Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Determining the DUF55-domain structure of human thymocyte nuclear protein 1 from crystals partially twinned by tetartohedry

Feng Yu,[a,e] Aixin Song,[b] Chunyan Xu,[a] Lihua Sun,[a,e] Jian Li,[a,e] Lin Tang,[a] Minmin Yu,[c] Todd O. Yeates,[d] Hongyu Hu[b] and Jianhua He[a]*

[a]Shanghai Institute of Applied Physics, Chinese Academy of Sciences, People's Republic of China, [b]State Key Laboratory of Molecular Biology, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, People's Republic of China, [c]Physical Biosciences Division, Lawrence Berkeley National Laboratory, USA, [d]Department of Chemistry and Biochemistry, University of California, USA, and [e]Graduate School, Chinese Academy of Sciences, People's Republic of China

Correspondence e-mail: hejh@sinap.ac.cn

Received 2 October 2008
Accepted 8 December 2008

PDB Reference: DUF55 domain of human thymocyte nuclear protein 1, 3eop, r3eopsf.

Human thymocyte nuclear protein 1 contains a unique DUF55 domain consisting of 167 residues (55–221), but its cellular function remains unclear. Crystals of DUF55 belonged to the trigonal space group $P3_1$, but twinning caused the data to approach apparent 622 symmetry. Two data sets were collected to 2.3 Å resolution. Statistical analysis confirmed that both data sets were partially twinned by tetartohedry. Tetartohedral twin fractions were estimated. After the structure had been determined, only one twofold axis of rotational pseudosymmetry was found in the crystal structure. Using the *DALI* program, a YTH domain, which is a potential RNA-binding domain from human YTH-domain-containing protein 2, was identified as having the most similar three-dimensional fold to that of DUF55. It is thus implied that DUF55 might be a potential RNA-related domain.

1. Introduction

Partly as a consequence of the large-scale use of synchrotron radiation and improvements in software tools, thousands of new protein structures are being determined each year. However, not all protein crystal structures can be determined using routine procedures; crystal twinning leads to many difficult cases (Yeates, 1997; Parsons, 2003). The causes of crystal twinning are not fully understood, but more than half of the reported cases occur in combination with pseudo-symmetry (Lebedev *et al.*, 2006). In recent years, several interesting and complicated cases have been successfully determined (*e.g.* Gayathri *et al.*, 2007; MacRae & Doudna, 2007; Anand *et al.*, 2007; see also the review by Dauter *et al.*, 2005). In the case of merohedral twinning, the multiple twin domains within a single-crystal specimen cannot usually be distinguished by visual microscopic examination. Furthermore, because the reciprocal lattices of the twin domains overlap exactly in merohedral twinning, the observed diffraction pattern is typically unremarkable. Therefore, in order to detect twinning, the statistics of the intensity data must be examined (Stanley, 1972; Rees, 1980; Yeates, 1988; Padilla & Yeates, 2003).

To determine structures from twinned crystals, it is essential to identify the true space group, determine the underlying twin operation(s) [*i.e.* the symmetry operation(s) relating the twin domains] and estimate the twin fractions (the volume ratios of the component twin domains). Once the true space group has been confirmed, data reduction can be carried out as for ordinary diffraction data. Twinning does not usually prevent structure determination by molecular replacement (Chandra *et al.*, 1999; Yeates, 1997). Phasing highly twinned data using

MIR/MAD/SAD methods tends to be more challenging; however, this has also been accomplished successfully in several cases (Barends *et al.*, 2005; Yang *et al.*, 2000; Dauter, 2003; Yeates & Rees, 1987; Sultana *et al.*, 2007; Larsen & Harrison, 2004; Toms *et al.*, 2004; MacRae & Doudna, 2007; Rudolph *et al.*, 2003). In cases of twinning, atomic refinement can be carried out in various ways. For example, the effects of twinning can be applied to the model intensities or the observed intensities can be corrected or 'detwinned' in order to obtain estimates of the true crystallographic intensities for comparison to the model intensities. Most reported cases of twinning in macromolecular crystals only involve two twin-domain orientations (hemihedral twinning), but tetartohedral twinning involving four twin domains (*e.g. P*3 twinned towards apparent *P*622) is also possible (Rosendal *et al.*, 2004; Barends *et al.*, 2005; Gayathri *et al.*, 2007; Anand *et al.*, 2007); an even higher category, ogdohedral twinning, with eight twin domains (*e.g. P*3 twinned towards apparent *P*6/*mmm*), is also possible, although only for achiral or racemic mixtures of molecules. In the case of tetartohedral twinning, four twin domains are related by three twin operations and the observed intensity of each reflection is a weighted sum of four twin-related reflection intensities,

$$J_{\rm obs} = \alpha_1 I_1 + \alpha_2 I_2 + \alpha_3 I_3 + \alpha_4 I_4, \qquad (1)$$

where $J_{\rm obs}$ is an observed (twinned) intensity, the I_k are true reflection intensities and the α_k are the four tetartohedral twin fractions (which must sum to unity). This paper reports the structure of the DUF55 domain of human thymocyte nuclear protein 1 (hTHYN1), as determined from partially tetartohedrally twinned crystals using molecular replacement and detwinned data.

hTHYN1 was first identified in human CD34$^+$ haematopoietic stem/progenitor cells and was previously named HSPC144 (Zhang *et al.*, 2000); its biological function is unknown. It contains 225 amino-acid residues and is highly conserved in a wide range of species including yeast, plants and vertebrates. Previous studies of chicken and mouse THYN1 proteins have indicated that THYN1 plays a role in apoptosis, but the mechanism is still unclear (Compton *et al.*, 2001; Jiang, Toyota, Yoshimoto *et al.*, 2003). There are some differences in the tissue distributions of chicken and mouse THYN1 (Miyaji *et al.*, 2002), but the THYN1 protein mainly exists in the nucleus (Jiang, Toyota, Takada *et al.*, 2003). The recombinant full-length hTHYN1 protein is unstable at high concentrations and is therefore not well suited for structural analysis. Using limited proteolysis and mass spectrometry, two stable domains, HSPC144-P (residues 44–225) and DUF55 (residues 54–221; formerly DUF589), have been identified (Song *et al.*, 2005). DUF55 is a unique domain of undefined function (according to the Pfam database; Finn *et al.*, 2006) which exhibits a highly conserved sequence in eukaryotes. Elucidation of the domain structure of hTHYN1 may help us to understand the function of these proteins in apoptosis.

