were grown in 0.6 *M* sodium citrate buffered to pH 7.5 with HEPES by the hanging-drop vapour-diffusion method at 293 K. These crystals diffracted beyond 2.0 Å resolution and belong to space group $P2_12_12_1$, with unit-cell parameters a = 84.2, b = 98.7, c = 302.8 Å. Three pseudo-origin peaks in the Patterson maps indicate an unusual packing compatible with the presence of three molecules in the crystal asymmetric unit and a solvent content of about 80% by volume.

1. Introduction

The bifunctional catalase–peroxidases are found in a wide variety of organisms including bacteria, archaebacteria and lower eukaryotes (Nicholls *et al.*, 2001). They are protective enzymes that degrade hydrogen peroxide either catalatically ($2H_2O_2 \rightarrow 2H_2O + O_2$) or peroxidatically ($H_2O_2 + 2AH \rightarrow 2H_2O + 2A^{\cdot}$), thereby preventing hydroxyl radical induced cellular damage. The rapidity of the catalatic reaction results in the catalase–peroxidases appearing to be predominantly catalases, despite their sequences suggesting a close relationship to the plant peroxidases. Indeed, the close relationship to the plant peroxidases was illustrated in the Trp105Phe variant of *E. coli* HPI (hydroperoxidase I), where the single-residue change converted the enzyme to a predominant peroxidase, with the catalatic activity decreasing 1000-fold and the peroxidatic activity increasing threefold (Hillar *et al.*, 2000).

The catalase–peroxidases gained significant notoriety in 1992 when it was confirmed that mutation of *katG*, encoding the *Mycobacterium tuberculosis* catalase peroxidase, imparted isoniazid resistance (Zhang *et al.*, 1992). Given the importance of isoniazid as an anti-tuberculosis drug and the prevalence of *katG*-induced resistance, interest in determining the structure of KatG grew, with several groups worldwide attempting to solve its structure. Unfortunately, attempts to crystallize a number of different catalase–peroxidases were without success until very

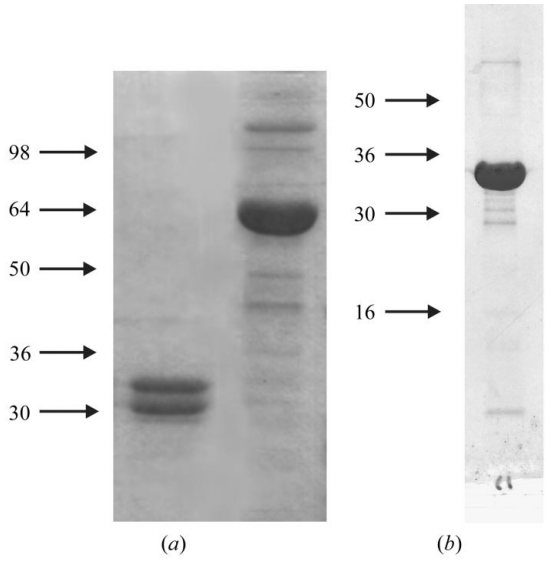

Figure 1
HPI was treated with trypsin as described in §2 and was separated on a 10% SDS–polyacrylamide gel (*a*). (*b*) HPI second domain purified by passage through a Ni–agarose column. The arrows indicate the location of size markers (values are in kDa).

recently, when the crystallization of the enzymes from *Haloarcula* (Yamada *et al.*, 2001) and *Synechococcus* (Wada *et al.*, 2002) were reported.

Catalase–peroxidase subunits appear to have arisen through a gene duplication and fusion event, with the two domains sharing sequence similarity (Welinder, 1991). The N-terminal domain is the active domain containing haem and active-site residues that when modified affect enzyme activity. The C-terminal domain has less sequence similarity, does not seem to bind haem and does not have the well conserved active-site motif characteristic of peroxidases. The haem occupancy for most catalase–peroxidases is only partial, with approximately one haem per two subunits contained in a mixture of dimers or tetramers with one, two and three haems (Hillar *et al.*, 2000). In this paper, we report that the two domains can be separated proteolytically and that the carboxy-domain appears to be the more stable domain. Because the sequence similarity between the N- and C-terminal domains suggested the possibility of some structural relatedness between the domains, a truncated clone of HPI (KatG) from *E. coli* encoding the C-terminal 305 residues was constructed and the expressed protein crystallized for diffraction analysis.

