Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Escherichia coli MltA: MAD phasing and refinement of a tetartohedrally twinned protein crystal structure

Thomas R. M. Barends,‡§ René M. de Jong,‡¶ Karin E. van Straaten,‡ Andy-Mark W. H. Thunnissen and Bauke W. Dijkstra*

Laboratory of Biophysical Chemistry, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

‡ These authors contributed equally to this paper.
§ Current address: Max-Planck Institute for Medical Research, Jahnstrasse 29, 69120 Heidelberg, Germany.
¶ Current address: Department of Chemistry, Columbia University, 3000 Broadway, MC 3153, New York, NY 10027, USA.

Correspondence e-mail: b.w.dijkstra@rug.nl

Received 22 September 2004
Accepted 22 February 2005

Crystals were grown of a mutant form of the bacterial cell-wall maintenance protein MltA that diffracted to 2.15 Å resolution. When phasing with molecular replacement using the native structure failed, selenium MAD was used to obtain initial phases. However, after MAD phasing the crystals were found to be tetartohedrally twinned, hampering correct space-group determination and refinement. A refinement protocol was designed to take tetartohedral twinning into account and was successfully applied to refine the structure. The refinement protocol is described and the reasons for the failure of molecular replacement and the success of MAD are discussed in terms of the effects of the tetartohedral twinning.

1. Introduction: twins and quadruplets

Twinned crystals are crystals that consist of a number of 'domains', each of which contributes to the diffraction pattern. In some cases, the diffraction patterns of the different domains overlap exactly. This may, for instance, happen when the orientations of the twin domains are related by a rotation that is part of the holohedry of the crystal lattice (the point group of highest symmetry of the crystal system). Such twins are known as merohedral twins. Solving a protein structure from such merohedrally twinned crystals can be difficult, but has been accomplished in several cases (Barends & Dijkstra, 2003; de Jong *et al.*, 2003; Declercq & Evrard, 2001; Rudolph *et al.*, 2003; Terwisscha van Scheltinga *et al.*, 2001).

Most merohedrally twinned protein crystals reported in the literature consist of two domains. These are called hemihedral twins. In hemihedral twinning, the intensity of a reflection $I_{hkl}^{\rm hemihedral}$ is the sum of two contributions weighted by the twinning fraction α, which is related to the relative size of the individual twin domains,

$$I_{hkl}^{\rm hemihedral} = (1-\alpha)I_{hkl} + \alpha T I_{hkl}, \qquad (1)$$

where $0 < \alpha < 0.5$ and the indices of the two contributing intensities are related by the twinning operator T, which is the reciprocal-space equivalent of the relationship between the orientation of the twin domains.

Since each diffraction intensity arising from a hemihedral twin is the sum of two contributions, the twinning has an 'averaging' effect on the intensities. Thus, compared with data from a single crystal, data from a twinned crystal will contain less very strong and less very weak reflections. Because of this, the shape of the cumulative intensity distribution will be sigmoidal if twinning is at play. Intensity statistics can also be conveniently used to diagnose twinning. If a crystal is twinned, the values of $\langle E^2 - 1\rangle$, $\langle E\rangle$, $\langle E^3\rangle$ and $\langle E^4\rangle$ will deviate from their theoretical values derived from Wilson statistics because

of the 'averaging' of intensities. Strong twinning also adds extra symmetry to the diffraction data, which can confuse the space-group assignment. Higher apparent symmetry (especially when it extends to high resolution) can be a strong indicator of twinning (Chandra *et al.*, 1999; Yeates, 1997; Yeates & Fam, 1999).

There are different ways to deal with twinned data for crystal structure solution. Sometimes, if α is sufficiently small, the data can be 'detwinned' mathematically, after which phasing and refinement can proceed as usual (Terwisscha van Scheltinga *et al.*, 2003; Yeates, 1997). Twinned data can, however, also be used for structure solution directly, *i.e.* without detwinning. For instance, in molecular replacement the extra rotational symmetry caused by the twinning may cause multiple solutions to be found that are related by the twinning operator and one can then build a model from the solution(s) from one twin domain. Also, it has been shown that it is possible to use MAD to phase twinned data directly (Dauter, 2003; Rudolph *et al.*, 2003; Yang *et al.*, 2000).

Refining a model against twinned data, however, requires a special protocol, since the refinement target needs to model the effects of twinning in order to fully explain the observed intensities. Otherwise, the refinement will converge at unacceptably high R factors. In principle, adapting the refinement protocol to hemihedral twinning is fairly simple, since if the twinning fraction and twinning operator are known then it is possible to calculate intensities from the model whilst taking twinning into account using (1). Using calculated 'twinned' intensities from the model, one can refine the model against the twinned data. Such protocols have been implemented in software packages such as *SHELXL* (Sheldrick & Schneider, 1997) and *CNS* (Brünger *et al.*, 1998) and a number of crystal structures have been successfully refined against hemihedrally twinned data; see, for example, Ito *et al.* (1995) and Valegård *et al.* (1998).

A more complicated form of twinning is tetartohedral twinning. In such crystals, the diffraction from four domains contributes to each reflection,

$$I_{hkl}^{\text{tetartohedral}} = \alpha I_{hkl} + \alpha' T' I_{hkl} + \alpha'' T'' I_{hkl} + \alpha''' T''' I_{hkl}. \quad (2)$$

Note that we have now defined four twinning fractions and three twinning operations and that $\alpha + \alpha' + \alpha'' + \alpha''' = 1$; or rather, these twins are not really twins, they are quadruplets. A similar averaging effect on the intensities is to be expected as with hemihedral twinning and the same statistics that are used for the detection of hemihedral twinning should also be useful to detect tetartohedral twinning. Also, if the tetartohedral twinning is strong, there should clearly be extra symmetry in the diffraction data. In a case described by Rosendal *et al.* (2004), tetartohedral twinning added 222 symmetry to the $P3$ space-group symmetry, resulting in apparent 622 point-group symmetry of the crystal.

In principle, it should also be possible to refine a structure from a tetartohedral twin by adapting the refinement protocol analogously to the adaptations made for hemihedral twins, but using (2). In this paper, we describe a refinement protocol adapted to tetartohedral twinning and its successful application in the refinement of the D308A mutant of a soluble form of the lytic transglycosylase A (MltA) in complex with a hexasaccharide substrate analogue against data from a tetartohedrally twinned crystal. MltA is a muramidase involved in bacterial cell-wall maintenance. It cleaves peptidoglycan (a polymer of alternating units of *N*-acetyl glucosamine and *N*-acetyl muramic acid cross-linked by small peptides), forming non-reducing anhydromuropeptides (van Straaten *et al.*, 2004). The structure of the MltA–hexasaccharide complex and its implications will be the subject of another paper (van Straaten *et al.*, in preparation).

