crystallization communications

Preliminary X-ray analysis of twinned crystals of the Q88Y25_Lacpl esterase from *Lactobacillus plantarum* WCFS1

Yanaisis Álvarez,[a,b] María Esteban-Torres,[c] Iván Acebrón,[a] Blanca de las Rivas,[c] Rosario Muñoz,[c] Martín Martínez-Ripoll[a] and José M. Mancheño[a]*

[a]Grupo de Cristalografía y Biología Estructural, Instituto Rocasolano, CSIC, Serrano 119, 28006 Madrid, Spain, [b]Centro de Estudios Avanzados de Cuba, CITMA, Carretera de S. Antonio km 2, 17100 Habana, Cuba, and [c]Laboratorio de Biotecnología Bacteriana, Instituto Ciencia y Tecnología de Alimentos y Nutrición (ICTAN), CSIC, Juan de la Cierva 3, 28006 Madrid, Spain

Correspondence e-mail: xjosemi@iqfr.csic.es

Q88Y25_Lacpl is an esterase produced by the lactic acid bacterium *Lactobacillus plantarum* WCFS1 that shows amino-acid sequence similarity to carboxyl-esterases from the hormone-sensitive lipase family, in particular the AFEST esterase from the archaeon *Archaeoglobus fulgidus* and the hyperthermophilic esterase EstEI isolated from a metagenomic library. N-terminally His$_6$-tagged Q88Y25_Lacpl has been overexpressed in *Escherichia coli* BL21 (DE3) cells, purified and crystallized at 291 K using the hanging-drop vapour-diffusion method. Mass spectrometry was used to determine the purity and homogeneity of the enzyme. Crystals of His$_6$-tagged Q88Y25_Lacpl were prepared in a solution containing 2.8 *M* sodium acetate trihydrate pH 7.0. X-ray diffraction data were collected to 2.24 Å resolution on beamline ID29 at the ESRF. The apparent crystal point group was 422; however, initial global analysis of the intensity statistics (data processed with high symmetry in space group *I*422) and subsequent tests on data processed with low symmetry (space group *I*4) showed that the crystals were almost perfectly merohedrally twinned. Most probably, the true space group is *I*4, with unit-cell parameters $a = 169.05$, $b = 169.05$, $c = 183.62$ Å.

1. Introduction

Carboxylester hydrolases are widespread enzymes that are produced by bacteria, archaea and eukarya (Hemilä *et al.*, 1994). According to their specificity, these proteins are commonly classified as true lipases (EC 3.1.1.3), which display maximum activity towards water-insoluble long-chain triglycerides, carboxylesterases (EC 3.1.1.1), which hydro-lyze (synthesize) ester-containing molecules that are at least partially soluble in water (*e.g.* below their critical micellar concentration), and various types of phospholipases (Arpigny & Jaeger, 1999). Generally, these proteins are classified into three families based on sequence homology: the C group, which includes cholinesterases and fungal lipases, the L group, which includes lipoprotein lipases and bacterial lipases, and the HSL group, whose name derives from the mammalian hormone-sensitive lipase (Holm *et al.*, 1988, 2000). The latter group currently includes several members with known crystal structures, such as brefeldin A from *Bacillus subtilis* (PDB entry 1jkm; Wei *et al.*, 1999), EST2 from *Alicyclobacillus acidocaldarius* (PDB entry 1evq; De Simone *et al.*, 2000), the AFEST esterase from the archaeon *Archaeoglobus fulgidus* (PDB entry 1jji; De Simone *et al.*, 2001) and the hyperthermophilic esterase EstEI isolated from a metagenomic library (PDB entry 2c7b; Byun *et al.*, 2007). Other members of this family have been identified by sequence analysis (Hemilä *et al.*, 1994). A *BLAST* search against the PDB using the sequence of Q88Y25_Lacpl from *Lactobacillus plantarum* WCFS1 as a query revealed 25% sequence identity (42% similarity) to EST2 and AFEST and 25% sequence identity (40% similarity) to EstEI.

