Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Direct phasing at low resolution of a protein copurified with human paraoxonase (PON1)

A. Fokine,[a] R. Morales,[b] C. Contreras-Martel,[b] P. Carpentier,[b] F. Renault,[c] D. Rochu[c] and E. Chabriere[a]*

[a]Laboratoire de Cristallographie et Modélisation des Matériaux Minéraux et Biologiques, Unité Mixte de Recherche 7036 CNRS, Université Henri Poincaré Nancy I, BP 239, 54506 Vandoeuvre-lès-Nancy, France, [b]Laboratoire de Cristallographie et Cristallogenèse des Protéines, Institut de Biologie Structurale Jean-Pierre Ebel CEA–CNRS, Université Joseph Fourier, 41 Rue Jules Horowitz, 38027 Grenoble CEDEX 1, France, and [c]Centre de Recherches du Service de Santé des Armées, Unité d'Enzymologie, 24 Avenue des Maquis de Grésivaudan, 38702 La Tronche, France

Correspondence e-mail: chabrier@lcm3b.uhp-nancy.fr

Received 19 April 2003
Accepted 8 August 2003

In this paper, the low-resolution structure of a previously unknown protein copurified with human paraoxonase (PON1) is reported. The structure of this protein was very difficult to solve using classical crystallographic methods. Progress was made using a new phasing method based on topological analysis. From the experimental point of view, this method has the advantage of requiring only a simple low-resolution X-ray data set. The program used and the different steps of the data-processing and phasing procedure are described. The results provided an insight into the failure of previous molecular-replacement attempts. The low-resolution shape of the protein which was presented with confidence is compared with and confirmed by the structure at 1.8 Å solved subsequently using classical methods. This work shows that this direct-phasing method could be used systematically in difficult cases: it provides low-resolution structural information comparable with that obtainable by electron microscopy.

1. Introduction

Human paraoxonase (PON1; EC 3.1.8.1) is a glycosylated lipoprotein of 354 residues associated with high-density lipoprotein in plasma. Although no physiological function has been assigned to this enzyme, it seems to play a role in preventing the development of atherosclerosis (Watson *et al.*, 1995; Shih *et al.*, 1998; Jakubowski, 2000). Moreover, it is capable of hydrolyzing organophosphorus (OP) insecticides and nerve agents (Li *et al.*, 1993). Natural protection against OPs by PON1 has been clearly established in mammals (Costa *et al.*, 1987, 1990; Li *et al.*, 1995; Shih *et al.*, 1998; Haley, 2000). Thus, PON1 is toxicologically relevant because of its catalytic scavenger activity toward OPs.

In order to solve the three-dimensional structure of PON1 by X-ray crystallography, we sought to crystallize the enzyme. PON1 was purified from human plasma using the protocol described by Gan *et al.* (1991). Crystals belonging to two different space groups, monoclinic $C2$ and orthorhombic $C222_1$, were obtained. Although the crystals are small ($\sim$0.001 mm^3), they are of good quality (Fig. 1). A diffraction data set was collected at the European Synchrotron Radiation Facility (ESRF, Grenoble, France) at 1.9 Å (Chabriere *et al.*, 2000).

In an attempt to obtain phases by multiple isomorphous replacement (MIR), we tried to prepare heavy-atom derivatives. We anticipated that this step would be not be too difficult as in solution it is possible to substitute one of the two calcium ions of PON1 by terbium (Josse *et al.*, 1999) and to bind one Hg atom owing to the presence of a free cysteine (Sorenson *et*

al., 1995). Surprisingly, we did not succeed in binding Tb, other lanthanide or Hg atoms in the crystals. Soaking and co-crystallization methods were assayed. In the co-crystallization solution PON1 loses its aryl esterase activity, which strongly suggests that heavy atoms are bound to the protein. However, the Patterson maps obtained from the crystals grown in the co-crystallization solution did not reveal any heavy atoms.