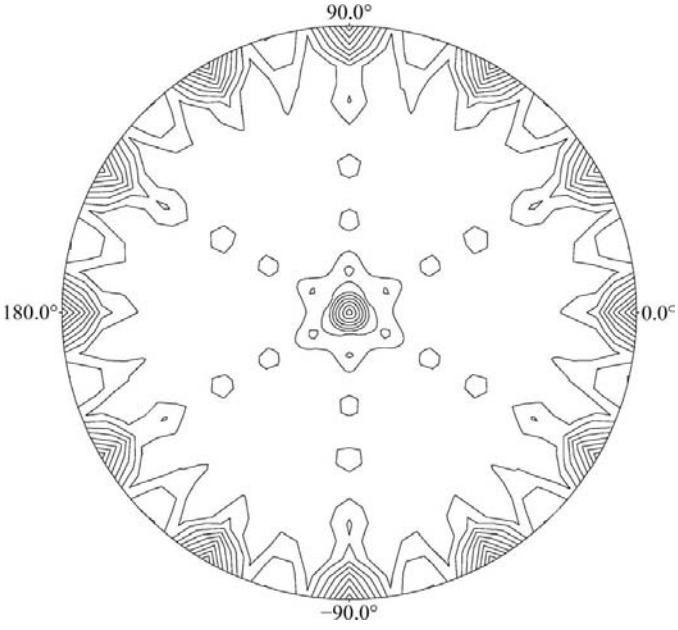

Figure 1
Self-rotation function plot of data set 1 ($\kappa = 180^\circ$).

2. Materials and methods

2.1. Protein expression and purification

DUF55 was expressed and purified as described previously (Song *et al.*, 2005). Briefly, the DUF55 DNA sequence (residues 55–221) was cloned into plasmid pET22b, which produces a C-terminally His-tagged protein. DUF55 was purified using metal-chelating chromatography, SP cation-exchange chromatography (Amersham) and gel filtration on Amersham HiLoad Superdex 75. Finally, the protein was concentrated in a buffer consisting of 20 m*M* HEPES pH 7.2, 150 m*M* NaCl, 5% glycerol and 2 m*M* DTT.

2.2. Protein crystallization

DUF55 at a concentration of 10 mg ml^{-1} was crystallized using the hanging-drop vapour-diffusion method at 291 or 277 K. After screening numerous conditions, the following crystallization condition was identified: a hanging drop containing 1 µl protein solution, 1 µl reservoir solution (0.1 *M* sodium acetate pH 4.0–4.8, 28% PEG 2000 MME, 200 m*M* ammonium sulfate) and 0.2 µl 30% 1,6-diaminohexane as an additive (1,6-diaminohexane is a basic reagent, so that the actual pH is about 10.9). After a week, fan-like crystals appeared. The dimensions of the crystals obtained at 291 K (0.45 × 0.15 × 0.08 mm) were notably larger than those obtained at 277 K (0.30 × 0.07 × 0.02 mm). The crystals were soaked in a cryoprotectant solution (0.1 *M* sodium acetate pH 4.0–4.8, 42% PEG 2000 MME, 75 m*M* ammonium sulfate, 75 m*M* NaCl, 15% glycerol, 3% 1,6-diaminohexane) and flash-frozen in liquid nitrogen.

2.3. Data collection

The diffraction of the DUF55 crystals was too weak for the collection of usable data sets using a rotating-anode X-ray generator. Two data sets (named data set 1 and data set 2)

Table 1
Results of molecular replacement of data set 1.

	$P3_1$	$P3_121$	$P3_112$	$P6_4$
LLG	182	49	13	5
Before refinement				
$R_{\rm twin}$†	0.3083	0.4569	0.4515	0.4475
Free $R_{\rm twin}$	0.3140	0.4937	0.4633	0.4302
After twinning refinement‡				
$R_{\rm twin}$	0.2400	0.3815	0.3892	0.3620
Free $R_{\rm twin}$	0.2984	0.5077	0.4743	0.4384

† In this paper, $R_{\rm twin}$ for hemihedral twinning was calculated as $R_{\rm twin} = \sum ||F_{\rm obs}| - [(1-\alpha)|F_{\rm calc}+F_{\rm bulk}|^2 + \alpha T|F_{\rm calc}+F_{\rm bulk}|^2]^{1/2}| / \sum |F_{\rm obs}|$. $R_{\rm twin}$ for tetartohedral twinning was calculated as $R_{\rm twin} = \sum ||F_{\rm obs}| - (\alpha_1|F_{\rm calc}+F_{\rm bulk}|^2 + \alpha_2 T_1|F_{\rm calc}+F_{\rm bulk}|^2 + \alpha_3 T_2|F_{\rm calc}+F_{\rm bulk}|^2 + \alpha_4 T_3|F_{\rm calc}+F_{\rm bulk}|^2)^{1/2}| / \sum |F_{\rm obs}|$, where T_1, T_2 and T_3 are the twin operations that apply to the structure factors, $F_{\rm calc}$ is the structure-factor array for the macromolecular model including ordered solvent and $F_{\rm bulk}$ is the structure-factor array for an appropriate model of disordered solvent (Barends *et al.*, 2005). ‡ Perfect twinning was assumed in all cases. For hemihedral twinning, refinements were carried out with *CNS* and `anneal_twin.inp/bindividual_twin.inp`. For tetartohedral twinning, refinement scripts for perfect tetartohedral twinning were developed by Barends *et al.* (2005).

were collected to 2.3 Å resolution on beamline 3W1A at the Beijing Synchrotron Radiation Facility and processed using *MOSFLM* (Leslie, 2006) and *SCALA* (Collaborative Computational Project, Number 4, 1994).

