Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Acetobacter turbidans α-amino acid ester hydrolase: merohedral twinning in $P2_1$ obscured by pseudo-translational NCS

Thomas R. M. Barends and Bauke W. Dijkstra*

University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

Correspondence e-mail: bauke@chem.rug.nl

Received 18 July 2003
Accepted 19 September 2003

The structure elucidation of the α-amino acid ester hydrolase from *Acetobacter turbidans* by molecular replacement is described. In the monoclinic crystal, the molecules are related by both rotational and pseudo-crystallographic translational NCS (non-crystallographic symmetry). Refinement of the structure converged at unacceptably high *R* factors. After re-evaluation of the data, it was found that the crystal was merohedrally twinned, with a high twinning fraction. It is shown that the pseudo-crystallographic NCS causes aberrant behaviour of conventional twinning indicators, which explains why the twinning was only realized at the refinement stage.

1. Introduction

The α-amino acid ester hydrolases (AEHs) constitute an enzyme family capable of hydrolyzing esters and amides of α-amino acids (Kato, 1980; Kato *et al.*, 1980*a*,*b*; Barends, Polderman-Tijmes *et al.*, 2003). They can also catalyze the reverse reaction, which makes them useful for the production of certain semi-synthetic β-lactam antibiotics, such as ampicillin, amoxycillin, cephalexin and cephadroxil (Takahashi *et al.*, 1972; van der Does, 2002). In particular, the AEH from *Acetobacter turbidans* (Takahashi *et al.*, 1974) has recently been investigated in this respect (Fernandez-Lafuente *et al.*, 2001; Polderman-Tijmes, Jekel, Jeronimus-Stratingh *et al.*, 2002; Polderman-Tijmes, Jekel, van Merode *et al.*, 2002) and crystallization conditions for this enzyme have previously been reported by this laboratory (Barends, Hensgens *et al.*, 2003).

The structure of this enzyme has now been solved and will be described in detail in another paper. We report here on the structure-elucidation process that was followed with the *A. turbidans* AEH. The structure determination was severely hampered by a rare form of merohedral twinning, which was obscured by non-crystallographic translational symmetry. This case illustrates the need for vigilance on the part of the crystallographer with regard to unusual features of crystals that may, if unchecked, ruin a crystallographic structure determination and render any biological inferences made from its results untrustworthy.

2. Methods

2.1. Crystallization and data collection

Crystals were prepared as described previously (Barends, Hensgens *et al.*, 2003). Briefly, recombinant *Acetobacter turbidans* AEH labelled with a *C*-terminal *myc*-epitope and His_6 tag (Polderman-Tijmes, Jekel, van Merode *et al.*, 2002) was concentrated to 5 mg ml^{-1} in 50 m*M* sodium phosphate

Table 1
Data-collection statistics for the AEH data obtained after processing in $P2_1$ (correct space group) and $P222$ (incorrect space group).

Data were collected in the resolution range 40–2.0 Å. Values in parentheses are for the highest resolution shell (2.07–2.00 Å).

	Processing in $P2_1$ (correct space group)	Processing in $P222$ (incorrect space group)
Unit-cell parameters		
a (Å)	98.4	98.4
b (Å)	275.6	197.0
c (Å)	197.0	275.6
β (°)	90.1	
No. of observations	1655436	1637985
No. of unique reflections	598744	253518
$R_{\rm merge}$† (%)	9.1 (34.1)	9.6 (39.2)
Completeness (%)	85.5 (67.8)	89.8 (76.0)
Redundancy	2.8 (0.7)	5.0 (3.3)
$\langle I/\sigma(I)\rangle$	6.6 (1.5)	10.3 (2.2)

† $R_{\rm merge} = \sum|I - \langle I\rangle|/\sum I$.