At about the same time as our heavy-atom experiments, the structure of diisopropylfluorophosphatase (DFPase) from *Loligo vulgaris* was solved (Scharff *et al.*, 2001). This phophotriesterase displays several features in common with PON1. DFPase also contains two calcium ions, one of which can be substituted by other bivalent ions, as is the case with PON1. Despite the low identity score between the amino-acid sequences (20.4%) given by *ALIGN* (Myers & Miller, 1998), the alignment suggests similarity between the three-dimensional structures of PON1 and DFPase. Many residues involved in the binding of calcium ions or in the enzymatic reaction are conserved or homologous (data not shown). Moreover, submission of the PON1 sequence (P27169) to *3D-PSSM* (a fast web-based method for protein-fold recognition using one-dimensional and three-dimensional sequence profiles coupled with secondary-structure and solvation-potential information; Kelley *et al.*, 2000) gave a very high and contrasting solution; the result also strongly suggests DFPase as a model for molecular replacement. Despite the strong topological homology predicted, we did not obtain any solution using the programs *CNS* (Brünger *et al.*, 1998) or *AMoRe* (Navaza, 1994).

In our case, the three main methods for solving the phase problem (molecular replacement, MIR/MAD or high-resolution direct phasing) were not helpful. Therefore, we decided to test a new phasing program, *GENMEM* (Lunina & Lunin, personal communication). This program is based on topological analysis of Fourier syntheses and allows solution of the phase problem at low resolution (Lunin, Lunina & Urzhumtsev, 2000). From low-resolution images, we aimed to understand the reasons for our earlier failures and to obtain the first structural biological information on human paraoxonase.

2. Materials and methods

2.1. Data collection

The program *GENMEM* finds phases from analysis of Fourier syntheses calculated at very low resolution (see §3). To prevent bias and distortion of the Fourier syntheses, it is necessary to collect complete data at the lowest resolution.

A special data set was collected on the BM14 beamline of the European Synchrotron Radiation Facility with a 135 mm MAR CCD detector.

Particular attention was paid to ensure good measurement of the diffraction spots at very low resolution, avoiding any spot overlaps. For this purpose, the beam was tuned at the relatively long wavelength of 1.55 Å (a good compromise with the photon flux decreasing with energy) to expand the reciprocal lattice. The largest possible crystal-to-detector distance (350 mm) was then used. The resolution at the detector edge was about 8 Å. The smallest beam stop available (0.750 mm) was used and was positioned near the detector in order to avoid obscuring very low resolution reflections. In order to collect a complete data set, an additional cradle χ was joined to the conventional goniometer head, allowing rotation of the crystal around an axis perpendicular to φ. During the first part of the experiment, 180 images with a $\Delta\varphi$ oscillation of 1° were collected at $\chi = 0°$ (*i.e.* without the χ cradle). 110 additional images (110° around φ) were then collected at $\chi = 90°$ in order to reach the blind region ($\chi = 0$). Collision of the χ cradle with the incident beam collimator and obstruction by the χ cradle in the diffraction zone were avoided.

In this way, a complete data set to 8 Å resolution (without any missing reflections) was collected (Table 1). The data were processed using the *XDS* package (Kabsch, 1993).

The intensities of the low-resolution reflections of our data set are weak (see Table 1). Thus, in our case, the classical automatic procedure from *XSCALE* to *XDSCONV* cannot be used. At this point in the data processing, the first 18 low-resolution reflections were carefully manually examined. Owing to the geometry of the data collection, some reflections near the centre of the detector were very badly measured (beam-stop shadow, Lorenz correction, flatness of Ewald sphere *etc.*). When possible from the redundancy, reflections that revealed a low signal-to-noise ratio [$I/\sigma(I) < 2$] were discarded. Discarding these few reflections had a dramatic

Figure 1
Crystal of a protein copurified with human paraoxonase and used in the present study.

Table 1
Summary of data collection.

Resolution (Å)	∞–8	∞–25
λ† (Å)	1.54981	
Space group	$C222_1$	
Unit-cell parameters	$a = 86.8$, $b = 96.2$, $c = 89.9$	
No. of observed reflections	4005	84
No. of unique reflections	459	18
No. of possible reflections	459	18
Completeness (%)	100	100
Redundancy	8.7	4.7
$\langle I/\sigma(I)\rangle$	32.31	3.8
R_{sym}‡ (%)	6.4	44.6

† Beam-stop diameter 0.75 mm at ∼350 mm distance. ‡ The high R_{sym} value at 25 Å arises from some weak reflections that show poor internal agreement. For the stronger half of the reflections, the R_{sym} value falls to 10%.

effect on the integration quality of this weak data set: we obtained a final $I/\sigma(I)$ mean ratio of 3.8 at 25 Å resolution (Table 1). Once the atomic model had been obtained, this low intensity was modelled with the packing and the solvent effect (Fokine *et al.*, 2002).