2. Experimental methods and results

2.1. Crystallization, data collection and data processing

Preparation of the plasmid coding for the mutant and the expression and purification of MltA D308A will be described elsewhere (van Straaten *et al.*, in preparation). Briefly, the QuickChange kit (Stratagene) was used to make the D308A mutation in the pMSS expression vector. *Escherichia coli* cells carrying the mutant plasmid were expressed either in LB medium to obtain unlabelled protein or in M9 medium supplemented with selenomethionine and all the amino acids apart from methionine to obtain SeMet-labelled protein. Unlabelled and SeMet-labelled protein were purified using ion-exchange chromatography and stored in 30 m*M* NaCl, 10 m*M* $MgCl_2$ and 10 m*M* Tris–malonate pH 5.2, as described in van Straaten *et al.* (2004).

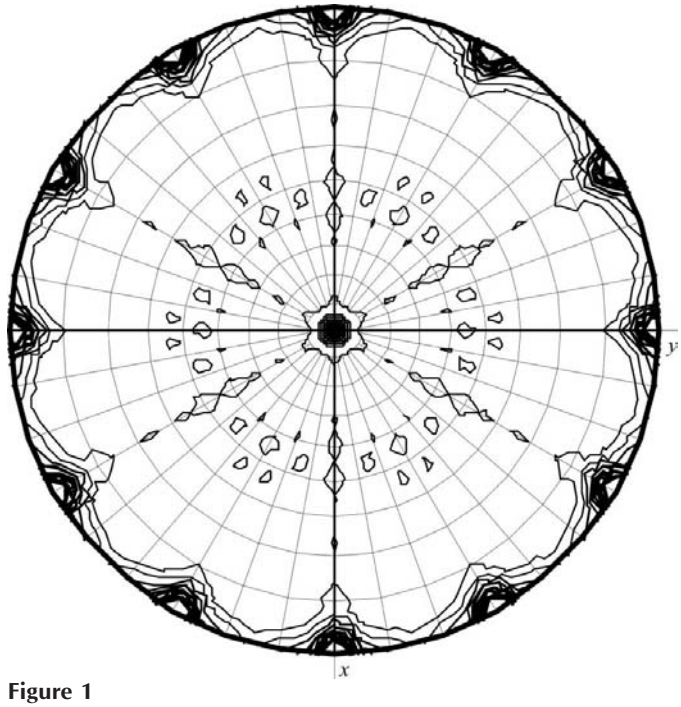

Figure 1
$\kappa = 180^\circ$ section of the self-rotation function of the MltA data collected at beamline BW7A of the EMBL Outstation, Hamburg, Germany. The apparent 622 point-group symmetry of the crystal generates a twofold symmetry axis perpendicular to the paper and six twofold symmetry axes in the plane of the paper. This figure was prepared using *MOLREP* (Vagin & Teplyakov, 1997).

Crystals of both unlabelled and SeMet-labelled protein were grown at room temperature using the modified microbatch method (D'Arcy *et al.*, 1996) by mixing 1.5 µl 15 mg ml^{-1} protein that had been incubated with 15 mg ml^{-1} hexa-*N*-acetyl glucosamine with 1.5 µl 1.5–2.0 *M* ammonium sulfate in 100 m*M* phosphate/citrate buffer pH 4.2 under Al's Oil. Hexagonal bipyramidal crystals grew within 7 d to a longest dimension of 0.25 mm. Crystals were flash-frozen in liquid nitrogen after cryoprotection in mother liquor with 26% glycerol.

A 3.5 Å resolution data set was collected in-house from an unlabelled crystal on a Bruker–Nonius FR591 rotating-anode generator equipped with Osmic mirrors and a MAR Research CCD detector. A three-wavelength MAD data set was collected to 2.5 Å resolution from a SeMet-labelled crystal on beamline BW7A of the EMBL Outstation at the DESY synchrotron in Hamburg, Germany (Pohl *et al.*, 2001). A high-resolution data set was collected to 2.15 Å resolution on the ID14-1 beamline of the ESRF in Grenoble. All data were collected at 100 K and processed with *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997). Statistics are reported in Table 1. At first sight, all data sets could be integrated assuming a primitive hexagonal lattice with $a = b \simeq 91$, $c \simeq 187$ Å (this cell will be called the 'small cell'). However, closer inspection of the diffraction patterns revealed lines of very weak reflections in between lines of much stronger reflections that were not predicted using the small cell. Assuming a much larger cell, in which $a = b \simeq 157$, $c \simeq 187$ Å, these very weak reflections were also predicted. Given that the weak reflections probably arise from one or more pseudo-crystallographic translation operations and given their low intensity, it was decided to initially work with the small cell only and to include the information from the weak class of reflections at a later stage.

2.2. Space-group determination and structure solution

The highest symmetry space groups in which the small-cell in-house data could be scaled and which were consistent with the systematic absences were $P6_422$ and $P6_222$. Initially, molecular replacement was used to attempt to phase these data using the native structure (van Straaten *et al.*, in preparation) as a search model. However, molecular replacement failed to yield solutions in $P6_422$, $P6_222$ or in lower symmetry space groups. At the time, it was thought that perhaps a large conformational rearrangement induced by the hexa-*N*-acetyl glucosamine substrate analogue precluded structure solution by molecular replacement. Therefore, SeMet-labelled protein was prepared and a MAD data set was collected.