2.4. Data analysis and molecular replacement

Initially, the data sets were processed in space group $P6_222$/$P6_422$ according to the self-rotation function plots (Fig. 1) and the detected systematic absences. However, crystal twinning was indicated by an impossibly low value of the Matthews coefficient in this symmetry (1.16 Å^3 Da^{-1} for one molecule per asymmetric unit). The presence of twinning was confirmed by examination of the cumulative intensity distribution (Fig. 2) and Padilla–Yeates local intensity statistics (Fig. 3). Plausible lower symmetry space groups included $P3_1$, $P3_2$, $P3_112$/$P3_212$, $P3_121$/$P3_221$ and $P6_2$/$P6_4$. Molecular-replacement searches were therefore carried out under all possible symmetries using

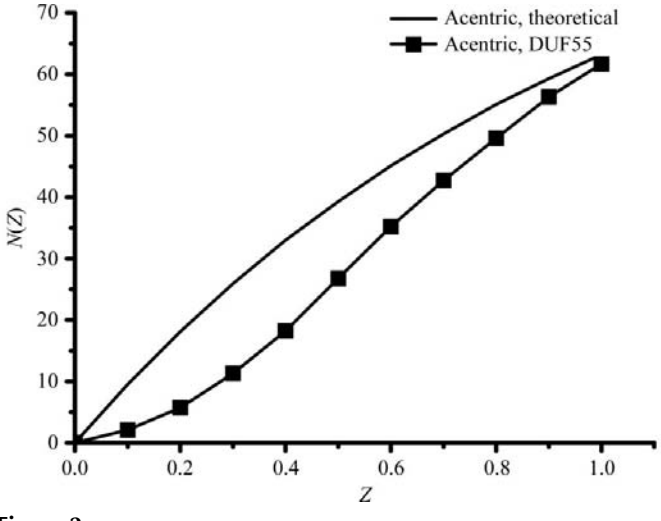

Figure 2
Cumulative intensity distributions for acentric reflections (cumulative Wilson distributions). For the DUF55 data (data set 1) the distribution is strongly sigmoidal, which is indicative of twinning.

Table 2
Twinning tests.

	Data set 1	Data set 2
Perfect twinning test		
$\langle\|I\|^2\rangle/(\langle\|I\|\rangle)^2$†	1.5665	1.4530
$(\langle\|F\|\rangle)^2/\langle\|F\|^2\rangle$‡	0.8818	0.9022
H test§		
$-h, -k, l$	0.327	0.447
$k, h, -l$	0.460	0.449
$-k, -h, -l$	0.327	0.439
Estimated tetartohedral twin fractions (Yeates & Yu, 2008)		
Solution 1 ($\alpha_1, \alpha_2, \alpha_3, \alpha_4$)	0.424, 0.300, 0.134, 0.142	0.349, 0.218, 0.209, 0.224
Solution 2 ($\alpha_1, \alpha_2, \alpha_3, \alpha_4$)	0.076, 0.200, 0.366, 0.358	0.151, 0.282, 0.291, 0.276

† 2.0 for untwinned, 1.5 for perfectly hemihedrally twinned and 1.25 for perfectly tetartohedrally twinned (Stanley, 1972). $\langle|I|^2\rangle/(\langle|I|\rangle)^2$ for perfect tetartohedral twinning was calculated as $\int_0^{+\infty} z^2 \times [4^4/\Gamma(4)]z^3 \exp(-4z)\,dz = \int_0^{+\infty} (128/3)z^5 \exp(-4z)\,dz = 1.25$. ‡ 0.785 for untwinned, 0.885 for perfectly hemihedrally twinned and 0.940 for perfectly tetartohedrally twinned (Stanley, 1972). The Wilson ratio for perfect tetartohedral twinning was calculated as $[\int_0^{+\infty} z^{1/2} \times [4^4/\Gamma(4)]z^3 \exp(-4z)\,dz]^2 = [\int_0^{+\infty} (128/3)z^{7/2} \exp(-4z)\,dz]^2 = 0.940$. § The values listed represent estimates of the hemihedral twin fraction α assuming hemihedral twinning about the specified operator. These values do not translate directly to tetartohedral twin fractions.

the program *Phaser* (McCoy *et al.*, 2007) and data set 1. The search model was PDB entry 2ar1, which had 43% sequence identity (Arakaki *et al.*, 2006). Although potential molecular-replacement solutions were found in $P3_112$, $P3_121$ and $P6_4$, they were judged to be unreliable because of low log-likelihood gain (LLG) values (13, 49 and 5, respectively). A more reliable solution was found in $P3_1$, with an LLG value of 182. We tested out all four solutions in *CNS* (Brünger *et al.*, 1998) and after one cycle of hemihedral/tetartohedral twin-

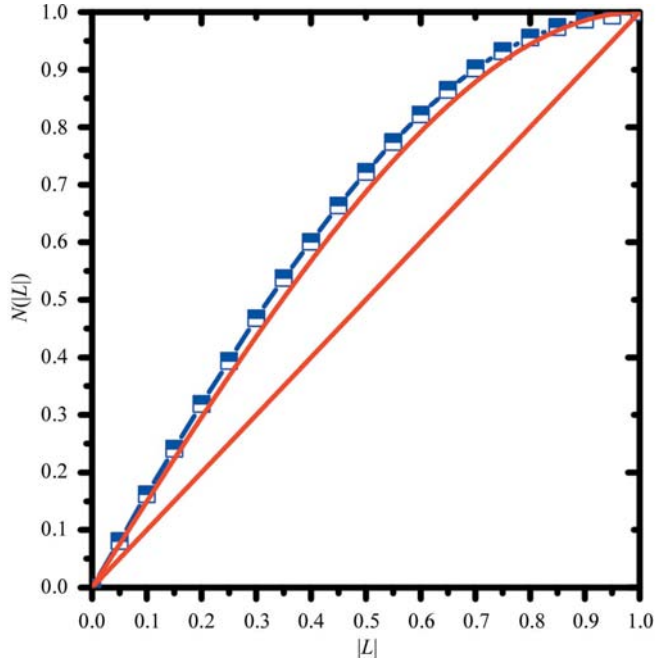

Figure 3
Analysis of Padilla–Yeates local intensity statistics (Padilla & Yeates, 2003). Theoretical distributions for untwinned acentric data are shown by the red straight line and those for perfectly hemihedrally twinned (acentric) data are shown by the red curve; the distribution for the DUF55 acentric data (data set 1) is shown by the blue curve.

Table 3
Comparison of partial refinements under the eight possible twin-fraction solutions for data set 1.