Lipolytic enzymes from the C and L groups are included in the structural superfamily of α/β hydrolases (Ollis *et al.*, 1992; Schrag & Cygler, 1997; Heikinheimo *et al.*, 1999; Nardini & Dijkstra, 1999). The α/β-hydrolase fold exhibits a central mixed β-sheet surrounded on both sides by α-helices. The diverse enzymatic activities of the wide variety of enzymes comprising this superfamily rely mainly on a catalytic triad formed by Ser, His and Asp (Glu) residues, reminiscent of serine proteinases, that is responsible for nucleophilic attack on the

carbonyl C atom of the scissile ester bond. Typically, the catalytic serine residue usually appears in a conserved Gly-Xaa-Ser-Xaa-Gly peptapeptide at the apex of a sharp turn: the nucleophilic elbow (Derewenda & Sharp, 1993). In contrast to proteins from the C and L groups, members of the HSL family contain a cap domain consisting of two separate helical regions located in the upper region of the central β-sheet.

In this work, we report the production, purification, crystallization and preliminary X-ray crystallographic analysis of His$_6$-tagged Q88Y25_Lacpl. The results obtained indicate that His$_6$-tagged Q88Y25_Lacpl crystals present almost perfect merohedral twinning.

2. Materials and methods

2.1. Cloning and overexpression

The *lp_0973* gene (coding for the Q88Y25_Lacpl esterase) from *L. plantarum* WCFS1 was PCR-amplified by HS Prime Start DNA polymerase (Takara) using the primers 563 (5′-*GGTGAAAACC-TGTATTTCCAGGGC*atgccaacaattaattcgattcaaa-3′) and 564 (5′-*A-TCGATAAGCTTAGTTAGCTATTA*ctaaattaacgcggccgccatcact-3′) (the nucleotides pairing the expression-vector sequence are indicated in italics and the nucleotides pairing the *lp_0973* gene sequence are written in lowercase letters). The 1.1 kb purified PCR product was inserted into the pURI3-TEV vector using the restriction enzyme-free and ligation-free cloning strategy described by Geiser *et al.* (2001). Expression vector pURI3-TEV was constructed based on the commercial expression vector pT7-7 (USB) but contained the leader sequence MGGSHHHHHHGENLYFQG consisting of an N-terminal methionine followed by three spacer amino acids, a His$_6$ tag, a spacer glycine residue and the TEV recognition site (Curiel *et al.*, 2011). Thus, the final recombinant His$_6$-tagged Q88Y25_Lacpl consisted of 355 amino-acid residues with a molecular weight of 38.7 kDa. *Escherichia coli* DH5α cells were transformed and recombinant plasmids containing the correct insert (pURI3TEV-Q88Y25_Lacpl) as identified by restriction-enzyme analysis and further verified by DNA sequencing were isolated. Subsequently, pURI3TEV-Q88Y25_Lacpl and pGro7 (Takara Bio Inc.) vectors were co-transformed into *E. coli* BL21 (DE3) competent cells for co-overexpression of GroES/GroEL and the esterase Q88Y25_Lacpl. Ampicillin and chloramphenicol were used as selection markers.

Cells carrying the latter two plasmids (pURI3TEV-Q88Y25_Lacpl and pGro7) were grown at 310 K in Luria–Bertani medium containing ampicillin (100 µg ml^{-1}), chloramphenicol (20 µg ml^{-1}) and L-arabinose (2 mg ml^{-1}) until they reached an optical density at 600 nm of 0.4–0.6. Induction of Q88Y25_Lacpl overexpression was carried out by addition of IPTG (0.4 m*M* final concentration). After induction, the cells were grown at 295 K for 20 h and collected by centrifugation using a Beckman Coulter J-25 Avanti centrifuge (7500*g* for 15 min at 277 K).