2. Experimental results

2.1. Domain characterization and expression

After many unsuccessful attempts to crystallize HPI (KatG) from *E. coli*, small crystalline aggregates were observed after partial proteolysis of the protein. Systematic trypsin digestions, conducted at 298 K in 45 m*M* Tris–HCl pH 8.0 buffer at a 1/100 ratio of trypsin to HPI for up to 1 h, produced two well defined protein bands on SDS–polyacrylamide gel of approximately 32 and 34 kDa (Fig. 1). Amino-terminal sequencing revealed that both protein fragments started at residue Tyr422, suggesting the existence of a compact protease-resistant globular C-terminal domain in KatG.

The portion of *katG* encoding Tyr422–Leu726 (the C-terminal residue of KatG), named KatG_2D, was cloned into pGFPCR by PCR amplification using the primers GAAGGTTCTCATATGTACATCGGGCCGGAAGTG and ATCCAGCGTCGACTCCAGATAAGTGTGAGCACAACCAGGC, which included *Nde*I and *Sal*I restriction sites, respectively. The fragment was subcloned into pET28a(+), an expression vector that adds an amino-terminal hexa-His tag to the expressed protein, using the *Nde*I and *Sal*I restriction enzymes. A culture of *E. coli* BL21 p*lysS* strain transformed with the KatG_2D encoding plasmid was grown at 293 K to an A_{600} of 0.3–0.4, induced with 0.2 m*M* IPTG and grown overnight. KatG_2D was purified in a single step by elution from a Hi-Trap column equilibrated with 20 m*M* phosphate, 500 m*M* NaCl, 10 m*M* imidazole pH 7.4, to which a gradient of 10–500 m*M* imidazole in the same buffer was applied. The protein was concentrated to 30 mg ml^{-1} in 20 m*M* phosphate, 100 m*M* NaCl pH 7.0 buffer.

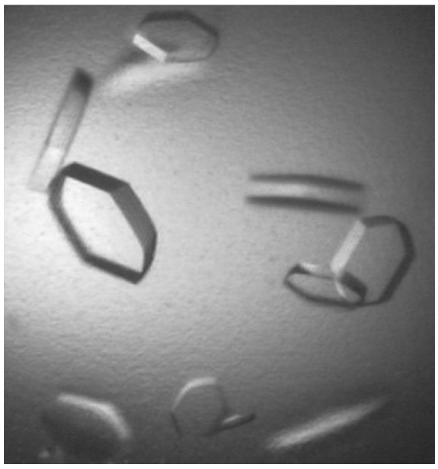

Figure 2
Crystals of the second domain of HPI. The longest *c* axis is perpendicular to the largest crystal faces.

2.2. Crystallization and crystal-packing analysis

Plate-like crystals from the KatG_2D protein were obtained by the hanging-drop vapour-diffusion method in 0.6 *M* sodium citrate buffered to pH 7.5 with HEPES at 293 K (Fig. 2). The crystals diffracted beyond 2.0 Å resolution and belong to the orthorhombic space group $P2_12_12_1$, with unit-cell parameters $a = 84.2$, $b = 98.7$, $c = 302.8$ Å. The longest *c* axis was perpendicular to the