Refinements were carried out with detwinned data. All twin-related reflections were kept together in either the test set or the refinement set. The reported values are from preliminary refinements based on detwinning without the aid of calculated model structure factors (see text). Final refinement values are given in Table 5.

	$R_{\rm detwin}$/free $R_{\rm detwin}$ (%)		$R_{\rm twin}$/free $R_{\rm twin}$ (%)	
Twin fractions	Before refinement	After refinement	Before refinement	After refinement
Solution 1				
0.424, 0.300, 0.134, 0.142	**46.19/44.82**	**37.50/42.76**	**31.45/32.07**	**25.88/28.60**
0.300, 0.424, 0.142, 0.134	46.97/45.78	41.38/47.98	31.43/32.05	27.67/31.65
0.134, 0.142, 0.424, 0.300	50.84/51.12	46.54/52.89	32.75/33.14	30.80/32.28
0.142, 0.134, 0.300, 0.424	50.46/51.51	46.48/51.64	32.72/33.09	30.31/33.81
Solution 2				
0.076, 0.200, 0.366, 0.358	50.13/51.35	48.17/53.17	33.04/33.27	31.37/34.34
0.200, 0.076, 0.358, 0.366	49.15/48.75	43.78/47.68	32.77/33.13	29.62/31.45
0.366, 0.358, 0.076, 0.200	45.91/45.39	38.51/45.95	31.67/32.36	26.74/30.42
0.358, 0.366, 0.200, 0.076	46.08/44.74	39.11/45.13	31.53/32.20	26.65/30.28

ning refinement only the solution in $P3_1$ led to improved $R_{\rm work}$ and $R_{\rm free}$ values simultaneously (Table 1); the others led to a poorer $R_{\rm free}$ although $R_{\rm work}$ clearly decreased. The true space group was therefore assigned as $P3_1$, implying that the DUF55 crystals were tetartohedrally twinned. Similar results were obtained using data set 2 (data not shown).

2.5. Estimation of tetartohedral twin fractions

Overall intensity statistics were evaluated in order to test for high or perfect twinning (Yeates, 1997) and the results confirmed severe twinning (Table 2). In order to identify the twin operator and estimate the twin fraction, the data (reduced in $P3_1$) were evaluated using the *H*-test (Yeates, 1988) under the three possible twin operators in $P3_1$. The results (Table 2) suggested that data set 1 was partially twinned by tetartohedry and that data set 2 might be nearly perfectly twinned by tetartohedry. We therefore attempted to refine the structure against data set 2 using *CNS* and available scripts (Barends *et al.*, 2005). However, it was not possible to obtain a satisfactory refinement. One possible explanation was that data set 2 was also partially (but not perfectly) tetartohedrally twinned and that the correct twin fractions therefore needed to be incorporated into the refinement. To this end, a new method for estimating tetartohedral twin fractions (Yeates & Yu, 2008) was applied. Two unique and equally plausible solutions for the four twin-fraction values were obtained from this method; the correct solution had to be distinguished by further analysis. The potential solutions for the twin fractions for data sets 1 and 2 are shown in Table 2.

In the case of tetartohedral twinning, the observed intensities (J_k) are related to the true crystallographic intensities (I_k) as follows:

$$\begin{cases} \alpha_1 I_1 + \alpha_2 I_2 + \alpha_3 I_3 + \alpha_4 I_4 = J_1 \\ \alpha_1 I_2 + \alpha_2 I_1 + \alpha_3 I_4 + \alpha_4 I_3 = J_2 \\ \alpha_1 I_3 + \alpha_2 I_4 + \alpha_3 I_1 + \alpha_4 I_2 = J_3 \\ \alpha_1 I_4 + \alpha_2 I_3 + \alpha_3 I_2 + \alpha_4 I_1 = J_4 \end{cases} . \qquad (2)$$

The assignment of subscripts to twin operations is arbitrary, so we can say that I_1 refers to $I(h, k, l)$, I_2 refers to $I(k, h, -l)$, I_3 refers to $I(-k, -h, -l)$ and I_4 refers to $I(-h, -k, l)$, according to the three underlying twin operations. If the twin fractions (α_1, α_2, α_3, α_4) have been estimated, then the intensities can be detwinned (*i.e.* one can solve the equations above, given the observed intensities, to obtain the true intensities, arbitrarily setting $I_{\rm true} = I_1$). If the data have been detwinned successfully, then the resulting intensities should obey exponential statistics (Wilson, 1949). On the other hand, if an incorrect solution for the twin fractions has been chosen then the detwinned intensities may not follow the correct distribution. In this way, it is possible to distinguish between correct and incorrect solutions for the twin fractions (Yeates & Yu, 2008). This approach was used to analyze the potential twin-fraction solutions for data sets 1 and 2 (Fig. 4). In the case of data set 1, the detwinned data calculated under solution 1 followed a somewhat more ideal distribution than those calculated under solution 2. In the case of data set 2, solution 2 was much better.

It is important to note here that for each solution for the twin-fraction values there are four different permutations of the four twin fractions which are equally correct (Yeates & Yu, 2008); different permutations simply correspond to exchanging the assignments of I_1, I_2, I_3 and I_4 in (2). The permutation of the twin fractions that is chosen is therefore arbitrary, unless it is necessary to obtain agreement with a previously defined set of intensities. This was the case here, as a molecular-replacement model (and its calculated intensities) had already been obtained (*i.e.* prior to detwinning). In this work, therefore, detwinning under all four allowed permutations of the twin-fraction solution was performed and the correct solution was decided by the behaviour of atomic refinement and by the inspection of electron-density maps. For completeness, the four permutations of the alternate (less plausible) solution for the twin fractions were also tested. The same procedure was applied to data set 2. Thus, eight separate preliminary refinements were conducted using each of the two data sets (the results for data set 1 are shown in Table 3). Solution 1 for data set 1 was (0.424, 0.300, 0.134, 0.142), with the alternate orderings (0.300, 0.424, 0.142, 0.134), (0.134, 0.142, 0.424, 0.300) and (0.142, 0.134, 0.300, 0.424) being equally possible. It was verified that solution 1 of data set 1 (0.424, 0.300, 0.134, 0.142) and solution 2 of data set 2 (0.291, 0.276, 0.151, 0.282) were correct. This was consistent with the previous examination of detwinned intensity statistics (Fig. 4).

