Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Use of multiple anomalous dispersion to phase highly merohedrally twinned crystals of interleukin-1β

Markus G. Rudolph,[a]† Matthew S. Kelker,[a] Thomas R. Schneider,[b] Todd O. Yeates,[c] Vanessa Oseroff,[d] David K. Heidary,[d] Patricia A. Jennings[d] and Ian A. Wilson[a]*

[a]Department of Molecular Biology and The Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, CA 92037, USA, [b]Department of Structural Chemistry, University of Göttingen, 37077 Göttingen, Germany, [c]Department of Chemistry and Biochemistry, University of California, Los Angeles, CA 90095, USA, and [d]Department of Chemistry, University of California, San Diego, La Jolla, CA 92307, USA

† Present address: Department of Molecular Structural Biology, Institute for Microbiology and Genetics, Georg-August University, 37077 Göttingen, Germany.

Correspondence e-mail: wilson@scripps.edu

Received 23 October 2002
Accepted 22 November 2002

PDB Reference: F42W/W120F interleukin-1β, 1l2h, r1l2hsf.

The crystal structure at 1.54 Å resolution of a double mutant of interleukin-1β (F42W/W120F), a cytokine secreted by macrophages, was determined by multiple-wavelength anomalous dispersion (MAD) using data from highly twinned selenomethionine-modified crystals. The space group is $P4_3$, with unit-cell parameters $a = b = 53.9$, $c = 77.4$ Å. Self-rotation function analysis and various intensity statistics revealed the presence of merohedral twinning in crystals of both the native (twinning fraction $\alpha \simeq 0.35$) and SeMet ($\alpha \simeq 0.40$) forms. Structure determination and refinement are discussed with emphasis on the possible reasons for successful phasing using untreated twinned MAD data.

1. Introduction

Twinning is a crystal-growth peculiarity, common in small-molecule crystals but apparently less frequent in macromolecular crystals, where two or more crystals of different relative orientations intergrow to form a larger aggregate (Catti & Ferraris, 1976; Donnay & Donnay, 1974; Friedel, 1926). The lattices of the individual crystals may overlap in two or three dimensions, leading to epitaxic or merohedral twinning, respectively. While epitaxic twinning gives rise to separate diffraction patterns, one for each crystal, and is therefore easily detectable and may be accounted for (Terwisscha van Scheltinga *et al.*, 2001; Yeates & Fam, 1999), merohedral twinning produces a single diffraction pattern and its diagnosis is less obvious (Yeates, 1997; Yeates & Fam, 1999). Merohedrally twinned crystals of chiral molecules may be composed of two (hemihedral) or four (tetratohedral) different twin domains (Chandra *et al.*, 1999; Yeates, 1997; Yeates & Fam, 1999) and their diffraction pattern is therefore a superposition of the diffraction pattern of two or four separate twin domains (Liang *et al.*, 1996; Lietzke *et al.*, 1996). Hemihedrally twinned crystals are described by the symmetry operation necessary to superimpose the two individual twin domains (the twinning operator) and the fractional volume of the less populated twin domain in the crystal (the twin fraction α). As a consequence of the twinning operator, additional symmetry is introduced that is not part of the Laue symmetry group of the crystal. The twinning operator relates reflections that are not symmetry equivalent in the given space group, but are convoluted in the diffraction pattern of the twinned crystal such that the observed spot intensities are equal to the weighted average of these twin-related intensities,

$$I_{\rm obs}(\mathbf{h}_1) = (1-\alpha)I(\mathbf{h}_1) + \alpha I(\mathbf{h}_2), \quad (1)$$

$$I_{\rm obs}(\mathbf{h}_2) = \alpha I(\mathbf{h}_2) + (1-\alpha)I(\mathbf{h}_2). \quad (2)$$

$I_{obs}(\mathbf{h}_i)$ and $I(\mathbf{h}_i)$ are the measured and true crystallographic intensities, respectively, of the twin-related reflections $\mathbf{h}_i$. This linear system of two independent equations in two unknowns can be solved for the intensities $I(\mathbf{h}_i)$,

$$I(\mathbf{h}_1) = [(1-\alpha)I_{obs}(\mathbf{h}_1) - \alpha I_{obs}(\mathbf{h}_2)]/(1-2\alpha), \quad (3)$$

$$I(\mathbf{h}_2) = [(1-\alpha)I_{obs}(\mathbf{h}_2) - \alpha I_{obs}(\mathbf{h}_1)]/(1-2\alpha). \quad (4)$$

