Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Arabidopsis thaliana peroxidase N: structure of a novel neutral peroxidase

Osman Mirza,[a] Anette Henriksen,[a]† Lars Østergaard,[b]‡ Karen G. Welinder[b]§ and Michael Gajhede[a]*

[a]Protein Structure Group, Department of Chemistry, University of Copenhagen, Universitetsparken 5, DK-2100 København Ø, Denmark, and [b]Department of Protein Chemistry, Institute of Molecular Biology, University of Copenhagen, Ø. Farimagsgade 2A, DK-1353 København K, Denmark

† Present address: Carlsberg Laboratory, Department of Chemistry, Gamle Carlsbergvej 10, DK-2500 Valby, Denmark.
‡ Present address: Department of Plant Physiology, Institute of Molecular Biology, University of Copenhagen, Ø. Farimagsgade 2A, DK-1353 København K, Denmark.
§ Present address: Department of Life Science, University of Aalborg, Sohngaardsholmsvej 49/57, DK-9000 Aalborg, Denmark.

Correspondence e-mail: gajhede@jerne.ki.ku.dk

Received 21 September 1999
Accepted 14 December 1999

PDB Reference: ATP N, 1qgj.

The structure of the neutral peroxidase from *Arabidopsis thaliana* (ATP N) has been determined to a resolution of 1.9 Å and a free *R* value of 20.5%. ATP N has the expected characteristic fold of the class III peroxidases, with a C^{α} r.m.s.d. of 0.82 Å when compared with horseradish peroxidase C (HRP C). HRP C is 54% identical to ATP N in sequence. When the structures of four class III plant peroxidases are superimposed, the regions with structural differences are non-randomly distributed; all are located in one half of the molecule. The architecture of the haem pocket of ATP N is very similar to that of HRP C, in agreement with the low small-molecule substrate specificity of all class III peroxidases. The structure of ATP N suggests that the pH dependence of the substrate turnover will differ from that of HRP C owing to differences in polarity of the residues in the substrate-access channel. Since there are fewer hydrogen bonds to haem C17 propionate O atoms in ATP N than in HRP C, it is suggested that ATP N will lose haem more easily than HRP C. Unlike almost all other class III plant peroxidases, ATP N has a free cysteine residue at a similar position to the suggested secondary substrate-binding site in lignin peroxidase.

1. Introduction

Peroxidases are oxidoreductases that catalyze the oxidation of substrates using hydrogen peroxide as the electron acceptor. They are widespread in nature and are expressed by both prokaryotic and eukaryotic cells. Unlike the copper-containing laccases, which all have very similar three-dimensional structures, significant covalent and structural variation is seen among the haem-containing peroxidases. This has lead to a sequence-based definition of superfamilies (Welinder, 1992). Class I peroxidases are intracellular peroxidases of prokaryotic origin and include, for example, mitochondrial yeast cytochrome *c* peroxidase. Class II consists of secreted fungal peroxidases exemplified by lignin peroxidase. Class III peroxidases are the higher plant peroxidases targeted for the secretory pathway with horseradish peroxidase C as the most thoroughly studied example.

The many peroxidase isozymes generally found in plants have given rise to debate and speculation on the significance and background of their diversity. It is obvious that the *Arabidopsis* sequencing project constitutes an opportunity to study the structure–function relationships and the significance of the full repertoire of class III peroxidases in a single species.

The ATP N protein has not been isolated from *A. thaliana* plants, but the gene has been cloned and overexpressed in *Escherichia coli* (Teilum *et al.*, 1999). The gene coding for a closely related peroxidase from horseradish roots HRP N (89% sequence identity) is also known (Bartonek-Roxå *et al.*, 1991). HRP N has never been isolated from horseradish (Bartonek-Roxå & Eriksson, 1994) and the enzyme differs from the known neutral peroxidases HRP B1–B3 (Aibara *et al.*, 1982).