buffer pH 6.2. The enzyme was crystallized by mixing 3 µl of protein solution with 3 µl 15–17% PEG 4000, 0.2 *M* ammonium acetate, 0.1 *M* sodium acetate pH 4.6, followed by equilibration against this PEG solution. Prism-shaped crystals of maximum dimensions 0.2 × 0.05 × 0.05 mm were obtained in one week. In an attempt to obtain a complex with a substrate, crystals were soaked for 30 min in mother liquor containing 10 mg ml^{-1} ampicillin. After soaking, the crystals were cryoprotected by soaking them for a few seconds in the same soaking solution containing 25% glycerol and were subsequently frozen in liquid nitrogen. Data from a single crystal were collected at beamline ID14-EH2 at the ESRF in Grenoble, France and were processed with *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997). Unless stated otherwise, all further calculations were carried out with *CCP*4 software (Collaborative Computational Project, Number 4, 1994).

3. Results

3.1. Space-group determination

Diffraction data from an AEH crystal were recorded to 2.0 Å resolution (Table 1). The crystal deteriorated strongly towards the end of the 180° rotation used to collect data, causing a completeness of less than 100%. It was found that the diffraction patterns could be indexed in a primitive orthorhombic lattice with unit-cell parameters 98, 197 and 275 Å. Systematic absences indicated a screw axis along the longest (275 Å) cell axis. Furthermore, alternating weak and strong reflections suggested pseudo-translational symmetry in the crystal. After integration, scaling and merging in $P222$ (see Table 1 for data statistics), clear peaks were found in the cross-rotation function with a search model based on the tetramer of the Y206A mutant of AEH (Barends *et al.*, 2003), but the translation function gave no clear solution with good packing for any of the orthorhombic space groups. Clearly, the true symmetry is lower than the apparent symmetry. Since systematic absences along the 275 Å cell axis suggested this axis to be associated with a crystallographic twofold screw axis, the data were processed again in space group $P2_1$ (see Table 1 for data statistics) with the 275 Å axis as the unique (b) axis. During post-refinement of the unit-cell parameters, β refined to 90.1°. The data showed a strong peak in the self-Patterson summation at (0, 0.005, 0.5), indicating a pseudo-crystallographic translation of ∼0.5 along the c axis. Interestingly, apart from the β angle being indistinguishable from 90°, $a^* = 2c^*$ in this crystal system.

In $P2_1$, clear solutions for both rotation and translation functions could be found and four AEH tetramers could be placed in the asymmetric unit. In doing so, the information about the pseudo-crystallographic translation was indispensable. Furthermore, non-crystallographic twofold rotation operations were found parallel to the a and c axes, seemingly explaining why the data could be merged in $P222$. The correctness of the solution could be confirmed by positive difference density for the Tyr206 side chain, since the structure of the Y206A mutant was used as a search model. No density for an ampicillin molecule was observed.

3.2. Unsuccessful refinement owing to pseudo-crystallographic translation symmetry?

Despite successful phasing in $P2_1$ with molecular replacement, refinement converged at unsatisfactorily high R and $R_{\rm free}$ values of 29 and 31%, respectively. Further refinement, rebuilding or addition of water molecules resulted in intolerably large differences between R and $R_{\rm free}$, indicating problems with the refinement process. At first, we believed the pseudo-crystallographic translation to be the problem. Since this translation gives rise to alternate weak and strong reflections, such symmetry has deleterious effects on refinement and structure quality (Vajdos *et al.*, 1997). Firstly, the weak reflections have relatively large errors, and secondly, these reflections lower the denominator in the fraction used to calculate the R factor ($\sum\big||F_{\rm obs}| - |F_{\rm calc}|\big|/\sum|F_{\rm obs}|$), giving rise to artificially high R factors. However, in the present case, the translational symmetry only affects low- and medium-resolution reflections, *i.e.* only low-resolution reflections of the zone $l = 2n + 1$ are weak. One result of this effect is that $\sum|F_{{\rm obs}(l=2n+1)}|$ is approximately as large as $\sum|F_{{\rm obs}(l=2n)}|$, meaning that the effect of the $l = 2n + 1$ reflections on the denominator $[\sum|F_{\rm obs}| = \sum|F_{{\rm obs}(l=2n+1)}| + \sum|F_{{\rm obs}(l=2n)}|]$ is small. Thus, the denominator is not significantly decreased by the $l = 2n + 1$ reflections, which rules out this explanation for the high R factors.