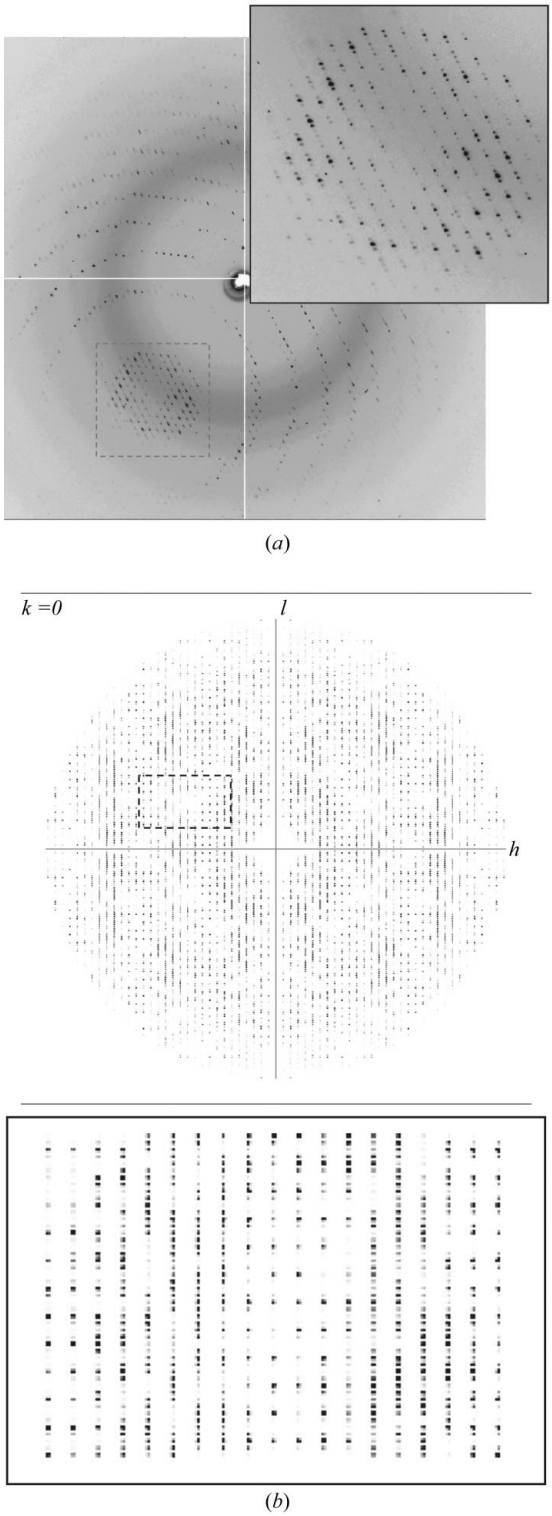

Figure 3
(*a*) 0.5° oscillation image obtained from the KatG_2D crystals collected at the ESRF ID14-2 station (Grenoble). Clustering of strong and weak reflections is evident in the inset. (*b*) A representation of the *h*0*l* reciprocal-space plane emphasizes the alternancy of layer lines perpendicular to the *c** direction with different periodicity. This pattern of intensity distribution suggests the presence of non-crystallographic translational symmetries (see text).

Table 1
Data-collection statistics.

Values in parentheses are for the highest resolution shell.

	KatG_D2	KatG_D2-Pt
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit-cell parameters (Å)	a = 84.2, b = 98.7, c = 302.8	a = 84.4, b = 99.6, c = 303.1
No. of unique reflections	135569	71298
Resolution range (Å)	20–2.1 (2.2–2.1)	20–2.7 (2.8–2.7)
Completeness (%)	92 (89)	99 (100)
$R_{\rm merge}$ (%)	8.7 (38.0)	9.5 (37.9)
Average $I/\sigma(I)$	7.2 (2.6)	6.2 (2.5)

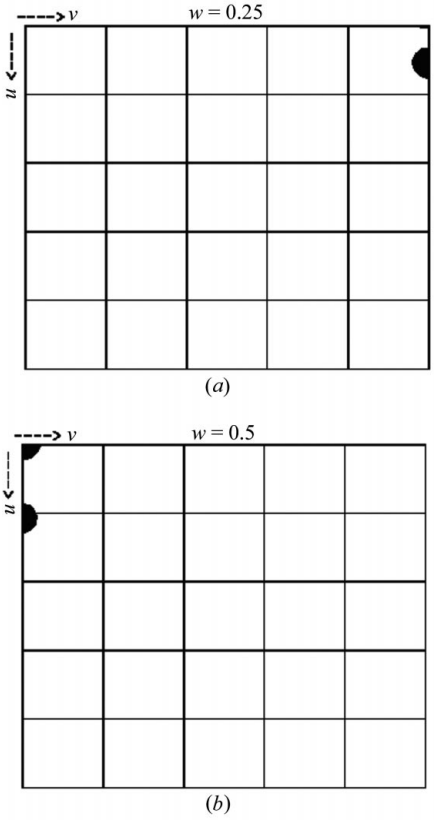

Figure 4
Native Patterson map of sections (*a*) w = 0.25 and (*b*) w = 0.5. The presence of three pseudo-origin peaks is evident at (u, v, w) positions of (0.053, 0.5, 0.25), (0.0, 0.0, 0.5) and (0.106, 0.0, 0.5), respectively.