2. Results

2.1. Crystallization

Expression in *E. coli* and folding of a novel neutral (pI 7.8) peroxidase from *A. thaliana* (ATP N) was optimized as reported previously (Teilum *et al.*, 1999). Crystals were grown by vapour diffusion from this enzyme preparation using hanging-drop experiments. A protein solution was dialyzed against 10 m*M* NaAc pH 4.8 and concentrated to 1.64 mg ml^{-1}. Crystals appeared after one week when drops containing 2 µl protein solution plus 2 µl 20% PEG 4000, 10% 2-propanol, 0.1 *M* HEPES pH 7.5 were equilibrated against a reservoir containing 20% PEG 4000, 10% 2-propanol, 0.1 *M* HEPES pH 7.5 at 277 K.

2.2. Data collection

Data were collected at the MAX-Lab synchrotron facility, beamline I711, at Lund University, Sweden. The crystal was flash-cooled to 120 K, preceded by a short wash in a

Table 1
Data collection and refinement statistics.

Space group	$P2_12_12_1$
Unit-cell parameters a, b, c (Å)	73.0, 87.5, 91.4
Maximum resolution (Å)	1.90
Highest resolution shell	1.97–1.90
Number of observed reflections	201921
Number of unique reflections	42515
Completeness (%)	92.6
Outer shell	94.8
R_{sym} (%)	6.5
Outer shell	16.8
$I/\sigma(I) > 2$ (%)	89.2
Outer shell	81.2

solution of 5% glycerol dissolved in the crystallization buffer. The crystal was rotated through 56 frames, with an oscillation range of 1.2° and 30 s exposure per frame. The crystal-to-detector distance was 300 mm. All data were processed using *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997). ATP N crystallizes in the space group $P2_12_12_1$, with unit-cell parameters a = 73.0, b = 87.5, c = 91.4 Å and two molecules in the asymmetric unit. Data-collection statistics are given in Table 1.

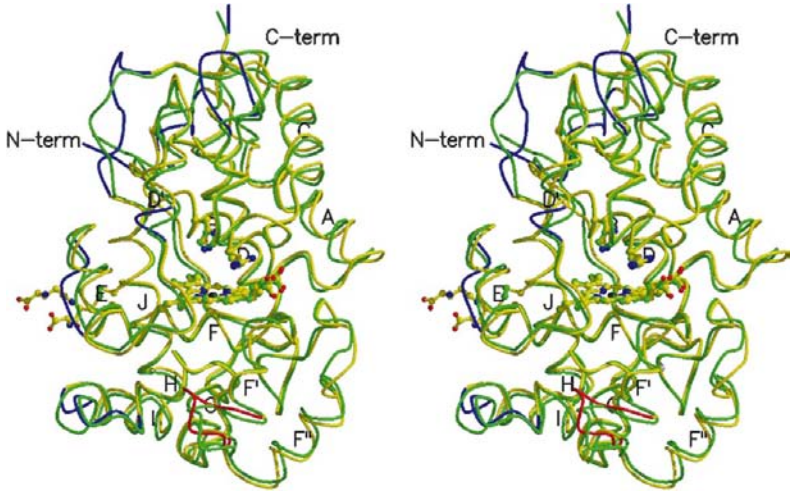

Figure 1
Stereo drawing of ATP N (green) with HRP C (yellow) superimposed. The part of the HRP C trace that corresponds to variable regions in presently known class III structures is coloured blue and red (the red part is the long stretch which is completely missing in the plant peroxidase insert of peanut peroxidase). The almost coincident side chains of the distal histidine and arginine are shown. The glutathione linked to the Cys275 of ATP N (corresponding to Glu279 in HRP C) is also shown. The figure was produced with *MOLSCRIPT* (Kraulis, 1991) and *Raster3D* (Merritt & Murphy, 1994).