3.3. Unsuccessful refinement owing to twinning?

Another possible cause for the observed problems is merohedral twinning, in which the diffraction patterns from two different 'twin domains' overlap. However, the intensity statistics normally calculated to check for merohedral twinning were closer to values expected for non-twinned crystals than to values for perfect twins (Table 2). Also, $P2_1$ does not normally allow merohedral twinning and therefore the possibility of twinning was not investigated further in the earlier

Table 2
Intensity statistics used for the detection of twinning calculated from the measured AEH data, simulated AEH data with a twinning fraction of 0.5 and observed data from a twinned crystal of the haloalcohol dehalogenase HheC.

The expected values for single crystal and perfectly twinned data are also indicated. Statistics were calculated using *TRUNCATE* and *ECALC* from the *CCP*4 package (Collaborative Computational Project, Number 4, 1994).

Statistic (all data acentric)	Expected value for a single crystal	Expected value for a perfectly twinned crystal	Observed value with AEH data	Calculated value with AEH coordinates (perfect twinning simulated)	Observed value with HheC data (near-perfect twin)
$\langle E \rangle$	0.886	0.94	0.903	0.916	0.931
$\langle E^3 \rangle$	1.329	1.175	1.289	1.256	1.202
$\langle E^4 \rangle$	2.0	1.5	1.891	1.787	1.586

stages of structure determination. However, in rare cases where fortuitous unit-cell parameters allow an extension of the crystallographic symmetry, twinning of monoclinic crystals can occur. In such cases, the symmetry of a higher-symmetry space group can be seen in the diffraction pattern if the twinning fraction is large (*i.e.* in cases of almost perfect or perfect twinning; Yeates & Fam, 1999). Thus, diffraction from highly twinned monoclinic crystals with $\beta \simeq 90^\circ$ may mimic a primitive orthorhombic symmetry (Larsen, 2002) and diffraction from highly twinned monoclinic crystals with $a^* \simeq c^*$ may mimic a *C*-centered orthorhombic symmetry (Declercq & Evrard, 2001). If such twinning results in perfectly overlapping lattices, as was reported by, for example, Ito *et al.* (1995), it is called merohedral twinning. If the overlap is imperfect, it is called pseudo-merohedral twinning (Yeates & Fam, 1999). Such pseudo-merohedral twinning gives rise to split spots at high resolution.

In our case, typical symptoms of strong merohedral twinning were observed. Firstly, only single spots occur at the highest resolution (Fig. 1). Secondly, the diffraction mimics