These MAD data, collected at the EMBL Outstation in Hamburg, were input into *SOLVE* (Terwilliger & Berendzen, 1999) for evaluation and phasing in space groups $P6_422$ and $P6_222$. Phasing seemed to be successful in $P6_422$. After solvent

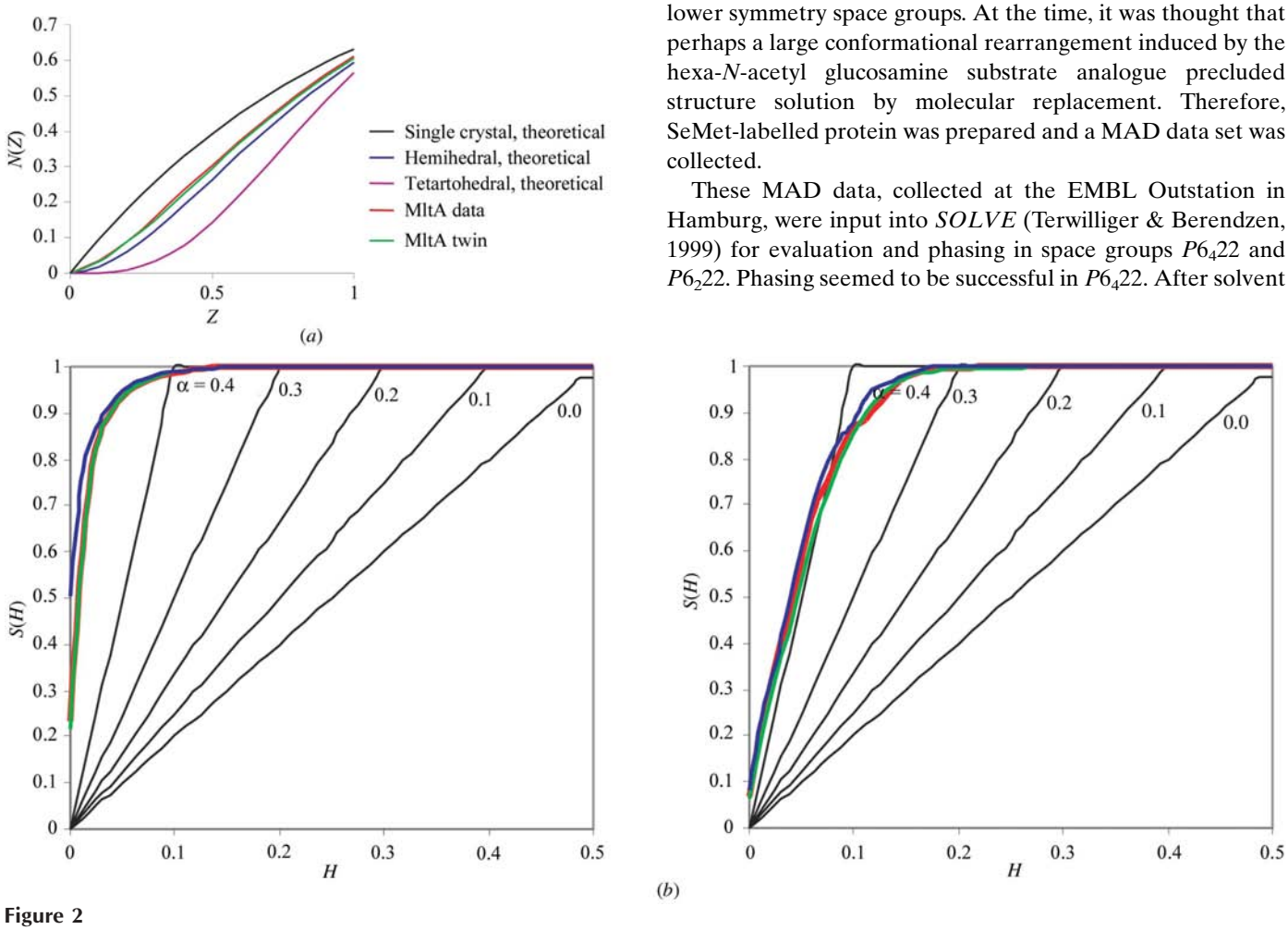

Figure 2
(*a*) Cumulative intensity distributions for acentric reflections. The sigmoidal character of the distribution of the MltA data is indicative of twinning. *Z*, relative intensity; *N*(*Z*), cumulative frequency of occurrence of reflections with intensity *Z*. (*b*) Yeates' twinning test, calculated using *DETWIN* (Collaborative Computational Project, Number 4, 1994), using the three possible twinning operators (red, green and blue line) in $P3_1$ for all reflections (left) and reflections between 2.3 and 2.15 Å resolution (right). Black lines indicate the expected behaviour for α = 0–0.4.

Table 1
Data-collection parameters of the in-house, Hamburg BW7A and ESRF data.

Values in parentheses are for the highest resolution shell.

			BW7A MAD data set			ID14-1	
	In-house		Peak	Inflection point	Remote	Small cell	Large cell
Wavelength (Å)	1.5418		0.9792	0.9794	0.9393	0.934	0.934
Space group	$P6_422$	$P3_1$	$P3_1$			$P3_1$	$P3_1$
Unit-cell parameters (Å)	$a = b = 90.4$, $c = 185.8$		$a = b = 91.1$, $c = 187.5$			$a = b = 91.0$, $c = 187.2$	$a = b = 157.5$, $c = 187.2$
Resolution (Å)	15–3.5		15–2.5	15–2.5	15–2.5	36–2.15	36–2.25
No. of observations	66596	65454	584170	346275	341552	278139	762486
No. of unique reflections	6287	21511	59614	59705	59754	93064	245812
Completeness (%)	99.7 (99.9)	98.7 (98.7)	100 (99.9)	100 (99.9)	100 (100)	100 (99.7)	100 (100)
$R_{\rm merge}$ (%)	14.5 (72.4)	12.8 (62.2)	5.8 (30.6)	4.9 (34.3)	4.7 (26.4)	9.2 (68.0†)	13.5 (65.5‡)
$I/\sigma(I)$	7.0 (3.7)	7.0 (1.9)	16.7 (7.7)	15.3 (5.3)	17.5 (6.4)	8.7 (1.6)	3.8 (2.0)
$R_{\rm merge}$, low resolution (%)	4.4	4.0	4.5	2.5	3.5	5.9	7.1
Resolution range (Å)	20–6.95	20–6.95	15–5.3	15–5.3	15–5.3	36–4.6	36–4.6

† Because of this high $R_{\rm merge}$ in the highest resolution shell, these data were used for refinement to 2.25 Å resolution only. ‡ Because of the very high $R_{\rm merge}$ above 2.5 Å resolution owing to the contribution of weak reflections, these data were used for refinement to 2.5 Å resolution only.

flattening and automated model building with *RESOLVE* (Terwilliger, 1999, 2003*a*, 2003*b*), a map and a small partial model were obtained. The electron-density map could be used to place several fragments of the native structure in the unit cell. However, an appreciable part of the structure could not be built and refinement stalled at high *R* factors. A possible explanation for this was incorrect space-group assignment. Therefore, the process of phasing, building and refinement was repeated using the lower symmetry space groups $P3_1$, $P3_121$, $P3_112$ and $P6_4$. For the trigonal space groups, refinement stalled at high *R* factors and for $P6_4$ the density did not permit the completion of the model. Thus, the correct space group was still unclear and the structure remained elusive.