The errors of the true intensities, $\sigma(\mathbf{h}_i)$, can be derived from the measured intensity errors $\sigma_{obs}(\mathbf{h}_i)$ through Gaussian error propagation,

$$\sigma_{obs}(\mathbf{h}_1) = \{[(1-\alpha)\sigma_{obs}(\mathbf{h}_1)]^2 + [\alpha\sigma_{obs}(\mathbf{h}_2)]^2\}^{1/2}/(1-2\alpha), \quad (5)$$

$$\sigma_{obs}(\mathbf{h}_2) = \{[(1-\alpha)\sigma_{obs}(\mathbf{h}_2)]^2 + [\alpha\sigma_{obs}(\mathbf{h}_1)]^2\}^{1/2}/(1-2\alpha). \quad (6)$$

If $\sigma_{obs}(\mathbf{h}_1)$ and $\sigma_{obs}(\mathbf{h}_2)$ are assumed to be equal, the error for the detwinned reflections increases by a factor $1/(1-2\alpha)$. The expressions for the intensities and the errors (equations 1–6) become indeterminate for $\alpha = 0.5$. Since the anomalous signal amounts to only a few percent of the reflection intensities (Hendrickson & Ogata, 1997), detwinning of anomalous data, by producing an increase in the intensity errors, will deteriorate the anomalous signal to an extent where the data may not be useable for structure determination by MAD. This effect is more pronounced at higher twin fractions. Twinned data have been used previously for structure determination by molecular replacement (Breyer *et al.*, 1999; Contreras-Martel *et al.*, 2001; Luecke *et al.*, 1998; Rabijns *et al.*, 2001; Redinbo & Yeates, 1993; Taylor *et al.*, 2000; Trame & McKay, 2001; Larsen *et al.*, 2002) and single or multiple isomorphous replacement (Ban *et al.*, 1999; Declercq & Evrard, 2001; Mueller, Muller *et al.*, 1999; Mueller, Schubel *et al.*, 1999; Terwisscha van Scheltinga *et al.*, 2001). Only recently has a successful structure determination been described that employs twinned SeMet MAD data for phasing (Yang *et al.*, 2000).

Interleukin-1β (IL-1β) is a cytokine secreted by macrophages in response to pathogens and is involved in numerous physiological processes, such as growth regulation and inflammation (Dinarello, 1996). IL-1β does not contain a classic leader sequence for export, but still has to cross the cell membrane in order to bind its extracellular receptor. It has been suggested that IL-1β is exported in a partially unfolded form, which initiated interest in the IL-1β folding pathway (Epps *et al.*, 1997; Yem *et al.*, 1992). The crystal structures of IL-1β and various mutants have been determined previously and refined to resolutions of 2.0–2.8 Å (Finzel *et al.*, 1989; Priestle *et al.*, 1989; Shaanan *et al.*, 1992; Treharne *et al.*, 1990; van Oostrum *et al.*, 1991; Veerapandian *et al.*, 1992; Yu *et al.*, 1999), revealing a solvent-exposed hydrophobic cavity of $\sim$85 Å^3 which hosts water molecules that bridge β-strands in the hydrophobic core of IL-1β. The essential role of these water molecules in IL-1β stability was recently established by the use of the fluorescence of the lone Trp120 residue (Covalt *et al.*, 2001). In addition, the cavity has been speculated to host mobile solvent molecules (Ernst *et al.*, 1995; Yu *et al.*, 1999) which may be important for the folding of IL-1β. In addition, residue Phe42 in the hydrophobic core plays a key role in early folding events of IL-1β. This residue is located close to β-strand 1, which is assumed to be unfolded during membrane-crossing of IL-1β (Yem *et al.*, 1992). In order to study the effect of the F42W substitution while at the same time preserving a single fluorescent probe, the IL-1β double mutant F42W/W120F was generated. Interestingly, the mutant was shown to fold faster than wild-type IL-1β (DKH and PAJ, unpublished results), indicating subtle structural changes in the hydrophobic core of this mutant compared with the wild type.

The crystal structure of the F42W/W120F IL-1β mutant was determined in order to ensure its three-dimensional integrity and to reveal any structural changes in both the protein and in the hydration of the central cavity relative to wild-type IL-1β. In order to exclude potential model bias by a molecular-replacement approach and to faithfully place water molecules, we redetermined the IL-1β structure to high resolution (1.54 Å) using SeMet MAD. Unfortunately, the crystals of the mutant exhibited a high degree of hemihedral twinning, but this did not hamper structure determination and refinement. In this study, we outline the phasing of the IL-1β diffraction data from hemihedrally twinned crystals using MAD, which represents the second example of a MAD structure determination using twinned data.

2. Materials and methods

2.1. Protein production and purification

The IL-1β F42W/W120F mutant and the SeMet-modified protein were generated, produced and purified using protocols similar to those described previously (Finke *et al.*, 2000; Heidary *et al.*, 1997). Buffers used for the purification of SeMet-modified proteins were supplemented with 5 m*M* dithiothreitol (DTT) and 3 m*M* 2-mercaptoethanol in order to prevent oxidation of the SeMet residues.