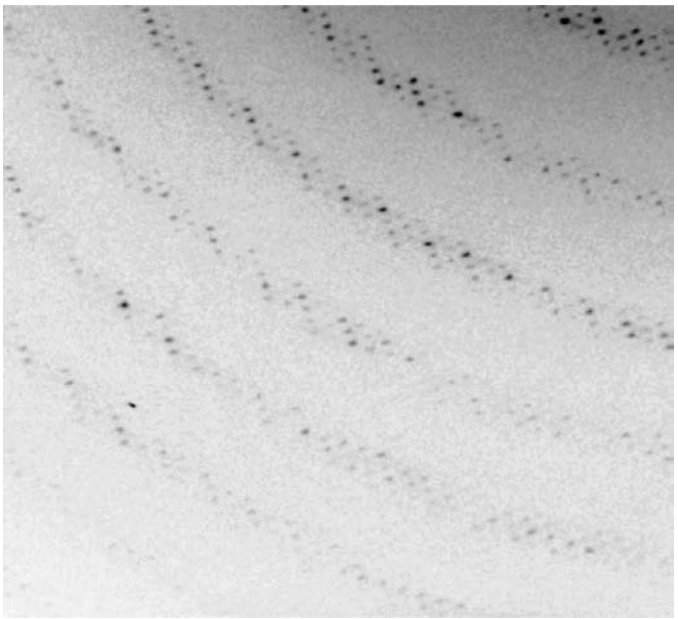

Figure 1
Typical diffraction from an AEH crystal, showing exclusively single spots at the high-resolution diffraction limit. The lower-left corner corresponds to 2.0 Å resolution.

222 symmetry, as witnessed by the relatively low merging *R* factor obtained from processing in *P*222. Thirdly, the structure could not be satisfactorily refined. In addition, the lattice may allow merohedral twinning in this special case, since β refined to 90° within experimental error. Indeed, upon refinement in a triclinic lattice, *i.e.* when no restrictions were applied to any of the unit-cell parameters, β refined to 89.96°. Thus, given the value of β and the apparent *P*222 symmetry, it is possible that the crystal was twinned by the twinning operator $(h,-k,-l)$.

3.3.1. Unexpected behaviour of twinning indicators. Indeed, in retrospect, the cumulative intensity distribution looked slightly but decidedly sigmoidal (Fig. 2), indicative of twinning. This sigmoidal character was not, however, as pronounced as might be expected when compared with data from a twinned crystal of the haloalcohol dehalogenase HheC, which displayed a twinning fraction of 0.48 (data courtesy of René de Jong, University of Groningen). Thus, the cumulative intensity distribution indicated a low degree of twinning at most.

Another way to diagnose twinning is by calculating the intensity statistics $\langle E \rangle$, $\langle E^3 \rangle$ and $\langle E^4 \rangle$. Table 2 shows the values of these statistics expected for non-twinned and perfectly twinned data, as well as the observed values for the AEH data and the twinned data from HheC. For the AEH data, all these indicators are closer to the values expected for non-twinned data than for twinned data, again hinting at a small twinning fraction.

The tests above are based on overall intensity statistics. Another way to diagnose twinning is to look for similarity between presumably twin-related reflections, assuming they are not already known to be equal owing to perfect twinning or higher symmetry. Such methods can be used to test for partial twinning, *i.e.* with a twinning fraction that is less than 0.5, and allow the researcher to estimate the twinning fraction. In principle, such tests are more sensitive, but they can be confused by pseudo-crystallographic NCS, which can also give rise to similarities between reflections. When the distribution of H [$H = \big||I_1| - |I_2|\big|/(|I_1| + |I_2|)$, where I_1 and I_2 are twin-related reflections (Yeates, 1997; Yeates & Fam, 1999)] was calculated using the program *DETWIN* (Collaborative Computational Project, Number 4, 1994) with the possible twinning operator $(h, -k, -l)$, a large twinning fraction of over 0.4 was indicated. To obtain a more accurate estimate of the twinning fraction, a Britton plot (Britton, 1972) was calculated. Such a plot shows the number of negative intensities after detwinning as a function of the twinning fraction. Using this plot, a twinning fraction of 0.43 was obtained by extrapolation of the linear part. Such a high twinning fraction is consistent with the low $R_{\rm merge}$ value obtained in space group *P*222.

To show that the high twinning fraction indicated by these latter two methods arises from twinning and not from pseudo-crystallographic NCS, both tests were repeated using only data with a resolution of between 2.5 and 2.0 Å. If the similarities between the potentially twin-related reflections are because of pseudo-crystallographic NCS, the similarity should break down at high resolution. However, with the high-resolution data both methods again indicated a high twinning fraction, showing that twinning is at play instead of pseudo-crystallographic NCS.