To determine the correct space group, we examined the height of relevant peaks in the self-rotation function of data processed in space group $P3_1$. This showed very strong peaks showing sixfold symmetry along *c*, as well as six twofold rotation axes in the *ab* plane, indicative of 622 point-group symmetry of the crystal (Fig. 1) as expected from the data processing. Given that all these peaks were of a height comparable to that of the origin peak and were still present in a rotation function calculated with data between 2.7 and 2.5 Å resolution, these symmetry elements are unlikely to be non-crystallographic symmetry operators. Thus, since the space group is clearly not $P6_422$, another explanation for the observations is required.

2.3. Detection of tetartohedral twinning

One possible explanation could be merohedral twinning. As mentioned in §1, this would explain both the high symmetry and the stalling of the refinement at high *R* factors in every possible space group. As expected for twinned data, the cumulative intensity distribution for acentric data (small cell) is clearly sigmoidal (Fig. 2*a*). Moreover, the value of $\langle E^2 - 1\rangle$ for acentric reflections, a strong diagnostic tool for diagnosing twinning, is 0.595, significantly lower than the value of 0.736 expected for a single crystal. Other intensity statistics also clearly indicated twinning (Table 2). Essentially, the same results were obtained with the high-resolution ESRF data.

Table 2
Intensity statistics of the MltA data from Hamburg and their theoretical values for a single crystal and a perfect (hemihedral) twin.

The statistics clearly indicate twinning. Estimated values for a perfect tetartohedral twin were deduced from a tetartohedrally twinned data set calculated from a set of random atoms.

Statistic	MltA data	Theoretical, single crystal	Theoretical, perfect hemihedral twin	Estimated, perfect tetartohedral twin
$\langle E\rangle$	0.93	0.89	0.94	0.97
$\langle E^3\rangle$	1.21	1.33	1.18	1.10
$\langle E^4\rangle$	1.6	2.0	1.5	1.3

In the present case, there are several different ways in which merohedral twinning could lead to the observed 622 point-group symmetry of the crystal. Firstly, the true space group could be $P6_4$, with hemihedral twinning about *a* (or *b*). Secondly, the true space group could be $P3_121$ or $P3_112$, with hemihedral twinning about *a* or *b* (depending on the space group) or about *c*. Thirdly, the space group could be $P3_1$, with tetartohedral twinning, *i.e.* with three twinning operators that are twofold operations about *a*, *b* and *c*. Moreover, since the peaks in the $\kappa = 180^\circ$ section of the self-rotation function are of approximately the same height as the origin peak, any twinning is likely to be perfect or near-perfect and will be difficult to distinguish from a crystallographic symmetry operation. However, it should be possible to identify the correct combination of space group and twinning operator by monitoring the progress of a refinement process that takes twinning into account, since refinement procotols with incorrect combinations of space group and twinning operator will fail to converge with acceptable *R* factors.

First, all the possibilities for hemihedral twinning were tested by refining the structure using established protocols for the refinement of hemihedral twins using *CNS* (Brünger *et al.*, 1998). Several refinement jobs were set up to test the different

Table 3
Validation of the 'tetartohedrally twinned *R* factor' (%).

Different *R* factors were calculated for the MltA model against calculated data, simulating a single crystal, a hemihedral twin and a tetartohedral twin. The data show that the *R* factor is very sensitive to the correct modelling of twinning.

Model used to calculate *R* factor	Calculated data, no twinning	Calculated data, perfect hemihedral twinning simulated	Calculated data, perfect tetartohedral twinning simulated
No twinning	0.0	23.8	26.2
Perfect hemihedral twinning	26.2	0.0	9.5
Perfect tetartohedral twinning	28.8	9.6	0.0

possibilities for hemihedral twinning. Although these jobs did converge with lower crystallographic *R* factors (taking the three possible ways of hemihedral twinning in $P6_4$ and $P3_121$ into account resulted in *R* factor/$R_{\rm free}$ values of 0.273/0.379, 0.249/0.394 and 0.250/0.408, respectively, compared with 0.358/0.398 without twinning), refinement still stalled with unacceptably large differences between the crystallographic *R* factor and $R_{\rm free}$ in all cases. Interestingly, in $P3_1$ incorporation of each of the three possible twinning laws into the refinement protocol improved the statistics significantly. Still, however, assuming any form of hemihedral twinning in $P3_1$ did not lead to an acceptable convergence of the refinement. Clearly, the assumption that there is only one twinning operator relating two twin domains (*i.e.* hemihedral twinning) is insufficient to explain the data, which leaves only the possibility of tetartohedral twinning in $P3_1$.

2.4. Further evidence of tetartohedral twinning

This hypothesis did, however, require further testing. For instance, while the intensity statistics of the MltA data strongly support hemihedral twinning, the theoretical values of these statistics for a tetartohedral twin would be expected to be different from those expected for a hemihedral twin. For instance, the cumulative intensity distribution for a tetartohedrally twinned data set may be expected to show an even more pronounced sigmoidal character than for a hemihedral twin. Indeed, if the theoretical distribution of intensities for a set of *n*-fold 'overlapping' reflections as derived by Stanley (1955) is calculated for the case $n = 4$, the cumulative intensity distribution for a tetartohedral twin is obtained as

$$N(z) = 1 - \left(\frac{64}{6}z^3 + 8z^2 + 4z + 1\right)\exp(-4z), \qquad (3)$$

which shows a more step-function-like shape than the distribution $[N(z) = 1 - (1 + 2z)\exp(-2z)]$ derived for a hemihedral twin (Fig. 2*a*). The MltA data closely follows the curve expected for a set of twofold overlapping reflections, *i.e.* the theoretical curve for hemihedrally twinned data (Fig. 2*a*). However, when the MltA intensities are subjected to fourfold averaging using *SFTOOLS* and the three possible twinning operators, *i.e.* when the data are artificially twinned tetartohedrally, the resulting data set again shows an intensity distribution that is practically identical to that expected for a hemihedral twin. Thus, it would seem that in the present case the intensity distribution statistics are not reliable in distinguishing between hemihedral and tetartohedral twinning, as even an artificial tetartohedral MltA twin displays a cumulative intensity distribution indicating hemihedral twinning. A possible reason for this is given below.

However, the twinning tests incorporated in the *CCP*4 program *DETWIN* (Collaborative Computational Project, Number 4, 1994) clearly indicated tetartohedral twinning. Upon evaluation of the MltA data, the Yeates (1997) twinning test showed almost perfect twinning for each of the three possible twinning operators (Fig. 2*b*), even when only data between 2.3 and 2.15 Å resolution were used (Fig. 2*b*). The same results were obtained with the Britton plot (not shown). We therefore believe that although this is not obvious from the intensity statistics, the data are perfectly or almost perfectly tetartohedrally twinned. Thus, a protocol for refinement against tetartohedrally twinned data was indeed required.