2.2. Crystallization and data collection

The F42W/W120F double mutant of IL-1β (11.6 mg ml^{-1}) was crystallized in hanging drops from 50 m*M* PIPES–NaOH pH 7.0, 2.0–2.2 *M* $(NH_4)_2SO_4$, 5 m*M* DTT. Crystals were cryoprotected in 25% glycerol, 1.5 *M* $(NH_4)_2SO_4$ and vitrified prior to data collection. An anomalous data set was collected at three wavelengths from a single crystal (Table 1) to 1.7 Å resolution at 93 K on beamline 9-2 at the Stanford Synchrotron Radiation Laboratory (SSRL). A 1.54 Å resolution native data set was collected on beamline 19-ID at the Advanced Photon Source (APS). Data reduction was performed using the *HKL* package (Otwinowski & Minor, 1997). The space group is $P4_3$, with unit-cell parameters $a = b = 53.9$, $c = 77.4$ Å. The Matthews coefficient (Matthews, 1968), $V_M = 2.9$ Å^3 Da^{-1}, suggested one molecule in the asymmetric unit, with a solvent content of 58%.

2.3. Structure determination and refinement

Self-rotation function analysis, cumulative intensity distributions, analysis of second moments and reduced $|E^2 - 1|$

Table 1
Crystallographic data collection, analysis and structure refinement.

Data collection†	SeMet			Native
Wavelength (Å)	0.9791 (f'')	0.9793 (f')	0.9184 (remote)	1.0332
Resolution range‡ (Å)	18.6–1.69 (1.73–1.69)	18.6–1.69 (1.73–1.69)	19.3–1.69 (1.73–1.69)	20.0–1.54 (1.60–1.54)
Unique reflections‡	52050 (3426)	52034 (3420)	52023 (3415)	31800 (3196)
Completeness‡ (%)	99.7 (99.4)	99.7 (99.4)	99.7 (99.4)	97.1 (98.2)
R_{sym}‡§¶, $P4$ (%)	[4.4] 5.3 (60.4)	[3.9] 4.7 (51.5)	[3.6] 4.6 (57.3)	[5.6] 7.4 (51.6)
R_{sym}‡§¶, $P422$ (%)	[7.4] 8.5 (65.3)	[6.7] 7.7 (55.9)	[6.5] 7.7 (61.9)	[9.4] 12.6 (57.2)
Average $I/\sigma(I)$‡	31.3 (2.0)	33.7 (2.4)	32.7 (2.1)	23.4 (2.6)
Redundancy‡	4.2 (4.2)	4.2 (4.3)	4.2 (4.2)	7.6 (5.2)
Mosaicity (°)	0.34	0.34	0.35	0.33
Mean $\lvert E^2 - 1\rvert$	0.575	0.575	0.583	0.584 (0.736)††
$\langle\lvert I\rvert^2\rangle/\langle\lvert I\rvert\rangle^2$	1.64	1.64	1.66	1.68 (2.0/1.5)‡‡
$\langle\lvert F\rvert^2\rangle/\langle\lvert F\rvert^2\rangle$	0.86	0.86	0.86	0.86 (0.785/0.865)‡‡
$\langle H\rangle$	0.078 (α = 0.42)	0.081 (α = 0.41)	0.077 (α = 0.42)	0.151 (α = 0.35)
$\langle H^2\rangle$	0.009 (α = 0.42)	0.010 (α = 0.42)	0.009 (α = 0.42)	0.033 (α = 0.34)
Refined f'/f''§§	−5.181/7.388	−6.078/3.609	−1.976/4.402	—
FOM after *SOLVE*¶¶	0.51/0.40			
FOM after *RESOLVE*¶¶	0.53/0.50			
$\langle\Delta\varphi\rangle$ (°)†††	60.8 and 47.1/59.5 and 44.2			
Refinement				
Resolution range (Å)				18–1.54
No. of reflections/test set				30261/1519
R_{cryst}/R_{free}‡‡‡ (%)				15.4/19.3
No. of residues/protein atoms				144/1168
No. of water O atoms				126
Coordinate error§§§ (Å)				0.2
Refined twin fraction				0.37
R.m.s. deviations from ideality				
Bonds (Å)				0.012
Angles (°)				2.6
Dihedrals (°)				28.3
Ramachandran plot¶¶¶				
Most favoured (%)				91.3
Additional allowed (%)				8.7
Average B values (Å^2)				
Protein				26.1 ± 11.0
Waters				34.3 ± 8.8