Thus, whereas the data have a high twinning fraction according to the distribution of H and the Britton plot, this was not at all obvious either from the cumulative intensity distribution or from $\langle E\rangle$, $\langle E^3\rangle$ or $\langle E^4\rangle$. However, the expected shape of the cumulative intensity distribution and the expected values of $\langle E\rangle$, $\langle E^3\rangle$ and $\langle E^4\rangle$ are all derived from the standard Wilson distribution of structure factors, which assumes a random distribution of atoms. Given the non-crystallographic translation symmetry, which gives rise to classes of strong and weak reflections, this assumption is invalid in the present case.

To investigate the effect of the non-crystallographic symmetry on the intensity statistics, a data set was calculated from the AEH model in which perfect twinning was simulated.

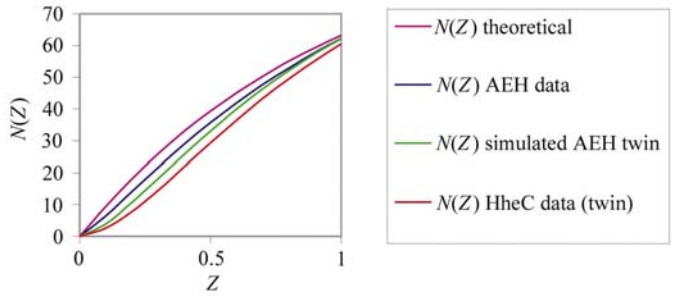

Figure 2
Theoretical and observed cumulative intensity distribution of the AEH data ($Z = I/\langle I\rangle$). A slight sigmoidal character is present in the distribution of observed intensities, which is much stronger in the intensity distribution of a twinned crystal of the protein HheC.

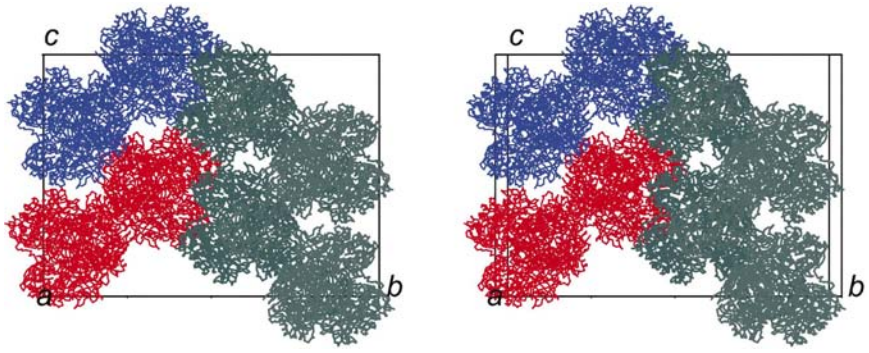

Figure 3
Stereofigure showing the packing of the AEH molecules in the unit cell. The asymmetric unit contains two pairs of tetramers, depicted in red and blue. The red pair and the blue pair are related by translation along *c*. Within a pair, the tetramers are related by twofold symmetry parallel to the *a* axis.

Indeed, the cumulative intensity distribution calculated from these data shows a less pronounced sigmoidal character than that for the HheC data. This fact is easily explained by the pseudo-translational symmetry, since the effect of twinning is to reduce the number of weak reflections, which causes a sigmoidal cumulative intensity distribution, while the effect of the pseudo-translation is to cause a large number of weak reflections, which counteracts the effect of the twinning and reduces any sigmoidal character in the cumulative intensity distribution.

Moreover, the values for $\langle E\rangle$, $\langle E^3\rangle$ and $\langle E^4\rangle$ from the calculated AEH perfect-twin data show marked deviations from the values for a perfect twin expected from a Wilson distribution. Indeed, the values have shifted significantly towards the values expected for a non-twinned data set. Since these data are calculated, this can only be a result of the crystal packing. Moreover, the observed values for these statistics are close to the values obtained with the simulated perfect twin, in accordance with a high degree of twinning in the observed data.