2.6. Refinement under partial tetartohedral twinning

Because all four twin-related reflections must be observed for detwinning to be carried out (2), data completeness is important for detwinning. We therefore refined the structure using data set 1. An atomic model was rebuilt using *Coot*

Table 4
Crystallographic data statistics for DUF55.

Values in parentheses are for the highest resolution shell.

	Data set 1	Data set 2
Space group	$P3_1$	$P3_1$
Unit-cell parameters (Å)	$a = b = 51.26$, $c = 122.40$	$a = b = 51.36$, $c = 122.75$
Resolution limits (Å)	44.41–2.30 (2.42–2.30)	44.50–2.30 (2.42–2.30)
$R_{\rm merge}$ (%)	5.6 (29.1)	5.4 (24.9)
Total no. of observations	77043 (10467)	122732 (15115)
Total no. of unique observations	15931 (2326)	15329 (2094)
$\langle I/\sigma(I)\rangle$	22.5 (5.4)	27.9 (6.7)
Completeness (%)	99.4 (98.8)	95.1 (88.8)
Multiplicity	4.8 (4.5)	8.0 (7.2)
Molecules per ASU	2	2

(Emsley & Cowtan, 2004) and simulated annealing and *B*-factor refinement were then performed employing non-crystallographic symmetry restraints using *CNS*. After about ten cycles of refinement, most of the amino-acid residues could be traced, including four residues which were introduced by plasmid construction. However, the $R_{\rm detwin}$ and free $R_{\rm detwin}$ values remained relatively high (29.84% and 34.92%, respectively). It was surmised that this was in part a consequence of the magnification of measurement errors caused by detwinning (*i.e.* by inverting the linear equations in equation 2). Consequently, after this point an alternate refinement strategy was employed. Instead of detwinning based on observed intensities alone, detwinning was performed using proportionality rules applied to the current values of the calculated model structure factors,

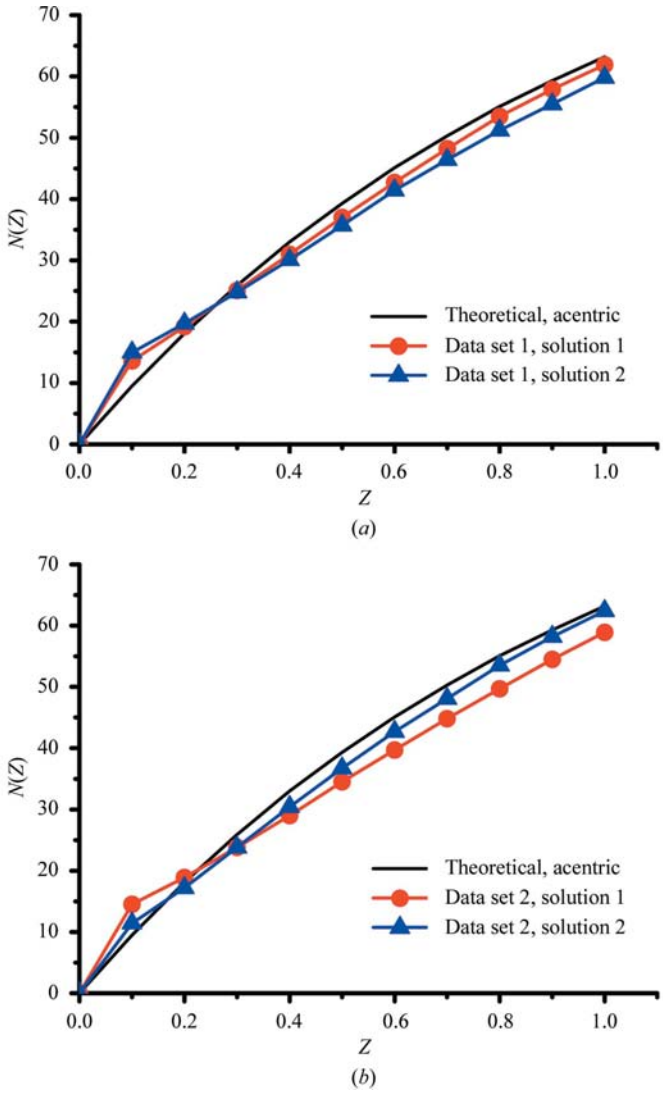

Figure 4
Comparison of cumulative intensity distributions (cumulative Wilson distributions) calculated under two possible solutions for the tetartohedral twin fractions. (*a*) Data set 1. The detwinned data calculated under solution 1 are better than those calculated under solution 2. (*b*) Data set 2. Solution 2 is much better.