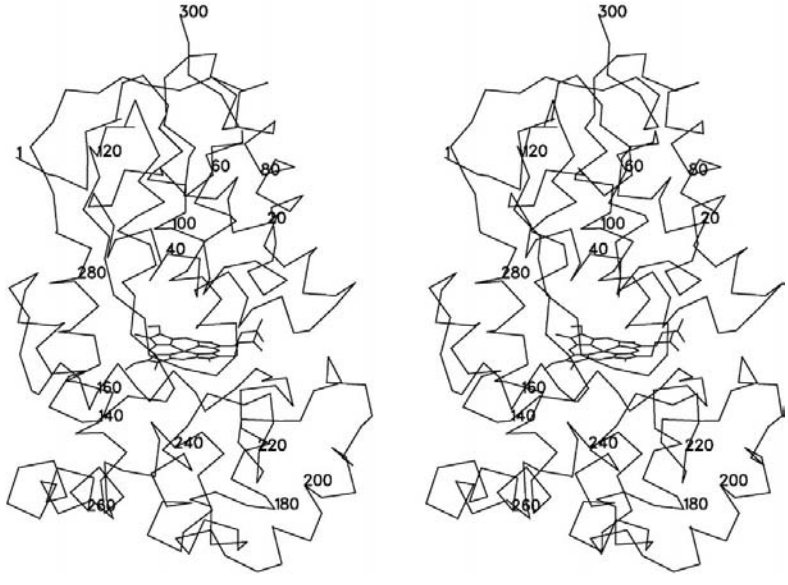

Figure 2
Stereo drawing of the ATP N C^{α} trace. The orientation is the same as in Fig. 1 and the first and every 20th C^{α} atom are numbered. This figure was produced with *SETOR* (Evans, 1993).

2.3. Structure solution and refinement

The structure was solved by molecular replacement using *AMoRe* (Navaza, 1994). The crystal structure of HRP C (Gajhede *et al.*, 1997) was used as a search model. The molecular-replacement searches, made in the resolution range 10–4 Å, gave two clear solutions related by twofold non-crystallographic symmetry. The subsequent rebuilding and refinement of the structure was performed using *O* (Jones *et al.*, 1991) and *CNS* (Brunger *et al.*, 1998) with the Engh–Huber force-field parameters (Engh & Huber, 1991). During the refinement, 5% of the data were set aside for calculation of the R_{free} value. After replacement with the correct residues, the structure was subjected to a 5000 K simulated annealing using torsion-angle refinement and the MLF target function in *CNS* to remove model bias. Cycles of simulated annealing from lower temperatures and manual fitting followed until the final model was obtained. The bulk-solvent correction was applied during all refinement stages. Two B factors per residue were initially refined and restrained individual B factors were finally refined. Strict NCS was applied in the early stages of refinement. All constraints were removed in the final part of refinement. Water molecules were found using *CNS* and were only kept in the model if they returned well shaped $2F_o - F_c$ density above 1σ and were involved in hydrogen bonding. Refinement statistics are given in Table 2.

ATP N contains one cysteine residue (Cys275) which is not involved in a disulfide bridge and excess electron density was observed in close proximity to the S atom of Cys275. This electron density could be described by a glutathione molecule (GSH) disulfide linked to Cys275. The initial coordinates and parameters for the GSH molecule were found in the HIC-UP database (Kleywegt & Jones, 1998). The attachment of the glutathione moiety is expected to arise from the refolding of the recombinant enzyme from inclusion bodies (Teilum *et al.*, 1999).