2.2. Purification

The cells were resuspended in 20 ml buffer *A* (50 m*M* sodium phosphate pH 7.0 containing 300 m*M* NaCl) per litre of cell culture and disrupted with a French press. The lysate was centrifuged at 17 400*g* for 40 min at 277 K using a Beckman Coulter J-25 Avanti centrifuge. After filtration through a 0.22 µm filter (Millipore), the supernatant was gently mixed for 20 min at room temperature with 1 ml TALON resin (Clontech) previously equilibrated in buffer *A*. The mixture was then loaded onto an Econo-Column (1.5 × 15 cm; Bio-Rad) and after sedimentation of the resin the unbound material was eluted. The resin was then washed with five column volumes

(CV) of buffer *A* and with another 5 CV of buffer *B* (50 m*M* sodium phosphate pH 7.0 containing 300 m*M* NaCl and 10 m*M* imidazole). The recombinant His$_6$-tagged protein was eluted with 5 CV of elution buffer (50 m*M* sodium phosphate pH 7.0 containing 300 m*M* NaCl and 150 m*M* imidazole). The eluate containing His$_6$-tagged Q88Y25_Lacpl was dialysed overnight at 277 K against 20 m*M* Tris–HCl pH 8.0 containing 100 m*M* NaCl and 0.04%(*w/v*) sodium azide and further concentrated by ultrafiltration with a YM-10 membrane (Amicon). A final polishing gel-filtration step was carried out on Superdex 75 prep-grade HiLoad 16/60 (GE Healthcare) with a Bio-Rad BioLogic DuoFlow FPLC system. Protein purity was checked by SDS–PAGE at the various stages of the purification process (Fig. 1). The His$_6$-tagged Q88Y25_Lacpl solution used for crystallization trials was finally prepared in 20 m*M* Tris pH 8.0 containing 0.1 *M* NaCl and 0.04%(*w/v*) sodium azide. The protein concentration was determined by UV–Vis absorbance measurements with a Nanodrop ND-1000 spectrophotometer, using an extinction coefficient of 1.03 (1 mg ml^{-1}, 1 cm, 280 nm).

2.3. Mass spectrometry

Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI–TOF MS) on a Finnigan LCQ Deca ion-trap mass spectrometer (Thermo Electron, San José, California, USA) was used to determine the molecular mass of Q88Y25_Lacpl. Mass spectra were recorded in full scan mode (m/z = 450–2000) and protein peaks detected were deconvoluted using the *BIOMASS* deconvolution tool from the *BioWorks* v.3.1 software (Thermo Fisher Scientific).

2.4. Crystallization

Initial crystallization conditions for His$_6$-tagged Q88Y25_Lacpl at 291 K were determined by the sparse-matrix method (Jancarik & Kim, 1991) with commercial screens from Hampton Research (Riverside, California, USA) and Qiagen in crystallization trials using the sitting-drop vapour-diffusion method in Innovaplate SD-2 96-well plates set up using a Nanodrop Innovadyne robot. A total of five plates (480 crystallization droplets) were used. Each drop contained
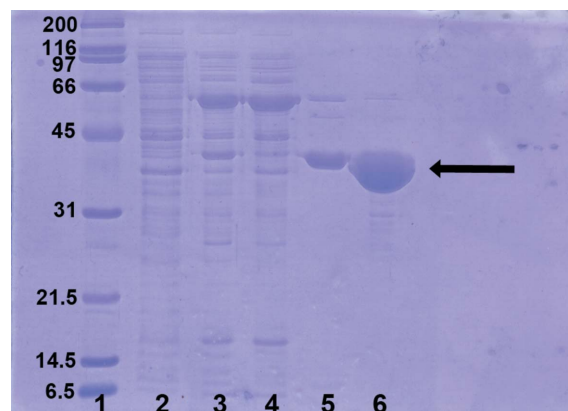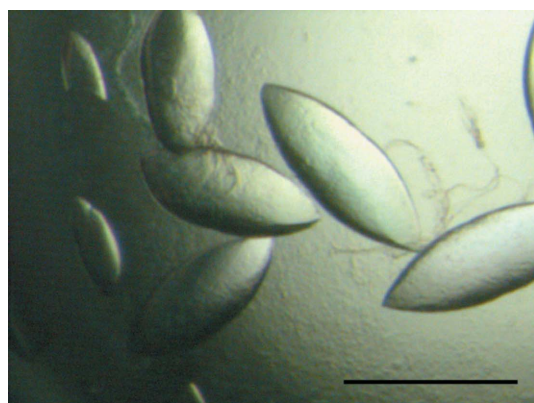