2.2. The *GENMEM* program and the phase problem at very low resolution

The program *GENMEM* (Lunina & Lunin, personal communication) performs direct phasing based on the topological properties of trial-and-error randomly phased Fourier syntheses (Lunin, Lunina & Urzhumtsev, 2000; Lunin, Lunina, Petrova *et al.*, 2000).

In the first stage of the phasing procedure, the program employs only a small number of reflections (10–40) at the lowest resolution. For these reflections, a large number of random phase sets (about 100 000) are generated with a uniform distribution in order to thoroughly sample phase space.

For each generated phase set, a Fourier synthesis is calculated using the experimental moduli and the topological proprieties of the syntheses are analysed in terms of the number and the size of connected domains (at a pre-selected cutoff density level) and the relative domain volumes. Phase sets giving syntheses satisfying a chosen topological criterion are selected and all other phase sets are rejected.

The selected Fourier syntheses are adjusted to a common unit-cell origin and enantiomorph choice (Lunin *et al.*, 1990; Lunin & Lunina, 1996) and the selected and adjusted phase sets are averaged. The averaging procedure assigns the mean phase value and its figure of merit to each reflection (Lunin, Lunina & Urzhumtsev, 2000). The figure of merit reflects the distribution of the phase value in the selected phase sets. The more the distribution is clustered, the larger the figure of merit is. In the case of a uniform distribution, it is obvious that one can conclude nothing about the correct phase, so the figure of merit must be low.

The mean phases with their figures of merit are used to calculate an *ab initio* phased density map with improved quality.

It is possible to increase the initial resolution by repeating the phasing procedure in a second stage with higher resolution reflections (usually 10–40) added into the process. For the new reflections uniform random phases are generated using the uniform distribution and for the previously phased reflections random phases are generated which are more or less remote from the first-stage averaged phases according to their figures of merit.

Details of the phasing method (Lunin, Lunina & Urzhumtsev, 2000) and test applications with experimental data sets from known structures (Lunin *et al.*, 2001; Chabriere *et al.*, 2001; Lunin *et al.*, 2002; Fokine *et al.*, 2002) have been described previously.