3. Discussion

Fig. 1 shows a superposition of ATP N and HRP C, which serves as a representative of the class III peroxidases. A stereoview of the ATP N C^{α} trace is shown in Fig. 2. It is clear that the two enzymes have a very similar fold, as indicated by the C^{α} superposition r.m.s.d. of 0.82 Å. In comparison, the r.m.s.d. between the two NCS-related ATP N molecules is 0.3 Å. However, in regions

corresponding to the sequences outside the boxes in the structural alignment in Fig. 2 significant differences are observed. Remarkably, all these divergent regions are situated in one half of the molecule. This also includes the deletion found in peanut peroxidase (Schuller *et al.*, 1996). It is possible that these differences are of importance with respect to peroxidase specificity towards macromolecular substrates (other proteins or cell-wall components) or associated molecules such as the dirigent enzymes (Gang *et al.*, 1999). However, four of these variable regions coincide with the glycosylation sites of wild-type HRP C (Welinder, 1979), which could indicate little mutational strain.

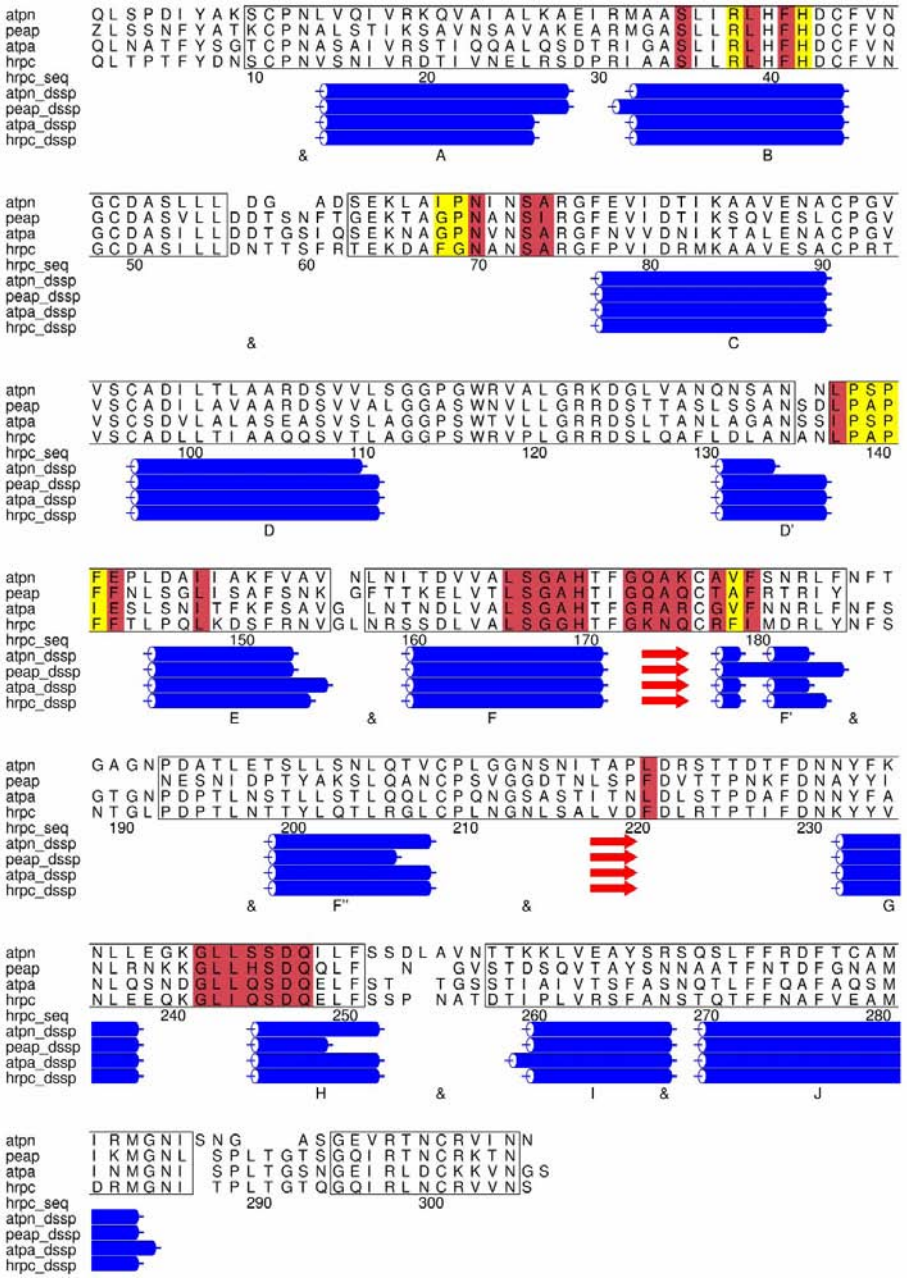

Figure 3
Structural alignment of ATP N and the class III peroxidases with known active enzyme structures. atpn denotes ATP N, peap is peanut peroxidase (Schuller *et al.*, 1996), atpa is ATP A2 (Østergaard *et al.*, 2000) and hrpc is HRP C. The alignment was produced using the programs *STAMP* (Russell & Barton, 1992) and *ALSCRIPT* (Barton, 1993). Stretches corresponding to genuine topological equivalence (Russell & Barton, 1992) are boxed (r.m.s.d. including the boxed 276 C^{α} atoms from each enzyme is 0.65 Å). Residues less than 12 Å from the *δ-meso* haem edge have red backgrounds and those that are known to also be in contact with substrates in the HRP C–BHA or HRP C–FA complexes have a yellow background. The HRP C sequence numbers and the *DSSP* assignments of secondary-structure elements (Kabsch & Sander, 1983) follow the alignment. The last line shows *α*-helix assignments; residues marked '&' are found at the glycosylated sites of wild-type HRP C (Welinder, 1979).