2.5. Refinement protocol for tetartohedrally twinned data

To refine the structure in $P3_1$ assuming tetartohedral twinning, the existing *CNS* scripts for hemihedral twin refinement were augmented to deal with the effects of tetartohedral twinning (see Appendix *A*). Specifically, those routines using calculated structure factors for the calculation of the least-squares refinement target and *R* factors for scaling and for the calculation of maps were adapted to comply with (2). Given that the twinning is perfect or near-perfect as shown by the self-rotation function, it was assumed that $\alpha = \alpha' = \alpha'' = \alpha''' = 1/4$. The use of the adapted routines enables simulated-annealing, energy-minimization and *B*-factor refinement against tetartohedrally twinned data, whilst allowing the progress of the refinement to be monitored using the *R* factors and electron-density maps.

To validate the adapted software, three sets of structure factors were calculated from the MltA model: one set of 'normal' calculated structure factors, one set of structure factors in which perfect hemihedral twinning was simulated by calculating intensities and applying (1) and a set of structure factors in which perfect tetartohedral twinning was simulated using (2). *R* factors were then calculated between the model and these three data sets while assuming no twinning, perfect hemihedral twinning and perfect tetartohedral twinning (Table 3). As expected, the *R* factors were zero when the non-twinned data were tested against the non-twinned model and when the correctly twinned model was used to explain twinned data. When the hemihedral twin model was used to explain the

tetartohedrally twinned data, the R factors were distinctly non-zero, as was the case when the tetartohedral twin model was used to explain hemihedrally twinned data.

2.6. Refinement of the tetartohedrally twinned MltA data

The new routines were then used to refine the MltA structure in $P3_1$ against the tetartohedrally twinned observed data. In accordance with standard practice in cases of near-perfect hemihedral twinning, the tetartohedral twinning in the MltA data was 'idealized' by averaging twin-related intensities. Furthermore, to enable meaningful cross-validation of the refinement process, a set of test reflections was chosen from the MltA data in such a way that any reflection and its three twin-related reflections would all be either in the test set or in the set used for refinement. The 'memory' of the previous test set was then 'erased' using simulated annealing. Refinement (including simulated annealing and energy minimization) of the MltA model using the adapted *CNS* routines resulted in an immediate drop of the R factor and $R_{\rm free}$ to below 30%. After further cycles of refinement and rebuilding and the modelling of the solvent structure, the R factor and $R_{\rm free}$ converged at 16.5 and 22.8%, respectively. Finally, using the same test set but expanded to higher resolution, the structure was refined against the high-resolution ESRF data set, resulting in an MltA model with good geometry and R factors of 16.1 and 19.5% at a resolution of 2.25 Å. The average B factors per monomer were 42.0, 42.7, 42.9 and 45.1 Å^2. After this successful refinement, the model was refined in the large cell against all observed intensities, leading to an R factor and $R_{\rm free}$ of 20.0 and 23.8%, respectively. NCS restraints were employed during refinement using both the large and the small cell data because of the low observation-to-parameter ratio caused by the averaging of twin-related reflections.

3. Discussion

3.1. Crystal packing: causes of twinning

Inspection of the crystal packing shows that MltA packs in layers of tetramers. Each tetramer (Fig. 3) shows approximate 222 point-group symmetry. This 222 symmetry is not exact, mainly because of conformational differences between the monomers in the tetramer. Thus, the packing of MltA in these crystals is comparable to the packing observed by Rosendal and coworkers in tetartohedrally twinned crystals of the crenarcheal SRP core (Rosendal *et al.*, 2004). As is the case in the SRP crystals, the approximate 222 symmetry of the MltA tetramer enables each tetramer to occupy four comparable but non-equivalent orientations, enabling the tetartohedral twinning. Thus, the MltA crystals provide a second example in which the packing of a particle with approximate 222 point-group symmetry in a trigonal lattice leads to tetartohedral twinning. The observation that the approximate 222 symmetry of the particle lines up with the twinning operators explains why the intensity statistics do not clearly show tetartohedral twinning because, as kindly pointed out by a referee, in cases where twinning is promoted by NCS the effects of twinning on the intensity statistics is reduced. This is because when NCS operators are approximately parallel to twinning operators, the averaging effect of the twinning is less pronounced, since the reflections averaged by the twinning were already strongly correlated owing to the NCS (Stanley, 1972).

In the 'large unit cell' there are two additional tetramers per asymmetric unit, related to the first tetramer by non-crystallographic translations of approximately 2/3, 1/3, 0 and 1/3, 2/3, 0.

3.2. Phasing of twinned data

Initial attempts to phase the data using molecular replacement failed when using the native structure as a search model. However, the current structure shows a remarkable conformational difference between the native structure and the mutant structure with hexa-*N*-acetyl glucosamine. MltA consists of a large and a small domain (van Straaten *et al.*, in preparation) and this conformational difference mainly involves a difference in relative orientation between the two domains. Remarkably, the failed attempts to phase the data with molecular replacement included searches with the separate domains. However, the high total symmetry of the crystal packing may be expected to severely complicate molecular replacement, explaining the failure of this technique in this case.

MAD phasing, on the other hand, was successful. This is unexpected, since the small differences in intensities that must be measured for a MAD (or MIR) experiment might be expected to be swamped by the averaging effects of twinning and in the case of MIR data are often detwinned prior to phasing. However, Yang and coworkers have described a case in which MAD was successfully applied to hemihedrally twinned data without detwinning (Yang *et al.*, 2000), as have Rudolph *et al.* (2003). Calculations by Dauter (2003) show that

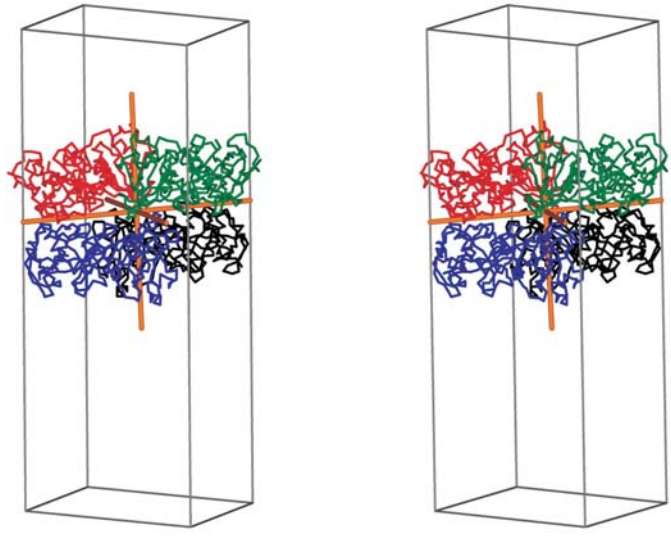

Figure 3
Stereofigure showing a tetramer of MltA in the small unit cell. Each monomer is coloured individually and the twofold NCS operators are shown in orange.