Figure 1
SDS–PAGE of His$_6$-tagged Q88Y25_Lacpl samples at different steps of the purification process. Lane 1, broad-range molecular-weight markers (Bio-Rad; labelled in kDa); lane 2, soluble fraction from *E. coli* BL21 (DE3) cells carrying the plasmid pURI3TEV; lane 3, soluble fraction from *E. coli* BL21 (DE3) cells carrying the plasmid pURI3TEV-Q88Y25_Lacpl after cell disruption; lane 4, flowthrough from the TALON step; lane 5, pooled fractions of the esterase eluted from TALON resin with buffer containing 150 m*M* imidazole; lane 6, pooled fractions from gel-filtration chromatography on Superdex 75 prep-grade HiLoad 16/60 (GE Healthcare).

250 nl protein (9 mg ml^{-1}) in Tris–HCl buffer [20 mM Tris–HCl pH 8.0 containing 0.1 M NaCl and 0.04%(w/v) sodium azide] and 250 nl reservoir solution. Drops were equilibrated against 65 µl reservoir solution. His$_6$-tagged Q88Y25_Lacpl crystals appeared in 2–3 d in condition Nos. 24 and 28 from the Index HT screen (Hampton Research), consisting of 2.8 M sodium acetate trihydrate pH 7.0 and 35%(v/v) Tacsimate pH 7.0, respectively. Scaling up of the crystallization conditions using hanging drops in 24-well VDX plates (1:1 protein:precipitant volume ratio; total volume 2 µl) rendered high-quality diffracting crystals (0.3–0.4 mm in the longest dimension) in 2.8 M sodium acetate trihydrate pH 7.0.

2.5. X-ray diffraction analysis and processing

For X-ray diffraction measurements, His$_6$-tagged Q88Y25_Lacpl crystals were transferred to an optimized cryoprotectant solution [2.8 M sodium acetate trihydrate pH 7.0 containing 20%(v/v) glycerol] for ~5 s and then flash-cooled at 100 K in a stream of liquid nitrogen. Diffraction data were collected using a Pilatus 6M detector (Dectris Ltd, Switzerland) on beamline ID29 at the European Synchrotron Radiation Facility (ESRF, Grenoble, France). The crystal-to-detector distance was maintained at 266.7 mm and the wavelength was 1.272 Å. The 1500 images collected at a speed of 5 Hz with an oscillation angle of 0.1° were processed using XDS (Kabsch, 2010) and scaled with $SCALA$ from the CCP4 program suite (Winn *et al.*, 2011). Data-collection statistics are given in Table 1.
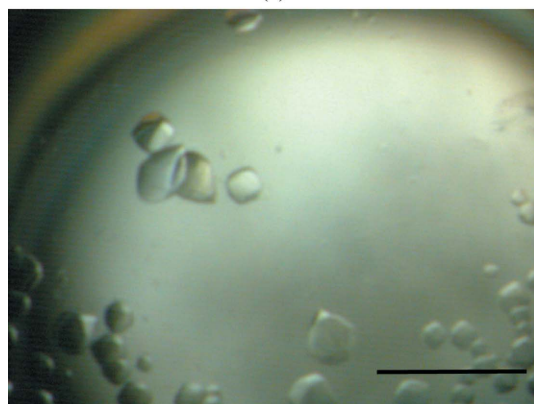
Figure 2
Crystals of recombinant His$_6$-tagged Q88Y25_Lacpl from *L. plantarum* WCFS1 grown at 291 K in (*a*) 2.8 M sodium acetate trihydrate pH 7.0 and (*b*) 35%(v/v) Tacsimate pH 7.0. The scale bar indicates 0.3 mm in both cases.

Table 1
Data-collection and processing statistics for the Q88Y25_Lacpl crystals.

The data correspond to a single data set processed in two different space groups. Values in parentheses are for the outermost shell.