Table 2
Refinement statistics.

Resolution limits (Å)	20.0–1.90
R_{cryst} (%)	17.0
R_{free} (%)	20.5
Number of non-H atoms in refinement	5937
Mean *B* value (Å^2)	11.4
Estimated error from cross-validated Luzzati plot (Å)	0.22
R.m.s.d. from ideality in bond lengths (Å)	0.009
R.m.s.d. from ideality in bond angles (°)	1.8
R.m.s.d. from ideality in improper angles (°)	1.1
Ramachandran statistics† (non-glycine residues)	
Number of residues in most favoured regions (%)	91.3
Number of residues in disallowed regions (%)	0

† Ramachandran statistics calculated with *PROCHECK* (Laskowski *et al.*, 1993).

A sphere with a radius of 12 Å from the *δ-meso* haem edge will include atoms from all residues of the substrate-access channel. As can be seen from Fig. 3, a total of 37 ATP N residues are within the sphere. Any residue interacting directly with small-molecule substrates is likely to be part of this subset. Of the subset, 14 ATP N residues differ from those found in HRP C. Disregarding conservative substitutions, five important changes remain. These are (using HRP C as a reference) Phe143Glu, Lys174Gln, Asn175Ala, Gln176Lys and Arg178Ala. From these substitutions, the electrostatic properties of ATP N's haem surroundings must be different from those of the HRP C haem. The pH dependence of substrate turnover must also be expected to be different in the two enzymes.

Those HRP C residues that have substrate contacts in the HRP C–benzhydroxamic acid (BHA; Henriksen *et al.*, 1998) and HRP C–ferulic acid (FA; Henriksen *et al.*, 2000) complexes are also marked in Fig. 2. Of these nine residues, four are changed in ATP N and all four changes are conservative. The most important substitutions seems to be Phe179Val, which leaves the active site slightly more open in ATP N, and Ala140Ser, which provides another possible hydrogen bond to substrate in ATP N.