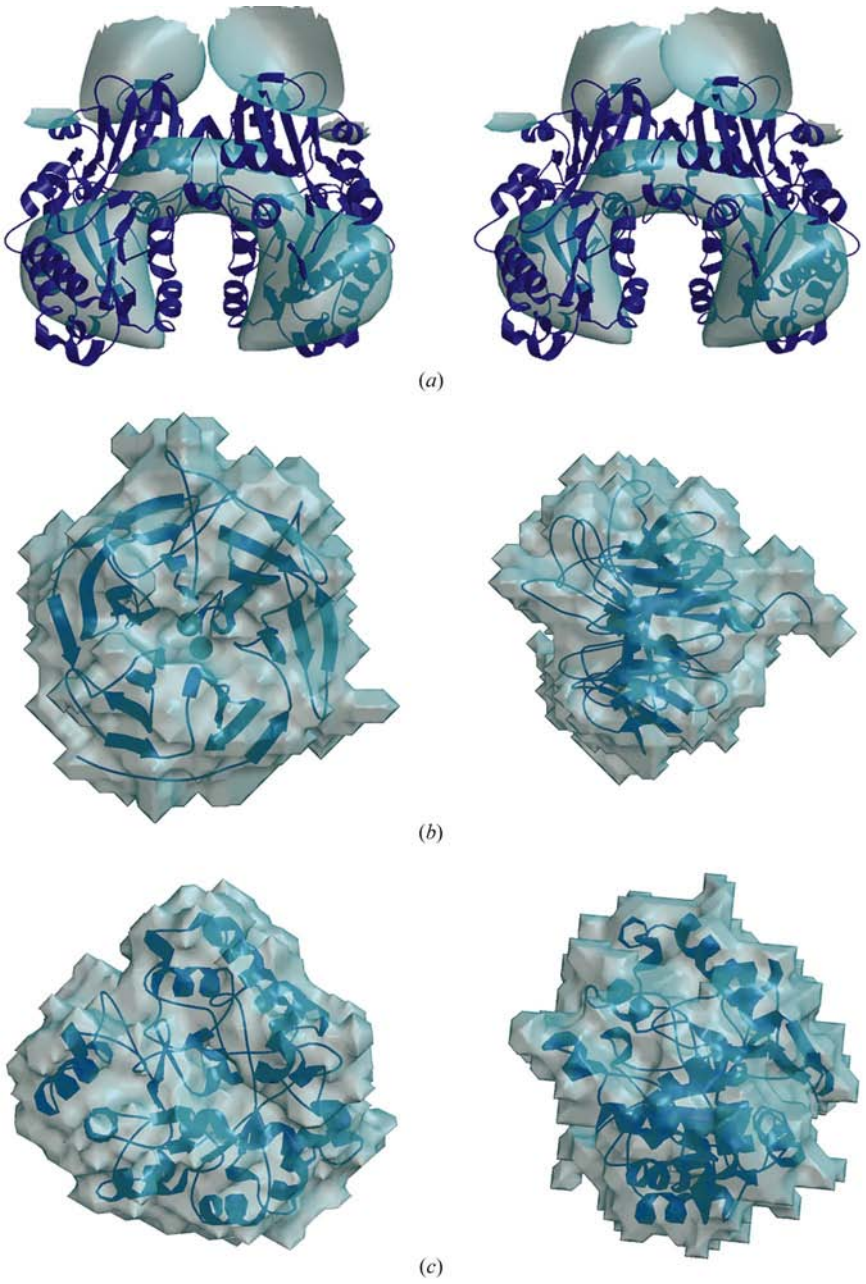

Figure 2
(*a*) Fourier synthesis calculated at 25 Å resolution with *ab initio* phases obtained by density topological trial and error. (*b*) Two perpendicular views of the DFPase structure and its envelope. (*c*) Two perpendicular views of the PTE structure and its envelope.

2.3. Phasing procedure

The phasing was performed at 25 Å resolution using 18 reflections. The only additional information used was the number of molecules in the unit cell. The crystals belong to space group $C222_1$; the unit cell contains eight asymmetric units and the Matthews coefficient shows that the crystals contain a single protein molecule per asymmetric unit. Thus, it was assumed that at high density-cutoff levels the synthesis calculated with correct phases should reveal eight isolated domains in the unit cell, each domain corresponding to one molecule. On this basis, the first criterion applied with *GENMEM* was formulated as the following. At the cutoff level corresponding to 10% of the highest density in the unit cell, the synthesis should reveal eight major connected domains of equal volume (*i.e.* one domain with the multiplicity of eight). Random phases were generated with the uniform distribution and 100 phase sets satisfying this criterion were selected from about 75 000 generated sets. The Fourier synthesis calculated with averaged phases revealed one major connected domain with a multiplicity of four, located on a twofold crystallographic axis.

The inconsistency between the selection criterion and the topology of the averaged-phases Fourier synthesis strongly suggests that the criterion was not optimal.

Therefore, the next criterion tested was one major connected domain with a multiplicity of four. In this case, the program easily found 100 sets satisfying the criterion from only 3781 generated sets and the Fourier synthesis (Fig. 2*a*) calculated with the averaged phases was consistent with the selection criterion. The small number of generated phase sets required to obtain 100 selected phase sets and the consistency between the topology of the final Fourier synthesis and the selection criterion strongly suggests the 25 Å resolution image (Fig. 2*a*) represents the actual structure of the protein (Morales *et al.*, 2002).

A large number of other criteria were formulated and tested. For some criteria, too many random phase sets (over one million) had to be generated to obtain 100 suitable phase sets; for others, the topology of the final Fourier syntheses was inconsistent with selection criteria, showing one major connected domain with a multiplicity of four.