Space group	I4	I422
Unit-cell parameters		
a (Å)	169.05	169.05
b (Å)	169.05	169.05
c (Å)	183.62	183.62
Resolution range (Å)	47.57–2.24 (2.36–2.24)	47.57–2.24 (2.36–2.24)
No. of measured reflections	695698 (103721)	690251 (103247)
No. of unique reflections	123334 (17985)	63786 (9204)
Completeness (%)	100 (100)	100 (100)
Multiplicity	5.6 (5.8)	10.8 (11.2)
Average $I/\sigma(I)$	13.3 (3.7)	14.3 (5.0)
R_{merge}† (%)	9.0 (46.3)	12.1 (49.7)
R_{meas}‡ (%)	9.9 (50.9)	12.7 (52.0)

† $R_{\mathrm{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl} \sum_i I_i(hkl)$. ‡ R_{meas} is the redundancy-independent R factor on intensities.

The *phenix_xtriage* module of the *PHENIX* program (Adams *et al.*, 2010) was used to analyse the collected data for twinning, which was first suspected from the intensity distribution of the processed data.

3. Results and discussion

The *lp_0973* gene from *L. plantarum* WCFS1 was cloned and the N-terminally His$_6$-tagged Q88Y25_Lacpl protein was overexpressed in *E. coli* BL21 (DE3) cells. Initial small-scale expression tests revealed that the production yield of soluble esterase increases significantly when it is co-overexpressed with the molecular chaperones GroES/GroEL from *E. coli* (Nishihara *et al.*, 2000). Expression levels of the esterase were highest ~20 h after induction with IPTG. The final production yield was approximately 5 mg per litre of culture.
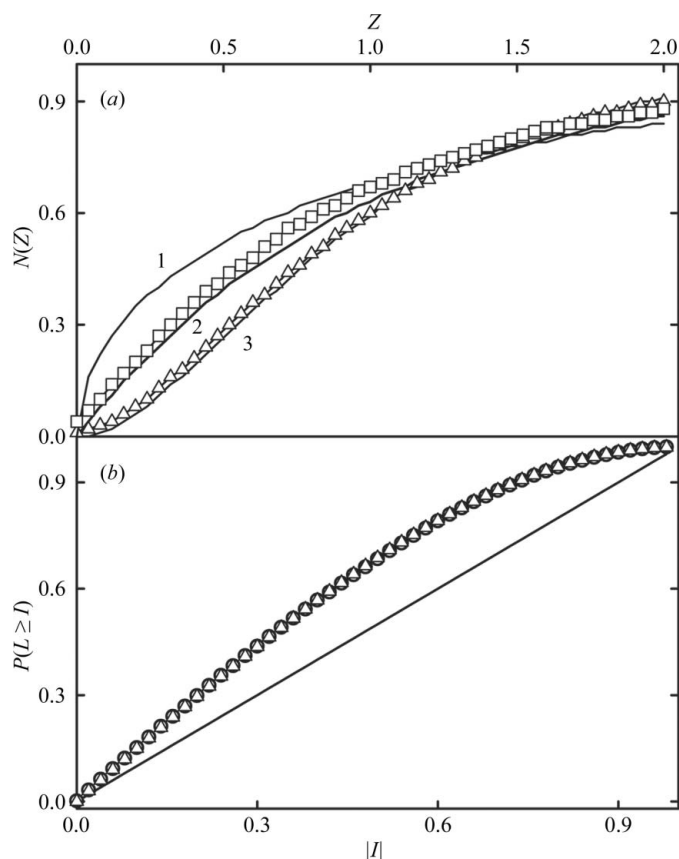
The protein was purified by immobilized metal-affinity chromatography on TALON and further polished on Superdex 75 prep-grade HiLoad 16/60 (GE Healthcare) (Fig. 1). The purity and the homogeneity of Q88Y25_Lacpl thus obtained were checked by MALDI–TOF mass spectrometry, which revealed a mass value of m/z 38 622 Da, which is in agreement with the calculated average mass of 38 628 Da for recombinant His-tagged Q88Y25_Lacpl without the amino-terminal Met residue. Therefore, the final recombinant esterase would contain 354 amino acids, namely the leader sequence (see §2) plus the complete Q88Y25_Lacpl sequence.