largest crystal face (Fig. 2) which, together with mosaicities above 0.7°, complicated data collection at high resolution. Diffraction data sets for both the native KatG_2D protein and a heavy-atom platinum derivative were collected using synchrotron radiation with flash-cooled crystals at resolutions of 2.1 and 2.7 Å, respectively. Reflections in the diffraction images presented a peculiar distribution of intensities, suggesting the existence of non-crystallographic translational symmetries (Fig. 3). Diffraction data were processed and scaled using *HKL*/*SCALEPACK* (Otwinowski, 1993) and programs from the *CCP*4 package (Collaborative Computational Project, Number 4, 1994) with the statistics shown in Table 1.

Patterson maps, computed either with the native or with the platinum-derivative diffraction data, showed three pseudo-origin peaks which confirmed the existence of translational symmetries (Fig. 4). The origin and peaks 1 (0.053, 0.5, 0.25) and 3 (0.106, 0.0, 0.5) are on a line at the same distance. The proximity of peaks 2 and 3 could explain changes in the relative heights observed for these peaks at different resolutions and seems compatible with assuming they should be of about the same height as pseudo-origin peak 1. Together with the absence of significant peaks in the self-rotation (Navaza & Saludjian, 1994) besides those corresponding to the twofold crystallographic axis, the pseudo-origin peaks in the Patterson maps would be compatible with the presence of three molecules in the crystal asymmetric unit and a solvent content of about 80% by volume (Matthews, 1968). The pseudo-translational symmetry also explains the high value obtained (above 3 at all resolutions) in the *SFCHECK* twinning test, where $\langle I^2\rangle/\langle I\rangle^2$ is analyzed (Vaguine *et al.*, 1999).

3. Conclusions

The C-terminal domain of HPI (KatG) from *E. coli* shows a 25% sequence identity with cytochrome *c* peroxidase (*CCP*) from *S. cerevisae* and has been characterized as a haemless compact globular structure resistant to proteolysis. The 305-residue protein corresponding to the C-terminal domain of KatG from Tyr422 to Leu726 was cloned and overexpressed to produce protein that remains soluble even at concentrations above 30 mg ml^{-1}. Crystals from this protein grown in 0.6 *M* sodium citrate diffracted beyond 2.0 Å and appear well suited to high-resolution studies. Three pseudo-origin peaks in the Patterson maps indicate an unusual packing compatible with the presence of three molecules in the crystal asymmetric unit and a solvent content of about 80% by volume (Matthews, 1968). It is anticipated that solution of the crystal structure of the C-terminal domain may provide insights into the crystal structure of the catalytically active N-terminal domain.

This work was supported by grants BIO099-0865 of DGICYT to IF and OGP9600 from the Natural Sciences and Engineering Research Council of Canada (NSERC) to PCL. Many thanks are given to Drs W. F. Ochoa and N. Verdaguer for their help with data collection.

References

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Hillar, A., Peters, B., Pauls, R., Loboda, A., Zhang, H., Mauk, A. G. & Loewen, P. C. (2000). *Biochemistry*, **39**, 5868–5875.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Navaza, J. & Saludjian, P. (1994). *Acta Cryst.* A**50**, 157–163.
Nicholls, P., Fita, I. & Loewen, P. C. (2001). *Adv. Inorg. Chem.* **51**, 51–106.
Otwinowski, Z. (1993). *Proceedings of the CCP4 Study Weekend. Data Collection and Processing*, edited by L. Sawyer, N. Isaacs & S. Bailey, pp. 56–62. Warrington: Daresbury Laboratory.
Vaguine, A. A., Richelle, J. & Wodak, S. J. (1999). *Acta Cryst.* D**55**, 191–205.
Wada, K., Tada, T., Nakamura, Y., Kinoshita, T., Tamoi, M., Sigeoka, S. & Nishimura, K. (2002). *Acta Cryst.* D**58**, 157–159.
Welinder, K. (1991). *Biochim. Biophys. Acta*, **1080**, 215–220.
Yamada, Y., Saijo, S., Sato, T., Igarashi, N., Usui, H., Fujiwara, T. & Tanaka, N. (2001). *Acta Cryst.* D**57**, 1157–1158.
Zhang, Y., Heym, B., Allen, B., Young, D. & Cole, S. (1992). *Nature (London)*, **358**, 591–593.