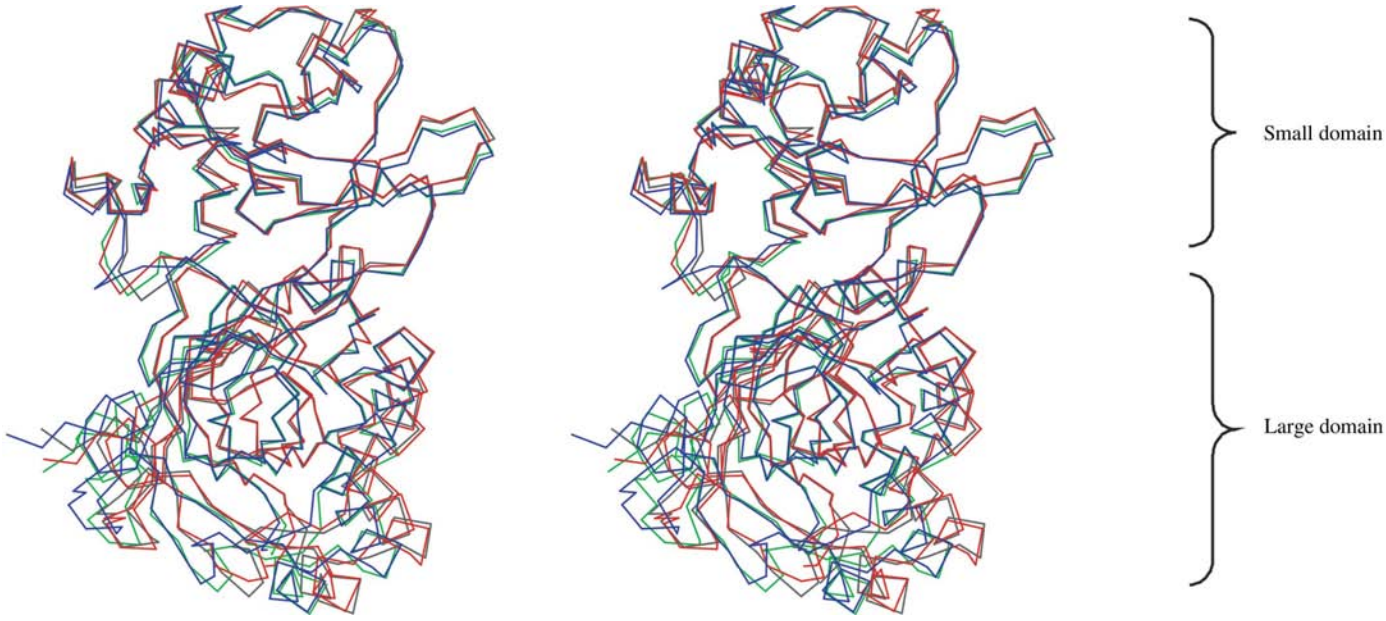

Figure 4
Stereofigure showing a superposition of four monomers, one from each of the four twin domains, constructed using the twinning operators. These operations lead to a reasonable superposition in the small domain of MltA (top), whereas for the large domain (bottom) the distances between corresponding atoms are on average much larger. In many parts of the structure, the green and blue domains are well aligned as are the orange and black domains, whereas in the same regions there are appreciable differences between the green and blue domains on one hand and the red and black domains on the other. This may explain the occurrence of doublet peaks for the Se atoms rather than quartets in the MAD maps.

SAD too may be feasible with untreated hemihedrally twinned data. MltA now provides the first example of a tetartohedrally twinned protein crystal structure solved using MAD.

The success of MAD in this case can be explained from the structure. Applying the real-space equivalents of the twinning operators (three perpendicular twofold rotations) to the refined structure yields sets of coordinates for the four twin domains in one coordinate system. Using these sets of coordinates, the diffraction experiment can be simulated, as the experiment records the sum of the intensities arising from the different twin domains in their relative orientations.

The resulting coordinate sets are very similar owing to the approximate 222 point-group symmetry of the tetramer, in the sense that the position of any atom from a certain monomer in a twin domain is very close to the positions of the same atom from another monomer in each of the other twin domains. A comparison of these coordinate sets shows that significant differences between such comparable but non-equivalent positions are mainly found in the large domain of MltA, whereas in the small domains such differences are much smaller (Fig. 4). Interestingly, the best parts of the MAD maps, in which partial models could be built automatically by *RESOLVE*, are all in the small domain of MltA. Thus, MAD phasing was most successful for those parts of the MltA molecules where the positions of semi-equivalent atoms in different twin domains are most similar.

As also shown in Fig. 4, two of the three twinning operations led to more remarkable differences between semi-equivalent atomic positions in the different twin domains than the other one. Thus, the rotated coordinate sets for the four twin domains form two pairs, with more similarity within one pair than between the two pairs. This effect is most obvious in the large domain of MltA. Interestingly, in this part of the structure some of the peaks in the MAD-phased anomalous difference Fourier map were doublets. Overlaying this map on

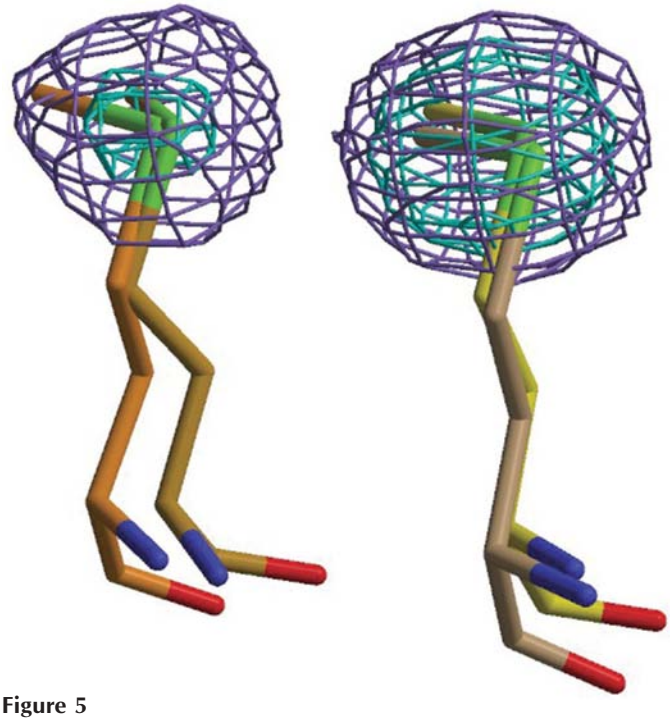

Figure 5
Double peak in the anomalous difference Fourier density phased with MAD phases. Four methionines, one each of the four twin domains, are shown. The blue and cyan contours denote contouring levels of 3σ and 4σ, respectively.

the superposition of the four twin domains explains this, as the two peaks of the doublets overlap almost perfectly with two pairs of selenium positions (Fig. 5).