Some attempts were made to increase the resolution of the map. At higher resolution, the map topology became complicated. Owing to increased error possibilities it was difficult to be confident in the results obtained. Therefore, we preferred to work with the clear map obtained at 25 Å resolution.

3. Results and discussion

The final image (Fig. 2*a*) shows a dimer of protein molecules which appear to be in close contact related by a twofold crystallographic axis. Each monomer has an elongated shape and the narrowing of the monomer density suggests that each monomer is constituted of two domains. These observations indicate that the protein has no structural similarities with the other known phospotriesterases, DFPase (Scharff *et al.*, 2001) and phophotriesterase from *Pseudomonas diminuta* (Benning *et al.*, 1995), both of which have globular shapes (Figs. 2*b* and 2*c*). This explains why molecular-replacement phasing with DFPase as a model failed. From the low-resolution image, it was possible to visualize the crystal packing, to locate the centre of mass of the molecule, to discard some possible models for molecular replacement and to determine the number of domains in the protein.

These results were subsequently confirmed by classical SIRAS structure determination at 1.8 Å resolution (Contreras-Martel *et al.*, in preparation) with one heavy-atom derivative, which was obtained using a new family of molecules specially designed to produce better heavy-atom derivatives (Kahn *et al.*, in preparation). The ribbon diagram of the high-resolution structure is shown superimposed on the low-resolution density map obtained *ab initio* in Fig. 2(*a*). Only the inversion operator was applied to superimpose the map on the atomic model, as the *GENMEM* program was not able to discriminate the correct enantiomorph.

During the model-building phase it was not possible to fit the sequence of human PON1 into the density map calculated with experimental phases at 1.8 Å, but owing to the excellent quality of the experimental map it was possible to sequence the protein crystallographically. To our great surprise, the sequence of the crystallized protein was very different from the PON1 sequence. This result was confirmed by N-terminal sequence analysis using the Edman method (Gagnon *et al.*, unpublished result). Therefore, the crystallized protein is not PON1 but another copurified protein. Crystals have been obtained from different purification pools and the presence of PON1 was positively identified in all these pools. The function of this previously unknown protein has been identified and will be reported in a paper of biological interest (Chabriere *et al.*, in preparation). All conclusions obtained by direct phasing at low resolution remain valid for this copurified protein.

4. Conclusion

We have carried out the first *ab initio* low-resolution structure determination using a new phasing method based on topological criteria for a previously unknown protein. The packing, the position, the shape and the number of domains of the protein were correctly determined without any prior information. The low-resolution (25 Å) results from density topological phasing were fully confirmed by a subsequent high-resolution (1.8 Å) structure analysis based on heavy-atom phasing. Moreover, this phasing procedure does not seem to require high-quality data: these results have been obtained despite an exceptionally poor $R_{\rm sym}$ at low resolution.

A crystallographic image obtained *via* density topological phasing, which requires only one complete low-resolution data set, can also provide valuable information for molecular-replacement phasing, allowing rejection of impossible models, determination of the number of molecules in the asymmetric unit or determination of the model and/or the location in the unit cell.

The authors thank V. Lunin (IMPB, Puschino, Russia), N. Lunina (IMPB, Puschino, Russia), J.-C. Fontecilla-Camps (IBS, Grenoble, France) and P. Masson (CRSSA, La Tronche, France) for constant encouragement and support. This work was partially supported by DGA/DSP/STTC grants (PEA990802-99CO029 and PEA010807).

References

Benning, M. M., Kuo, J. M., Raushel, F. M. & Holden, H. M. (1995). *Biochemistry*, **34**, 7973–7978.

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

Chabriere, E., Lunina, N. L., Lunin, V. Y. & Urzhumtsev, A. G. (2001). 20th European Crystallographic Meeting (ECM20), Krakow, Poland.

Chabriere, E., Viguie, N., Baud, D., Josse, D., Ferrer, J. L., Fontecilla-Camps, J. C. & Masson, P. (2000). In *Proceedings of the US Army Medical Defense Bioscience Review, Hunt Valley, USA*. Baltimore, MD, USA: US Army Medical Research and Material Command.