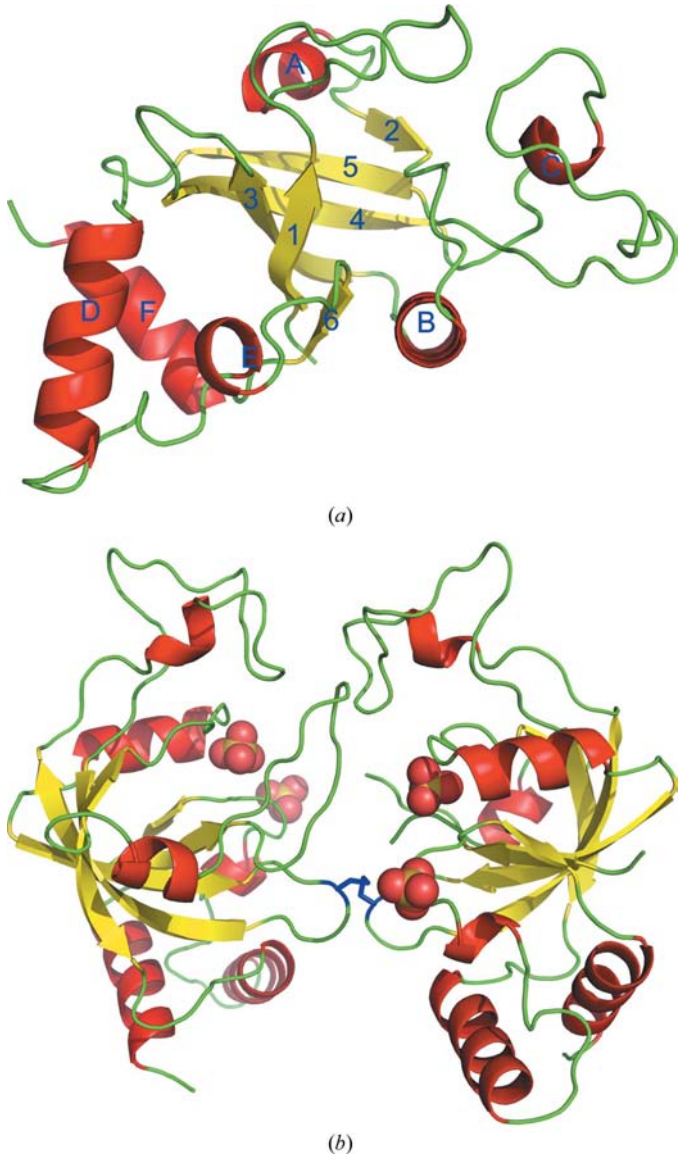

Figure 5
(*a*) A ribbon diagram of DUF55 showing secondary-structure elements with labels (the α-helices are lettered and the β-strands are numbered). (*b*) Dimer structure of DUF55. The intermolecular disulfide bond is coloured blue and four sulfate ions are also shown.

$$F_{\rm detwin} = \{[F_{\rm obs}^2 + 0.424k(F_{\rm calc} + F_{\rm bulk})^2 - 0.300kT_1(F_{\rm calc} + F_{\rm bulk})^2 - 0.134kT_2(F_{\rm calc} + F_{\rm bulk})^2 - 0.142kT_3(F_{\rm calc} + F_{\rm bulk})^2]/2\}^{1/2}, \quad (3)$$

where k is a scale factor, T_1, T_2 and T_3 indicate the twin operations that apply to the structure factors, $F_{\rm calc}$ is the structure-factor array for the macromolecular model including ordered solvent and $F_{\rm bulk}$ is the structure-factor array for an appropriate model of disordered solvent. In this way, detwinned (true) intensities can be obtained without the error magnification caused by inverting (2), although some model bias is likely to be introduced as a trade-off. After every round of refinement and model building, $F_{\rm detwin}$ values were recalculated in order to obtain more precise values. Following several cycles of recalculation of $F_{\rm detwin}$, refinement and model building, the final $R_{\rm detwin}$ and free $R_{\rm detwin}$ were 18.23% and 23.78%, respectively. After this structure had been determined, the final structure was refined against data set 2. The final $R_{\rm detwin}$ and free $R_{\rm detwin}$ of the second structure were 21.67% and 25.63%, respectively.

The model quality was analyzed using *PROCHECK* (Laskowski *et al.*, 1993). The figures were produced using *PyMOL*.

3. Results and discussion

3.1. Overall structure

The DUF55 crystal (data set 1) belonged to space group $P3_1$, with unit-cell parameters $a = b = 51.26$, $c = 122.40$ Å. Detailed crystallographic data statistics for DUF55 are shown in Table 4. There are 350 residues, including the His tag, in the asymmetric unit and 328 of these residues can be traced, excluding the first residues of both chains, region 64–71 of chain *B* and the His tag as no clear electron density was observed for these residues. The model is therefore virtually complete. The amino-acid sequence numbering of DUF55 was chosen to match that of full-length hTHYN1. Although 1,6-diaminohexane is important for crystallization, it does not appear in the electron-density maps. The refined structure contains 2719 non-H protein atoms, 37 water molecules and four sulfate ions. 82.9% of all residues are located in the most allowed regions of the Ramachandran diagram. The $R_{\rm detwin}$ and free $R_{\rm detwin}$ factors are 18.23% and 23.78%, respectively. The $R_{\rm twin}$ and free $R_{\rm twin}$ values (14.81% and 19.12%, respectively), which were calculated after refinement, are lower than $R_{\rm detwin}$ and free $R_{\rm detwin}$, as expected owing to the effects of statistical averaging of twinned intensities (Redinbo & Yeates, 1993). In order to avoid bias, quadruplets of twin-related reflections were kept together in either the test set or the refinement set throughout refinement. The detailed refinement statistics and quality of the model are shown in Table 5.

Table 5
Refinement statistics and quality of the model.

Values in parentheses are for the highest resolution shell.

$R_{\rm detwin}$	18.23 (25.13)
Free $R_{\rm detwin}$	23.78 (31.86)
$R_{\rm twin}$	14.78 (21.84)
Free $R_{\rm twin}$	19.10 (25.63)
Model quality	
No. of atoms	
Protein atoms	2719
Sulfate ions	4
Waters	37
Average *B* factors (Å^2)	
Mean *B* value	36.24
B value from Wilson plot	40.06
R.m.s. deviations from ideal values	
Bond lengths (Å)	0.0060
Bond angles (°)	1.1706
Residues in Ramachandran plot (%)	
Most allowed region	82.9
Allowed region	17.1
Generously allowed region	0.0
Disallowed region	0.0

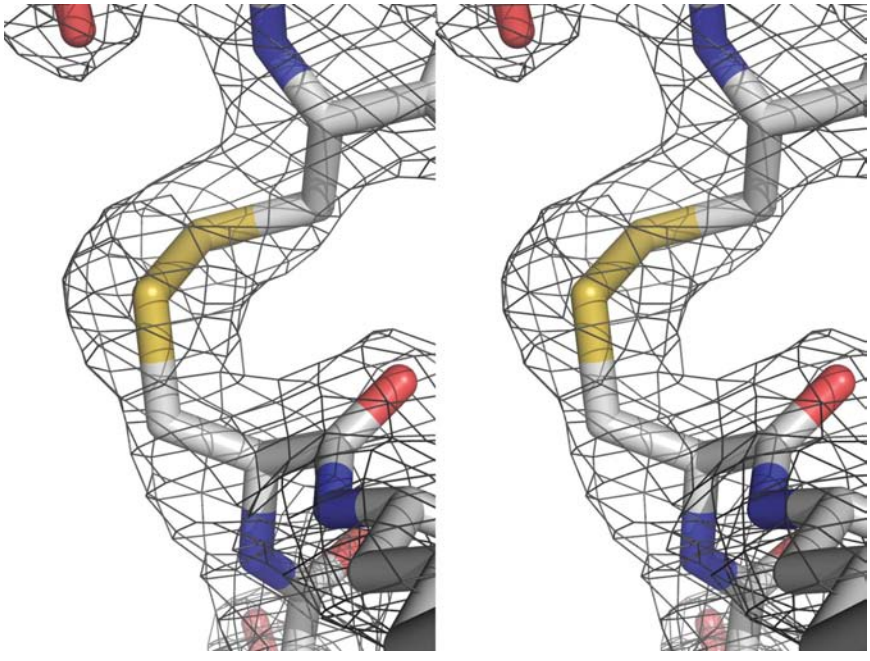

Figure 6
Stereoview of the intermolecular disulfide bond which is formed by Cys118 residues from both chains. The $2F_{\rm o} - F_{\rm c}$ map is set to 1σ (white, C; blue, N; red, O; yellow, S).