Such a cancelling out of the effects of twinning and crystal packing was recently described by Padilla & Yeates (2003), who showed that the statistics usually calculated for the detection of twinning may become useless in cases of anisotropy or pseudo-centring. They proposed a novel method for the detection of twinning based on local intensity differences which is indifferent towards the effects of classes of low- and high-intensity reflections. The present case illustrates the need for such statistics to be incorporated in the standard set of diagnostic tools at the disposal of the crystallographer.

3.3.2. Detwinning and refinement. As stated above, the twinning fraction as derived from the Britton plot was 0.43. Because at such high twinning fractions detwinning by the standard method for partially twinned data is impossible, the twinning was 'idealized', *i.e.* the data were turned into a data set with a 'perfect' twinning fraction of 0.5, by averaging the intensities of the twin-related reflections (Yeates, 1997). Subsequently, the data could easily be detwinned since a reasonable model of the structure was available. To this end, a procedure implemented in *CNS* (Brünger *et al.*, 1998) was used which essentially calculates the expected differences between twinned and untwinned data from the model and corrects the observed structure factors accordingly. Refinement against the detwinned data immediately resulted in R factors of ∼20%. To allow correct cross-validation of the model, a new test set was defined such that two twin-related reflections, (h, k, l) and $(h, -k, -l)$, are always in the same set. To erase the 'memory' of the previous test set, simulated annealing was performed in *CNS* (Brünger & Rice, 1997). The structure was then further refined using *REFMAC*5 (Murshudov *et al.*, 1997). Water molecules were automatically placed using *ARP*/*wARP* (Lamzin & Wilson, 1993). The

Table 3
Statistics for the current AEH model.

Values in parentheses are for the highest resolution shell.

Resolution range used in refinement (Å)	40–2.0
R factor (%)	19.7 (24.3)
No. of reflections in working set	504378
R_{free} (%)	23.8 (28.8)
No. of reflections	25420
R.m.s.d. bond lengths (Å)	0.007
R.m.s.d. bond angles (°)	1.3
No. of protein atoms	78 144
No. of solvent atoms	4138
No. of phosphate ions	4
Overall *B* value (Å^2)	
Main chain	17.4
Side chain	19.0
Solvent	34.7
Ramachandran plot, % residing in	
Core regions	86.0
Additional allowed regions	13.7
Disallowed regions	0.3†

† These include catalytic and structurally important residues.

current model displays the statistics shown in Table 3 and is being refined further.

4. Discussion and conclusions

In our current model of AEH, four approximately spherical 270 kDa tetramers of AEH occupy the asymmetric unit. They can be seen as two pairs of tetramers (Fig. 3). The two pairs are related by a pseudo-crystallographic translation along *c* (0.0, 0.005, 0.50) and the two tetramers within a pair are related by a twofold rotation about *a* coupled to a translation of approximately (−0.48, −0.07, 0.23). Thus, although the twinning operator is parallel to the NCS twofold axis along *a*, the two do not coincide, since the NCS twofold axis is shifted away from the *a* axis.

As a result of the pseudo-crystallographic translation, the unit cell can thus be seen as consisting of two half unit cells with *c* half the length of the *c* axis of the complete cell. Interestingly, in such a half-cell, the NCS operator between the pair of tetramers in the asymmetric unit would be ($x - 0.48$, $-y + 0.93$, $-z + 0.46$), the crystallographic symmetry operator would still be ($-x$, $y + 0.5$, z) and the combination of the two would be ($0.48 - x$, $0.43 - y$, $z - 0.46$). Thus, the three resulting operators are reminiscent of the crystallographic symmetry operations seen in primitive orthorhombic space groups that are shifted from their normal positions. Because of these shifts, the true symmetry is not orthorhombic, causing molecular replacement to fail in the orthorhombic space groups. A similar effect was reported for the twinned monoclinic crystals of Larsen (2002) and the AEH crystals thus provide another example of non-crystallographic rotational symmetry lined up with a twinning operator.