There are significant differences between ATP N and HRP C with respect to the hydrogen bonding to the C17 haem propionate closest to the substrate-access channel entry. In HRP C, the propionate is hydrogen bonded to Asn175 and Gln176, although the last bond is long (3.9 Å). The substitution Gln176Lys in ATP N still allows hydrogen bonding, but in HRP C the position of Gln176 is fixed by a hydrogen bond

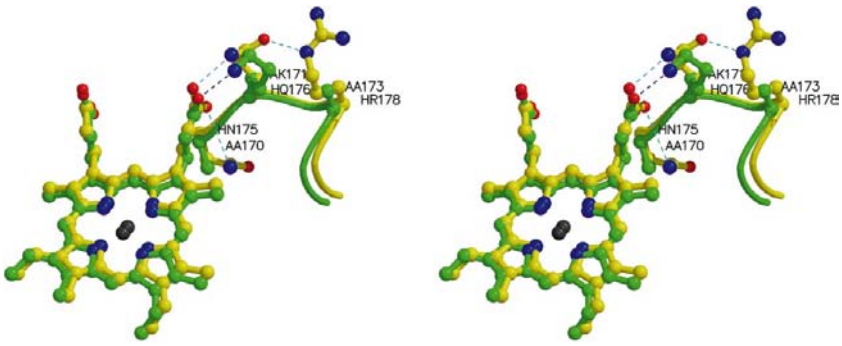

Figure 4
Stereo drawing of the hydrogen-bonding pattern around C17 haem propionate of HRP C and ATP N (superimposed). The C atoms of the HRP C residues His175, His176 and His178 and ATP N residues Ala170, Ala171 and Ala173 have the same colour coding as in Fig. 1. The HRP C residues are labelled with an H and the ATP N residues with an A. Both haem groups are shown in the corresponding colours. The HRP C hydrogen bonds are colored cyan; those of ATP N are coloured blue. The C17 propionate in HRP C is hydrogen bonded by residues 175 and 176. Furthermore, the position of the latter is fixed by hydrogen bonding to residue 178. In ATP N, only one hydrogen bond links the propionate and the apoprotein *via* Lys171. The residues corresponding to Asn175 and Gln171 are replaced by alanines in ATP N.

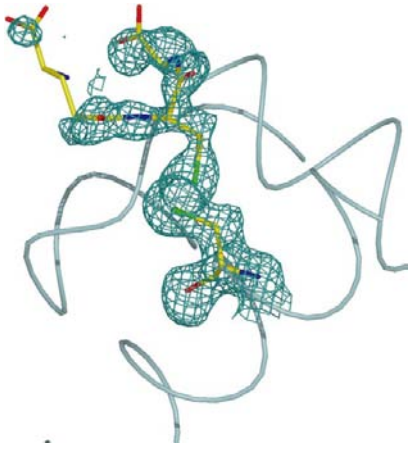

Figure 5
View of the 1σ $2F_o - F_c$ density corresponding to Cys275 and the bound glutathione moiety. This figure was produced using *DINO* (http://www.bioz.unibas.ch/~xray/dino).

to Arg178 N^{ε}. Arg178 is replaced by Ala in ATP N, so the lysine has full conformational freedom. Asn175 is also replaced by Ala in ATP N. These substitutions weaken the binding of the haem group in ATP N and ATP N might release haem more easily than HRP C. The hydrogen-bonding differences are shown in Fig. 4.

A large piece of coherent excess electron density is found in the distal haem pocket. This electron density is not associated with ATP N and extends from right above the Fe^{III} ion out into the substrate-access channel. The density is probably disordered water and buffer molecules. The distance from the closest peak in the electron density to the Fe atom is 2.7 Å. A water molecule is found at this position in HRP C (Gajhede *et al.*, 1997). An O atom only was included at this position in the final model of ATP N to describe the excess electron density. Thus, the coordination number of the iron appears to be five in ATP N, as in HRP C. This O atom is also within hydrogen-bonding distance of the distal histidine and the distal arginine. The disordered electron density is also within hydrogen-bonding distance of the carbonyl O atom of the conserved Pro134 (Pro139 in HRP C). This agrees with the need for a water molecule at this position in the proposed mechanism for substrate oxidation (Henriksen *et al.*, 2000).