The DUF55 structure consists of six α-helices and six β-strands that were well defined (Fig. 5*a*). α-Helices *A* and *B* and β-strands 1–6 form a notable surface cleft. Five homologous structures are known, of which four were determined using X-ray diffraction methods (PDB codes 2ar1, 2eve, 2g2x and 1zce). The r.m.s. deviations between DUF55 and the four homologous structures are 1.35, 1.25, 1.30 and 1.42 Å, respectively. Three of the homologous structures mentioned above contain ligands in this

cleft, which are apparently derived from the crystallization conditions or cryoprotectant solutions. 2ar1 contains a glycerol molecule in this cleft, while a MOPS molecule and sulfate ions exist in the same location in 2eve and 2g2x, respectively. Sulfate ions are also present in this cleft in the present structure of DUF55, but the locations of the sulfate ions are different. In DUF55, two of them are close to the location of the sulfonic acid group of the MOPS molecule in 2eve, whereas the other two are close to Arg200 and Arg202 (Fig. 5*b*). Unexpectedly, a well resolved intermolecular disulfide bond was found in both structures (Fig. 5*b*) involving Cys118 of both chains (Fig. 6).

There are two molecules in the asymmetric unit, which form a dimeric state through an intermolecular disulfide bond, while gel-filtration experiments show that the protein exists as a monomer in solution (Fig. 7). No intermolecular disulfide bonds are observed in the other structures, although a homologous Cys118 also exists in 2eve and 2g2x. In three of the four structures only one monomer exists in the asymmetric unit (2ar1, 2eve and 1zce). 2g2x is the exception; its asymmetric unit contains three molecules without apparent noncrystallographic point-group symmetry and the interactions between different pairs of molecules are distinct. Thus, the dimeric structure observed in the crystalline state in the present work may not be biologically significant.

3.2. Structure similarity to the YTH domain in YTH-domain-containing protein 2

A structural similarity search was performed using *DALI* (Holm & Sander, 1996). Most of the hits have no functional annotations. Among those proteins that have function annotations, the highest *Z* score (8.4) is from the YTH domain in the human YTH domain-containing protein 2 (PDB code 2yu6). YTH is a potential RNA-binding domain (Stoilov *et al.*, 2002). After superimposition of DUF55 with the YTH domain using secondary-structure matching, most of the α-helices and β-strands could be superimposed; the r.m.s. deviation of Cα atoms is 3.4 Å. Sequence alignment shows that Lys60, Ser61 and Trp90 are highly conserved in the homologous DUF55 domains and the YTH domain. The spatial locations of these three residues are also conserved in 2yu6 and DUF55. It is thus implied that DUF55 might be a potential RNA-related domain. Other structural neighbours of known function identified here had previously been reported in a search for folds similar to 2ar1 (Arakaki *et al.*, 2006). The highest *Z* score among those is 4.7 for the N-terminal domain of *E. coli* Lon protease (PDB code 2ane).

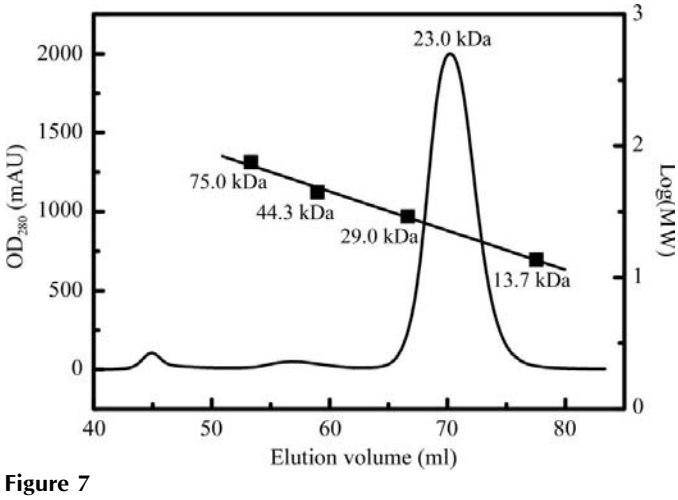

Figure 7
Analytical gel-filtration profile for DUF55. The blocks are the standard markers and the line is a linear fitting curve. The theoretical molecular weight of DUF55 with the His tag is 20.5 kDa; DUF55 therefore exists in a monomeric state in solution. The globular marker proteins are RNase A (13.7 kDa), carbonic anhydrase (29.0 kDa), ovalbumin (44.3 kDa) and conalbumin (75.0 kDa).

3.3. Crystal packing and twinning

Analyzing the DUF55 crystal packing, it is seen that the NCS twofold is nearly parallel to [100]. The NCS rotation matrix calculated using *CNS* was

$$\begin{pmatrix} 0.99998 & 0.00599 & -0.00221 \\ 0.00608 & -0.99915 & 0.04986 \\ -0.00197 & -0.04087 & -0.99916 \end{pmatrix} \simeq \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}. \quad (4)$$

This has been referred to as rotational pseudosymmetry (RPS; Zwart *et al.*, 2008). The presence of RPS apparently promotes the growth of twinned crystals owing to layering and low steric hindrance of molecules at the twin-domain interface. However, it is not a requirement for twin formation.