In conclusion, the structure determination of the AEH from *A. turbidans* illustrates the difficulties that may be encountered when twinning and pseudo-crystallographic translational symmetry are present in the same crystal. The problem in this particular case was confounded by the late realization on our part that the lattice in this special case allowed twinning. Thus, we hope that the present report contributes to the growing realization that twinning in macromolecular crystals is rather common, that it even occurs in space groups that do not usually allow it if the unit-cell parameters assume certain values and that it may be difficult to detect if other peculiarities, such as non-crystallographic translational symmetry, are combined with it.

References

Barends, T. R. M., Hensgens, C. M. H., Polderman-Tijmes, J. J., Jekel, P. A., de Vries, E. J., Janssen, D. B. & Dijkstra, B. W. (2003). *Acta Cryst.* D**59**, 158–160.

Barends, T. R. M., Polderman-Tijmes, J. J., Jekel, P. A. Hensgens, C. M. H., de Vries, E. J., Janssen, D. B. & Dijkstra, B. W. (2003). *J. Biol. Chem.* **278**, 23076–23084.

Barends, T. R. M., Polderman-Tijmes, J. J., Jekel, P. A., de Vries, E. J., Janssen, R. B., & Dijkstra, B. W. (2003). In preparation.

Britton, D. (1972). *Acta Cryst.* A**28**, 296–297.

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

Brünger, A. T. & Rice, L. M. (1997). *Methods Enzymol.* **276**, 243–269.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Declercq, J.-P. & Evrard, C. (2001). *Acta Cryst.* D**57**, 1829–1835.

Does, T. van der (2002). International Patent Publication PCT WO 02/20819 A2.

Fernandez-Lafuente, R., Hernández-Jústiz, O., Mateo, C., Terreni, M., Alonso, J., Garcia-López, J. L., Moreno, M. A. & Guisan, J. M. (2001). *J. Mol. Catal., B Enzym.* **11**, 633–638.

Ito, N., Komiyama, N. H. & Fermi, G. (1995). *J. Mol. Biol.* **250**, 648–658.

Kato, K. (1980). *Agric. Biol. Chem.* **44**, 1083–1088.

Kato, K., Kawahara, K., Takahashi, T. & Kakinuma, A. (1980*a*). *Agric. Biol. Chem.* **44**, 1069–1074.

Kato, K., Kawahara, K., Takahashi, T. & Kakinuma, A. (1980*b*). *Agric. Biol. Chem.* **44**, 1075–1081.

Lamzin, V. S. & Wilson, K. S. (1993). *Acta Cryst.* D**49**, 129–149.

Larsen, N. A. (2002). *Acta Cryst.* D**58**, 2055–2059.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Padilla, J. E. & Yeates, T. O. (2003). *Acta Cryst.* D**59**, 1124–1130.

Polderman-Tijmes, J. J., Jekel, P. A., Jeronimus-Stratingh, C. M., Bruins, A. P., van der Laan, J. M., Sonke, T. & Janssen, D. B. (2002). *J. Biol. Chem.* **277**, 28474–28482.

Polderman-Tijmes, J. J., Jekel, P. A., van Merode, A., Floris, T. A. G., van der Laan, J.-M., Sonke, T. & Janssen, D. B. (2002). *Appl. Environ. Microbiol.* **68**, 211–218.

Takahashi, T., Yamazaki, Y. & Kato, K. (1974). *Biochem. J.* **137**, 497–503.

Takahashi, T., Yamazaki, Y., Kato, K. & Isona, M. (1972). *J. Am. Chem. Soc.* **94**, 4035–4037.

Vajdos, F. F., Yoo, S., Houseweart, M., Sundquist, W. I. & Hill, C. P. (1997). *Protein Sci.* **6**, 2297–2307.

Yeates, T. O. (1997). *Methods Enzymol.* **276**, 344–358.

Yeates, T. O. & Fam, B. C. (1999). *Structure*, **7**, R25–R29.