Lens-shaped crystals of Q88Y25_Lacpl grew from precipitant solution consisting of 2.8 M sodium acetate trihydrate pH 7.0. Scaled and optimized trials resulted in large crystals approaching 0.3–0.4 mm in size (Fig. 2), which were subsequently analysed on beamline ID29 at the ESRF. Autoindexing performed with XDS (Kabsch, 2010) suggested a body-centred tetragonal lattice with unit-cell parameters $a = 169.05$, $b = 169.05$, $c = 183.62$ Å. The suggested point group of the lattice was 422, which is in agreement with the result obtained with *POINTLESS* (Evans, 2006).

Reasonable processing statistics could be obtained in space groups *I*422 and *I*4 (Table 1). However, intensity statistical indicators from *CTRUNCATE* showed the first indications of almost perfect twinning. For instance, the cumulative intensity distribution [$N(Z)$ test] for acentric and centric reflections (Fig. 3a) revealed an observed distribution superimposable with that expected for perfect twinning. Moreover, the values of the first and third moments of E (acentric reflections) were 0.935 (0.886 for untwinned and 0.940 for perfectly twinned) and 1.191 (1.329 for untwinned and 1.175 for perfectly

twinned), respectively. Other twinning tests were performed with the data set processed under low symmetry (*I*4) using the *phenix_xtriage* module of the *PHENIX* program (Adams *et al.*, 2010). The multivariate *Z* score from the *L*-test (Padilla & Yeates, 2003) was 12.1%, well above the expected value of 3.5%. A high value of the *Z* score shows that data are far from what would be expected for untwinned data. In fact, the *L* curve for acentric data processed under low symmetry (*I*4; Fig. 3*b*) provided a mean $|L|$ value of 0.375 (0.500 for untwinned and 0.375 for a perfect twin) and a mean L^2 value of 0.201 (0.333 for untwinned and 0.200 for a perfect twin), suggesting almost perfect twinning. Other twinning tests carried out with *PHENIX* are shown in Table 2. The most probable content of the asymmetric unit, as estimated using the Matthews probability calculator (Kantardjieff & Rupp, 2003) under the assumption that the true space group is *I*4, is 7–8 molecules, with a solvent content of between 49% (V_M = 2.42 Å^3 Da^{-1}) and 42% (V_M = 2.12 Å^3 Da^{-1}), respectively.

Since all attempts to solve the crystal structure of Q88Y25_Lacpl by molecular replacement using the atomic coordinates of EST2 (PDB entry 1evq; De Simone *et al.*, 2000), AFEST (PDB entry 1jji; De Simone *et al.*, 2001) and EstEI (PDB entry 2c7b; Byun *et al.*, 2007)

Table 2
Results of the twinning tests for the Q88Y25_Lacpl data processed in space group *I*4 carried out with *phenix_xtriage*.

The expected values for twinning tests are $\langle I^2 \rangle / \langle I \rangle^2$ of 2.000 for untwinned and 1.500 for a perfect twin; $\langle F \rangle^2 / \langle F^2 \rangle$ of 0.785 for untwinned and 0.885 for a perfect twin; $\langle |E^2 - 1| \rangle$ of 0.736 for untwinned and 0.541 for a perfect twin.

Likely space group	*I*4
Twinning tests, for acentric data	
$\langle I^2 \rangle / \langle I \rangle^2$	1.537
$\langle F \rangle^2 / \langle F^2 \rangle$	0.879
$\langle |E^2 - 1| \rangle$	0.543
Multivariate *Z* score from *L*-test	12.128
Possible twin operator	$-h, k, -l$

as search models failed, we are currently producing the SeMet derivative of Q88Y25_Lacpl in order to solve its structure by MAD. Nevertheless, considering the high degree of twinning of the obtained crystals and to avoid significant errors in the output amplitudes, we are also working on producing crystals of Q88Y25_Lacpl complexed with substrate analogues as well as new recombinant forms of the esterase that would permit us to prepare novel untwinned crystal forms.