4. Conclusion

As pointed out by Rosendal and coworkers, only particles with 222 point-group symmetry, such as the MltA tetrahedra, are likely to yield tetartohedrally twinned crystals (Rosendal *et al.*, 2004). Given that the only chiral space groups that can accommodate such twinning (Chandra *et al.*, 1999) account for only a small fraction of all protein crystals, one might expect that tetartohedral twinning is but an obscure rarity in protein crystallography. However, non-crystallographic 222 point-group symmetry is quite common in biological macromolecules and the possibility of tetartohedral twinning must therefore be kept in mind when working with such particles.

The MltA crystals do however show that tetartohedral twinning does not necessarily pose an insurmountable obstacle to structure solution. The 'overlap' of the selenium sites of the four twin domains probably facilitated the use of MAD phasing using the 'raw' twinned intensities. Moreover, owing to the versatility of the *CNS* language, existing protocols for hemihedral twin refinement could readily be adapted to accept the tetartohedrally twinned data. Thus, the case of MltA shows that even if protein crystals are tetartohedrally twinned, the data recorded from them can still be used for MAD phasing and a suitable refinement protocol is easily established. The altered *CNS* modules are available from the authors upon request or can be obtained from Crystallography Journals Online[1].

APPENDIX *A*
Adaptations made to *CNS* to accommodate tetartohedrally twinned data

The calculation of R factors whilst taking tetartohedral twinning into account was performed as follows,

$$R = \Bigg[\sum_{hkl} \Big| |F_{\rm obs}| - k\big(\alpha|F_{\rm calc}|^2 + \alpha' T'|F_{\rm calc}|^2 + \alpha'' T''|F_{\rm calc}|^2 + \alpha''' T'''|F_{\rm calc}|^2\big)^{1/2} \Big| \Bigg] \Big/ \sum_{hkl} |F_{\rm obs}|, \quad (4)$$

where T', T'' and T''' are the three twinning operators. In the case of MltA, all twinning fractions were set to 1/4 to model perfect tetartohedral twinning. Scale factors and partial structure factors modelling bulk solvent were also derived by adapting the relevant modules to use $(\alpha|F_{\rm calc}|^2 + \alpha' T'|F_{\rm calc}|^2 + \alpha'' T''|F_{\rm calc}|^2 + {\rm a}''' T'''|F_{\rm calc}|^2)^{1/2}$ instead of $|F_{\rm calc}|$.

The *CNS* refinement target for hemihedrally twinned data, t, and its derivative with respect to $F_{\rm calc}$, dt, can be adapted to comply with (2) as follows,

$$t = \frac{(|F_{\rm obs}| - k|F_{\rm calc}^{\rm tetartohedral}|)^2}{\sum_{hkl} |F_{\rm obs}|^2}, \quad (5)$$

$$\begin{aligned} dt = \frac{-2k}{\sum_{hkl} |F_{\rm obs}|^2} \Bigg\{ & \Bigg[(|F_{\rm obs}| - k|F_{\rm calc}^{\rm tetartohedral}|) \frac{\alpha F_{\rm calc}}{|F_{\rm calc}^{\rm tetartohedral}|} \Bigg] \\ & + T' \Bigg[(|F_{\rm obs}| - k|F_{\rm calc}^{\rm tetartohedral}|) \frac{\alpha' F_{\rm calc}}{|F_{\rm calc}^{\rm tetartohedral}|} \Bigg] \\ & + T'' \Bigg[(|F_{\rm obs}| - k|F_{\rm calc}^{\rm tetartohedral}|) \frac{\alpha'' F_{\rm calc}}{|F_{\rm calc}^{\rm tetartohedral}|} \Bigg] \\ & + T''' \Bigg[(|F_{\rm obs}| - k|F_{\rm calc}^{\rm tetartohedral}|) \frac{\alpha''' F_{\rm calc}}{|F_{\rm calc}^{\rm tetartohedral}|} \Bigg] \Bigg\} \\ & \times (|F_{\rm calc}^{\rm tetartohedral}| = [(\alpha|F_{\rm calc}|^2 + \alpha'|T' F_{\rm calc}|^2 + \alpha''|T'' F_{\rm calc}|^2 \\ & + \alpha'''|T''' F_{\rm calc}|^2)^{1/2}]. \end{aligned} \quad (6)$$

Note how the twinning operators also act on $|F_{\rm obs}|$ in the derivative dt. These formulae were built into the `refinementtarget_twin` module, to create `refinementtarget_twin_th`. The scripts for rigid-body refinement, energy minimization, B-factor refinement and simulated annealing were adapted to call the augmented routines.

To calculate maps from perfectly hemihedrally twinned data, *CNS* first detwins the observed structure factors, given the twinned data, the twinning operator and the model. We adapted the `data_detwin` module to account for tetartohedral twinning by using (7) to detwin the data,

$$\begin{aligned} F_{\rm obs}^{\rm detwinned} = (|F_{\rm obs}|^2 + k^2|F_{\rm calc}|^2 - k^2 T'|F_{\rm calc}|^2 - k^2 T''|F_{\rm calc}|^2 \\ - k^2 T'''|F_{\rm calc}|^2)^{1/2}/2^{1/2}. \end{aligned} \quad (7)$$

The resulting module `data_detwin_th` can then be called from the appropriately adapted script. These adaptations have been deposited as supplementary material.[1]

We would like to thank Drs A. Popov and P. Tucker (EMBL, Hamburg) for their help with collecting MAD data and Dr A. Brünger for useful comments.

References

Barends, T. R. M. & Dijkstra, B. W. (2003). *Acta Cryst.* D**59**, 2237–2241.

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.