Costa, L. G., McDonald, B. E., Murphy, S. D., Omenn, G. S., Richter, R. J., Motulsky, A. G. & Furlong, C. E. (1990). *Toxicol. Appl. Pharmacol.* **103**, 66–76.

Costa, L. G., Richter, R. J., Murphy, S. D., Omenn, G. S., Motulsky, A. G. & Furlong, C. E. (1987). *ATI ASU Ser.* **13**, 263–266.

Fokine, A., Lunina, N., Lunin, V. & Urzhumtsev, A. (2002). *Acta Cryst.* A**58**, C267.

Fokine, A. & Urzhumtsev, A. (2002). *Acta Cryst.* D**58**, 1387–1392.

Gan, K. N., Smolen, A., Eckerson, H. W. & La Du, B. N. (1991). *Drug Metab. Dispos.* **19**, 100–106.

Haley, R. W. (2000). *J. Psychopharmacol.* **14**, 87–88.

Jakubowski, H. (2000). *J. Biol. Chem.* **275**, 3957–3962.

Josse, D., Xie, W., Renault, F., Rochu, D., Schopfer, L. M., Masson, P. & Lockridge, O. (1999). *Biochemistry*, **38**, 2816–2825.

Kabsch, W. (1993). *J. Appl. Cryst.* **26**, 795–800.

Kelley, L. A., MacCallum, R. M. & Sternberg, M. J. E. (2000). *J. Mol. Biol.* **299**, 499–520.

Li, W. F., Costa, L. G. & Furlong, C. E. (1993). *J. Toxicol. Environ. Health*, **40**, 337–346.

Li, W. F., Furlong, C. E. & Costa, L. G. (1995). *Toxicol. Lett.* **76**, 219–226.

Lunin, V. Y. & Lunina, N. L. (1996). *Acta Cryst.* A**52**, 365–368.

Lunin, V. Y., Lunina, N. L., Petrova, T. E., Skovoroda, T. P., Urzhumtsev, A. G. & Podjarny, A. D. (2000). *Acta Cryst.* D**56**, 1223–1232.

Lunin, V. Y., Lunina, N., Podjarny, A., Bockmayr, A. & Urzhumtsev, A. (2002). *Z. Kristallogr.* **217**, 668–685.

Lunin, V. Y., Lunina, N. L., Ritter, S., Frey, I., Berg, A., Diederichs, K., Podjarny, A. D., Urzhumtsev, A. & Baumstark, M. W. (2001). *Acta Cryst.* D**57**, 108–121.

Lunin, V. Y., Lunina, N. L. & Urzhumtsev, A. G. (2000). *Acta Cryst.* A**56**, 375–382.

Lunin, V. Y., Urzhumtsev, A. G. & Skovoroda, T. P. (1990). *Acta Cryst.* A**46**, 540–544.

Morales, R., Contreras-Martel, C., Renaud, F., Chesne, M. L., Carpentier, P., Ferrer, J. L., Fokine, A., Fontecilla-Camps, J. C., Masson, P. & Chabriere, E. (2002). In *Proceedings of the US Army Medical Defense Bioscience Review, Hunt Valley, USA*. Baltimore, MD, USA: US Army Medical Research and Material Command.

Myers, E. W. & Miller, W. (1998). *Comput. Appl. Biosci.* **4**, 11–17.

Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.

Scharff, E. I., Koepke, J., Fritzsch, G., Lucke, C. & Ruterjans, H. (2001). *Structure*, **9**, 493–502.

Shih, D. M., Gu, L., Xia, Y., Navab, M., Li, W., Hama, S., Castellani, L. W., Furlong, C. E., Costa, L. G., Fogelman, A. M. & Lusis, A. J. (1998). *Nature (London)*, **394**, 284–287.

Sorenson, R. C., Primo-Parmo, S. L., Kuo, C. L., Adkins, S., Lockridge, O. & La Du, B. N. (1995). *Proc. Natl Acad. Sci. USA*, **92**, 7187–7191.

Watson, A. D., Berliner, J. A., Hama, S. Y., La Du, B. N., Faull, K. F., Fogelman, A. M. & Navab, M. (1995). *J. Clin. Invest.* **96**, 2882–2891.