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Arpigny, J. L. & Jaeger, K.-E. (1999). *Biochem. J.* **343**, 177–183.
Byun, J.-S., Rhee, J.-K., Kim, N. D., Yoon, J. H., Kim, D.-U., Koh, E., Oh, J.-W. & Cho, H.-S. (2007). *BMC Struct. Biol.* **7**, 47.
Curiel, J. A., de las Rivas, B., Mancheño, J. M. & Muñoz, R. (2011). *Protein Expr. Purif.* **76**, 44–53.
Derewenda, Z. S. & Sharp, A. M. (1993). *Trends Biochem. Sci.* **18**, 20–25.
De Simone, G., Galdiero, S., Manco, G., Lang, D., Rossi, M. & Pedone, C. (2000). *J. Mol. Biol.* **303**, 761–771.
De Simone, G., Menchise, V., Manco, G., Mandrich, L., Sorrentino, N., Lang, D., Rossi, M. & Pedone, C. (2001). *J. Mol. Biol.* **314**, 507–518.
Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.
Geiser, M., Cèbe, R., Drewello, D. & Schmitz, R. (2001). *Biotechniques*, **31**, 88–92.
Heikinheimo, P., Goldman, A., Jeffries, C. & Ollis, D. L. (1999). *Structure*, **7**, 141–146.
Hemilä, H., Koivula, T. T. & Palva, I. (1994). *Biochim. Biophys. Acta*, **1210**, 249–253.
Holm, C., Kirchgessner, T. G., Svenson, K. L., Fredrikson, G., Nilsson, S., Miller, C. G., Shively, J. E., Heinzmann, C., Sparkes, R. S., Mohandas, T., Lusis, A. J., Belfrage, P. & Schotz, M. C. (1988). *Science*, **241**, 1503–1506.
Holm, C., Osterlund, T., Laurell, H. & Contreras, J. A. (2000). *Annu. Rev. Nutr.* **20**, 365–393.
Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Kantardjieff, K. A. & Rupp, B. (2003). *Protein Sci.* **12**, 1865–1871.
Nardini, M. & Dijkstra, B. W. (1999). *Curr. Opin. Struct. Biol.* **9**, 732–737.
Nishihara, K., Kanemori, M., Yanagi, H. & Yura, T. (2000). *Appl. Environ. Microbiol.* **66**, 884–889.
Ollis, D. L., Cheah, E., Cygler, M., Dijkstra, B., Frolow, F., Franken, S. M., Harel, M., Remington, S. J., Silman, I., Schrag, J., Sussman, J. L., Verschueren, K. H. G. & Goldman, A. (1992). *Protein Eng.* **5**, 197–211.
Padilla, J. E. & Yeates, T. O. (2003). *Acta Cryst.* D**59**, 1124–1130.
Schrag, J. D. & Cygler, M. (1997). *Methods Enzymol.* **284**, 3–27.
Wei, Y., Contreras, J. A., Sheffield, P., Osterlund, T., Derewenda, U., Kneusel, R. E., Matern, U., Holm, C. & Derewenda, Z. S. (1999). *Nature Struct. Biol.* **6**, 340–345.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.

Figure 3
(*a*) Cumulative intensity distribution from *CTRUNCATE* for centric and acentric reflections processed in space group *I*4. *Z* represents normalized reduced intensity values and *N*(*Z*) is the normalized number of reflections > *Z*. Curve 1 represents the expected result for untwinned centric reflections; curves 2 and 3 represent the results expected for untwinned and perfectly twinned acentric reflections, respectively. Open squares represent the observed results for centric reflections and open triangles those obtained for acentric reflections. (*b*) Analysis of local intensity differences (Padilla & Yeates, 2003) for acentric reflections processed in space group *I*4. This panel represents the cumulative probability distribution, $P(L \geq l)$, of the parameter $|I|$ (or $|L|$ in Padilla & Yeates, 2003). Open circles represent the experimental data. The diagonal line represents the expected result for untwinned data and the curve with open triangles represents the expected result for perfectly twinned data. Values were generated with *phenix_xtriage*.