To our knowledge, four tetartohedrally twinned protein crystal structures have been determined prior to this work (PDB codes 1qzw, 2pi8, 2h6r and 2pk2). Three of them contained 222 rotational pseudosymmetry (1qzw, 2pi8 and 2h6r) occurring with noncrystallographic symmetry (NCS; Rosendal *et al.*, 2004; Barends *et al.*, 2005; Gayathri *et al.*, 2007). 2pk2 also has three mutually perpendicular NCS twofold symmetry axes which are nearly parallel to three pseudomerohedral twin operators (Anand *et al.*, 2007). Because of their fortunate orientations, layering would also occur, resulting in low steric hindrance at the twin-domain interface. Hence, the twofold NCS symmetry of 2pk2 has a similar function to 222 rotational pseudosymmetry. However, in the crystal of DUF55 there is no additional NCS twofold axis that would be required to generate local 222 symmetry. Layering does not seem to occur along the [001] and [120] directions, so a layering effect does not appear to promote twinning in the present case. It is easier to interpret the causes of twinning in the other four known tetartohedral twinning cases owing to their local 222 symmetry. DUF55 provides an example with only one axis of rotational pseudosymmetry. This suggests that a tetartohedrally twinned case without any NCS rotational pseudosymmetry might be found in the future.

We would like to thank Professor Zongxiang Xia, Professor Bauke W. Dijkstra and Karin van Straaten for help in the

structure refinement. Some of the diffraction data used in this study were collected at the University of Science and Technology of China; we are grateful to Professor Maikun Teng and Xiao Zhang. This work was supported by the Shanghai Natural Science Foundation (Grant No. 07JC14062) and the National Natural Science Foundation (Grant No. 10774155).

References

Anand, K., Schulte, A., Fujinaga, K., Scheffzek, K. & Geyer, M. (2007). *J. Mol. Biol.* **370**, 826–836.
Arakaki, T., Le Trong, I., Phizicky, E., Quartley, E., DeTitta, G., Luft, J., Lauricella, A., Anderson, L., Kalyuzhniy, O., Worthey, E., Myler, P. J., Kim, D., Baker, D., Hol, W. G. J. & Merritt, E. A. (2006). *Acta Cryst.* F**62**, 175–179.
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.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
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.
Compton, M. M., Thomson, J. M. & Icard, A. H. (2001). *Apoptosis*, **6**, 299–314.
Dauter, Z. (2003). *Acta Cryst.* D**59**, 2004–2016.
Dauter, Z., Botos, I., LaRonde-LeBlanc, N. & Wlodawer, A. (2005). *Acta Cryst.* D**61**, 967–975.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Finn, R. D., Mistry, J., Schuster-Bockler, B., Griffiths-Jones, S., Hollich, V., Lassmann, T., Moxon, S., Marshall, M., Khanna, A., Durbin, R., Eddy, S. R., Sonnhammer, E. L. L. & Bateman, A. (2006). *Nucleic Acids Res.* **34**, D247–D251.
Gayathri, P., Banerjee, M., Vijayalakshmi, A., Azeez, S., Balaram, H., Balaram, P. & Murthy, M. R. N. (2007). *Acta Cryst.* D**63**, 206–220.
Holm, L. & Sander, C. (1996). *Science*, **273**, 595–602.
Jiang, X. Z., Toyota, H., Takada, E., Yoshimoto, T., Kitamura, T., Yamada, J. & Mizuguchi, J. (2003). *Tissue Cell*, **35**, 471–478.
Jiang, X. Z., Toyota, H., Yoshimoto, T., Takada, E., Asakura, H. & Mizuguchi, J. (2003). *Apoptosis*, **8**, 509–519.
Larsen, N. A. & Harrison, S. C. (2004). *J. Mol. Biol.* **344**, 885–892.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Lebedev, A. A., Vagin, A. A. & Murshudov, G. N. (2006). *Acta Cryst.* D**62**, 83–95.
Leslie, A. G. W. (2006). *Acta Cryst.* D**62**, 48–57.
MacRae, I. J. & Doudna, J. A. (2007). *Acta Cryst.* D**63**, 993–999.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
Miyaji, H., Yoshimoto, T., Asakura, H., Komachi, A., Kamiya, S., Takasaki, M. & Mizuguchi, J. (2002). *Gene*, **297**, 189–196.
Padilla, J. E. & Yeates, T. O. (2003). *Acta Cryst.* D**59**, 1124–1130.
Parsons, S. (2003). *Acta Cryst.* D**59**, 1995–2003.
Redinbo, M. R. & Yeates, T. O. (1993). *Acta Cryst.* D**49**, 375–380.
Rees, D. C. (1980). *Acta Cryst.* A**36**, 578–581.
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.
Song, A. X., Chang, Y. G., Gao, Y. G., Lin, X. J., Shi, Y. H., Lin, D. H., Hang, Q. H. & Hu, H. Y. (2005). *Protein Expr. Purif.* **42**, 146–152.
Stanley, E. (1972). *J. Appl. Cryst.* **5**, 191–194.
Stoilov, P., Rafalska, I. & Stamm, S. (2002). *Trends Biochem. Sci.* **27**, 495–497.
Sultana, A., Alexeev, I., Kursula, I., Mäntsälä, P., Niemi, J. & Schneider, G. (2007). *Acta Cryst.* D**63**, 149–159.
Toms, A. V., Kinsland, C., McCloskey, D. E., Pegg, A. E. & Ealick, S. E. (2004). *J. Biol. Chem.* **279**, 33837–33846.
Wilson, A. J. C. (1949). *Acta Cryst.* **2**, 318–321.
Yang, F., Dauter, Z. & Wlodawer, A. (2000). *Acta Cryst.* D**56**, 959–964.
Yeates, T. O. (1988). *Acta Cryst.* A**44**, 142–144.
Yeates, T. O. (1997). *Methods Enzymol.* **276**, 344–358.
Yeates, T. O. & Rees, D. C. (1987). *Acta Cryst.* A**43**, 30–36.
Yeates, T. O. & Yu, F. (2008). *Acta Cryst.* D**64**, 1158–1164.
Zhang, Q. H. *et al.* (2000). *Genome Res.* **10**, 1546–1560.
Zwart, P. H., Grosse-Kunstleve, R. W., Lebedev, A. A., Murshudov, G. N. & Adams, P. D. (2008). *Acta Cryst.* D**64**, 99–107.