Chandra, N., Acharya, K. R. & Moody, P. C. E. (1999). *Acta Cryst.* D**55**, 1750–1758.

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

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: WD5026). Services for accessing these data are described at the back of the journal.

D'Arcy, A., Elmore, C., Stihle, M. & Johnston, J. E. (1996). *J. Cryst. Growth*, **168**, 175–180.
Dauter, Z. (2003). *Acta Cryst.* D**59**, 2004–2016.
Declercq, J.-P. & Evrard, C. (2001). *Acta Cryst.* D**57**, 1829–1835.
Ito, N., Komiyama, N. H. & Fermi, G. (1995). *J. Mol. Biol.* **250**, 648–658.
de Jong, R. M., Tiesinga, J. J. W., Rozeboom, H. J., Kalk, K. H., Tang, L., Janssen, D. B. & Dijkstra, B. W. (2003). *EMBO J.* **22**, 4933–4944.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pohl, E., Gonzalez, A., Hermes, C. & van Silfhout, R. G. (2001). *J. Synchrotron Rad.* **8**, 1113–1120.
Rosendal, K. R., Sinning, I. & Wild, K. (2004). *Acta Cryst.* D**60**, 140–143.
Rudolph, M. G., Kelker, M. S., Schneider, T. R., Yeates, T. O., Oseroff, V., Heidary, D. K., Jennings, P. A. & Wilson, I. A. (2003). *Acta Cryst.* D**59**, 290–298.
Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **277**, 319–343.
Stanley, E. (1955). *Acta Cryst.* **8**, 351–352.
Stanley, E. (1972). *J. Appl. Cryst.* **5**, 191–194.
van Straaten, K. E., Dijkstra, B. W. & Thunnissen, M. (2004). *Acta Cryst.* D**60**, 758–760.
Terwilliger, T. C. (1999). *Acta Cryst.* D**55**, 1863–1871.
Terwilliger, T. C. (2003*a*). *Acta Cryst.* D**59**, 38–44.
Terwilliger, T. C. (2003*b*). *Acta Cryst.* D**59**, 45–49.
Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.
Terwisscha van Scheltinga, A. C., Valegård, K., Hajdu, J. & Andersson, I. (2003). *Acta Cryst.* D**59**, 2017–2022.
Terwisscha van Scheltinga, A. C., Valegård, K., Ramaswamy, S., Hajdu, J. & Andersson, I. (2001). *Acta Cryst.* D**57**, 1776–1785.
Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.
Valegård, K., Terwisscha van Scheltinga, A. C., Lloyd, M. D., Hara, T., Ramaswamy, S., Perrakis, A., Thompson, A., Lee, H. J., Baldwin, J. E., Schofield, C. J., Hajdu, J. & Andersson, I. (1998). *Nature (London)*, **394**, 805–809.
Yang, F., Dauter, Z. & Wlodawer, A. (2000). *Acta Cryst.* D**56**, 959–964.
Yeates, T. O. (1997). *Methods Enzymol.* **276**, 344–358.
Yeates, T. O. & Fam, B. C. (1999). *Structure*, **7**, R25–R29.

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

addenda and errata

Escherichia coli MltA: MAD phasing and refinement of a tetartohedrally twinned protein crystal structure. Addendum and erratum

Thomas R. M. Barends,‡ René M. de Jong,§ Karin E. van Straaten, Andy-Mark W. H. Thunnissen and Bauke W. Dijkstra*

Laboratory of Biophysical Chemistry, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

‡ Current address: Max-Planck Institute for Medical Research, Jahnstrasse 29, 69120 Heidelberg, Germany.
§ Current address: Department of Chemistry, Columbia University, 3000 Broadway, MC 3153, New York, NY 10027, USA.

Correspondence e-mail: b.w.dijkstra@rug.nl

In the paper by Barends *et al.* [(2005), *Acta Cryst.* D**61**, 613–621] errors were published in equations (4) and (6). The correct versions of these equations are given in this article. Addenda to the caption for Fig. 2 and one of the references in the article are also given.

The correct versions of equations (4) and (6) from the paper by Barends *et al.* (2005) are as follows,

$$R = \left[\sum_{hkl} \left| |F_{\rm obs}| - k\left(\alpha|F_{\rm calc}|^2 + \alpha' T' |F_{\rm calc}|^2 + \alpha'' T'' |F_{\rm calc}|^2 + \alpha''' T''' |F_{\rm calc}|^2\right)^{1/2} \right| \right] \Big/ \sum_{hkl} |F_{\rm obs}|, \quad (4)$$

$$dt = \frac{-2k}{\sum_{hkl} |F_{\rm obs}|^2} \cdot \left\{ \left[(|F_{\rm obs}| - k|F_{\rm calc}^{\rm tetartohedral}|) \frac{\alpha F_{\rm calc}}{|F_{\rm calc}^{\rm tetartohedral}|} \right] + T' \left[(|F_{\rm obs}| - k|F_{\rm calc}^{\rm tetartohedral}|) \frac{\alpha' F_{\rm calc}}{|F_{\rm calc}^{\rm tetartohedral}|} \right] + T'' \left[(|F_{\rm obs}| - k|F_{\rm calc}^{\rm tetartohedral}|) \frac{\alpha'' F_{\rm calc}}{|F_{\rm calc}^{\rm tetartohedral}|} \right] + T''' \left[(|F_{\rm obs}| - k|F_{\rm calc}^{\rm tetartohedral}|) \frac{\alpha''' F_{\rm calc}}{|F_{\rm calc}^{\rm tetartohedral}|} \right] \right\} \quad (6)$$

where

$$|F_{\rm calc}^{\rm tetartohedral}| = \left(\alpha|F_{\rm calc}|^2 + \alpha'|T'F_{\rm calc}|^2 + \alpha''|T''F_{\rm calc}|^2 + \alpha'''|T'''F_{\rm calc}|^2\right)^{1/2}.$$

In Fig. 2(*b*) of the original paper, a cutoff of 3σ was used for reflections between 2.3 and 2.15 Å resolution because of the weak data at this high resolution. Also an addition is made to the reference for van Straaten *et al.* (2004) which is given in full in the reference list below.

References

Barends, T. R. M., de Jong, R. M., van Straaten, K. E., Thunnissen, A.-M. W. H. & Dijkstra, B. W. (2005). *Acta Cryst.* D**61**, 613–621.
van Straaten, K. E., Dijkstra, B. W. & Thunnissen, A.-M. W. H. (2004). *Acta Cryst.* D**60**, 758–760.

  doi:10.1107/S0907444905013533