Both the distal and proximal Ca^{2+} binding sites are very similar to the sites in HRP C. The distal site is seven-coordinated, with the ligands being carbonyl and side-chain O atoms and a water molecule. The proximal site is six-coordinated, with only carbonyl groups and side-chain O atoms as ligands.

The sequence of HRP C contains eight cysteine residues all engaged in disulfide bridges. In ATP N, an additional cysteine positioned at the centre of the *J* helix (Cys275) is bound to glutathione. Fig. 5 shows the $2F_o - F_c$ density corresponding to the glutathione moiety. It is noteworthy that superpositioning of ATP N and lignin peroxidase shows that this easily oxidizable amino acid is less than 5 Å from the position of the chemically modified tryptophan (Trp171) in lignin peroxidase (Choinowski *et al.*, 1999). Trp171 is suggested to be part of an alternative substrate-binding site in lignin peroxidase. However, the two positions are not related. They appear in different structurally conserved segments and only a few other peroxidases are found to have a cysteine residue at positions equivalent to ATP N 275. These peroxidases are all evolutionarily distant from ATP N.

The research is supported by a grant from the European Union's Biotechnology Program 'Towards Designer Peroxidases' (Contract No. BIO4-CT97-2031) and by the Danish Center for Synchrotron Radiation (DANSYNC).

References

Aibara, S., Kato, M., Mori, E., Morita, Y. & Yamashita, H. (1982). *J. Biochem.* **92**, 531–539.
Barton, G. J. (1993). *Protein Eng.* **6**, 37–40.
Bartonek-Roxå, E. & Eriksson, H. (1994). *J. Biotechnol.* **37**, 133–142.
Bartonek-Roxå, E., Eriksson, H. & Mattiasson, B. (1991). *Biochem. Biophys. Acta*, **1088**, 245–250.
Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Choinowski, T., Blodig, W., Winterhalter, K. H. & Piontek, K. (1999). *J. Mol. Biol.* **286**, 809–827.
Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.
Evans, S. V. (1993). *J. Mol. Graph.* **11**, 134–138.
Gajhede, M., Schuller, D. J., Henriksen, A., Smith, A. T. & Poulos, T. L. (1997). *Nature Struct. Biol.* **4**, 1032–1038.
Gang, D. R., Costa, M. A., Fujita, M., Dinkova-Kostova, A. T., Wang, H.-B., Burlat, V., Martin, W., Sarkanen, S., Davin, L. B. & Lewis, N. G. (1999). *Chem. Biol.* **6**, 143–151.
Henriksen, A., Schuller, D. J., Meno, K., Welinder, K. G., Smith, A. T. & Gajhede, M. (1998). *Biochemistry*, **37**, 8054–8060.
Henriksen, A., Smith, A. T. & Gajhede, M. (2000). In the press.
Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kabsch, W. & Sander, C. (1983). *Biopolymers*, **22**, 2577–2637.
Kleywegt, G. J. & Jones, T. A. (1998). *Acta Cryst.* D**54**, 1119–1131.
Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. (1993). *J. Appl. Cryst.* **26**, 283–289.
Merritt, E. A. & Murphy, M. E. (1994). *Acta Cryst.* D**50**, 869–873.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Østergaard, L., Teilum, K., Mirza, O., Mattson, O., Petersen, M., Welinder, K. G., Mundy, J., Gajhede, M. & Henriksen, A. (2000). Submitted.
Otwinowski, Z., & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Russell, R. B. & Barton, G. J. (1992). *Proteins*, **14**, 309–323.
Schuller, D. J., Ban, N., Huystee, R. B., McPherson, A. & Poulos, T. L. (1996). *Structure*, **4**, 311–321.
Teilum, K., Østergaard, L. & Welinder, K. G. (1999). *Proteins Expr. Purif.* **15**, 77–82.
Welinder, K. G. (1979). *Eur. J. Biochem.* **96**, 483–502.
Welinder, K. G. (1992). *Curr. Opin. Struct. Biol.* **2**, 388–393.