† All data-collection statistics were calculated with Friedel mates treated separately. ‡ Values in parentheses refer to the highest resolution shell. § $R_{sym} = 100\sum_h \sum_i |I_i(h) - \langle I(h)\rangle|/\sum_h I(h)$, where $I_i(h)$ is the ith measurement of reflection h and $\langle I(h)\rangle$ is the average value of the reflection intensity. ¶ Values in square brackets refer to the lowest resolution shell (20–4.07 Å for the SeMet data and 20–3.31 Å for the native data). †† The value in parentheses is the expected value for untwinned data. ‡‡ Values in parentheses are the expected values for untwinned/perfectly twinned data. §§ Refined using *XPREP*. ¶¶ Figure of merit for all reflections for the twinned/detwinned data. ††† Mean phase differences (resolution range 18–1.9 Å) between phases from *SOLVE* and *RESOLVE* derived from the twinned/detwinned data and phases derived from the final model without taking twinning into account. ‡‡‡ $R_{cryst} = \sum |F_o| - |F_c|/\sum |F_o|$, where F_o and F_c are the structure-factor amplitudes from the data and the model, respectively. R_{free} is R_{cryst} with 5% of test-set structure factors. §§§ Based on maximum likelihood. ¶¶¶ Calculated using *PROCHECK* (Laskowski *et al.*, 1993).

values indicated hemihedral twinning for both the native and SeMet data sets in space group $P4_3$ (Fig. 1, Table 1). The twin fractions of the native and MAD data sets were estimated to 0.36 and 0.40, respectively, by both partial and perfect twinning tests using *CNS* (Brünger *et al.*, 1998) and the twinning server (Yeates, 1997) after merging of Friedel mates. Despite the presence of hemihedral twinning, significant anomalous signal was present to 1.8 Å, as judged by the anomalous signal-to-noise ratio being >1.5 using *XPREP* (Bruker AXS). The substructure, consisting of six Se positions, was determined with *SOLVE* (Terwilliger & Berendzen, 1996) using the untreated twinned MAD data to a maximum resolution of 1.8 Å. The initial phases from *SOLVE* were improved by density modification using *RESOLVE* (Terwilliger, 2000), which resulted in overall figures of merit (FOMs) of 0.51 and 0.53, respectively. While the FOMs did not increase significantly during density modification, the electron-density maps were of significantly higher quality (Fig. 2). These 1.8 Å resolution maps and the twinned native data set to 1.54 Å resolution were used for initial automated model building (112 peptides, connectivity index 0.87) and refinement with *ARP*/*wARP* (Lamzin & Wilson, 1993). The resulting electron-density map was of excellent quality and facilitated manual building of the remaining residues and the side chains of residues 3–48, 55–104, 109–135 and 141–150 using *O* (Jones *et al.*, 1991).

Refinement to 1.54 Å resolution was continued in *CNS* against intensities using the protocols for twinned data, a fixed twin fraction of α = 0.35 and the maximum-likelihood target. The model was further rebuilt into σ_A-weighted $3F_o - 2F_c$ electron-density maps. Water molecules were assigned automatically in *CNS* at >3σ $F_o - F_c$ difference density peaks and were verified by manual inspection in *O*. After convergence in *CNS* (R_{cryst} = 19.7%, R_{free} = 24.3%), refinement of both the coordinates and the twin fraction was continued with *SHELXL* (Sheldrick & Schneider, 1997) and riding H atoms were included. Care was taken to keep the test set intact during all refinement steps. The final model (R_{cryst} = 15.4%, R_{free} = 19.3%; Table 1) is composed of all residues except residues 1–3, 54, 136–139 and 153–155, which showed no

interpretable electron density, and includes 126 water molecules.

3. Results and discussion

3.1. Detection of twinning in crystals of IL-1β

No split spots, which would indicate epitaxic twinning of IL-1β crystals, were observed in the diffraction pattern (data not shown). The first indication of possible hemihedral twinning was the observation that the data could be indexed, processed and scaled with good overall R_{sym} values in the $P4$ space group, but only reasonably well (especially in the low-resolution shell) in the $P422$ space group. Typical warning signs for twinning have been compiled (Herbst-Irmer & Sheldrick, 1998) and include apparent higher symmetry arising from the twin operator introducing additional symmetry that is not part of the Laue symmetry of the space group. Resolution-dependent inspection of R_{sym} revealed that while similar values were found in the high-resolution shells in $P4$ and $P422$, R_{sym} values in the lowest resolution shells were larger by almost a factor of two for the data scaled in $P422$ (Table 1). Calculation of the packing density assuming one molecule in the asymmetric unit yielded an unreasonably small Matthews parameter (Matthews, 1968) of $V_M = 1.45$ Å^3 Da^{-1} for the $P422$ case. IL-1β does not contain any possible perfect twofold symmetry (Finzel *et al.*, 1989; Priestle *et al.*, 1989), which could lead to an asymmetric unit harbouring only half a molecule with a concomitant V_M value of 2.9 Å^3 Da^{-1}. Thus, the scaling statistics and packing-density considerations already indicated the likely presence of hemi-

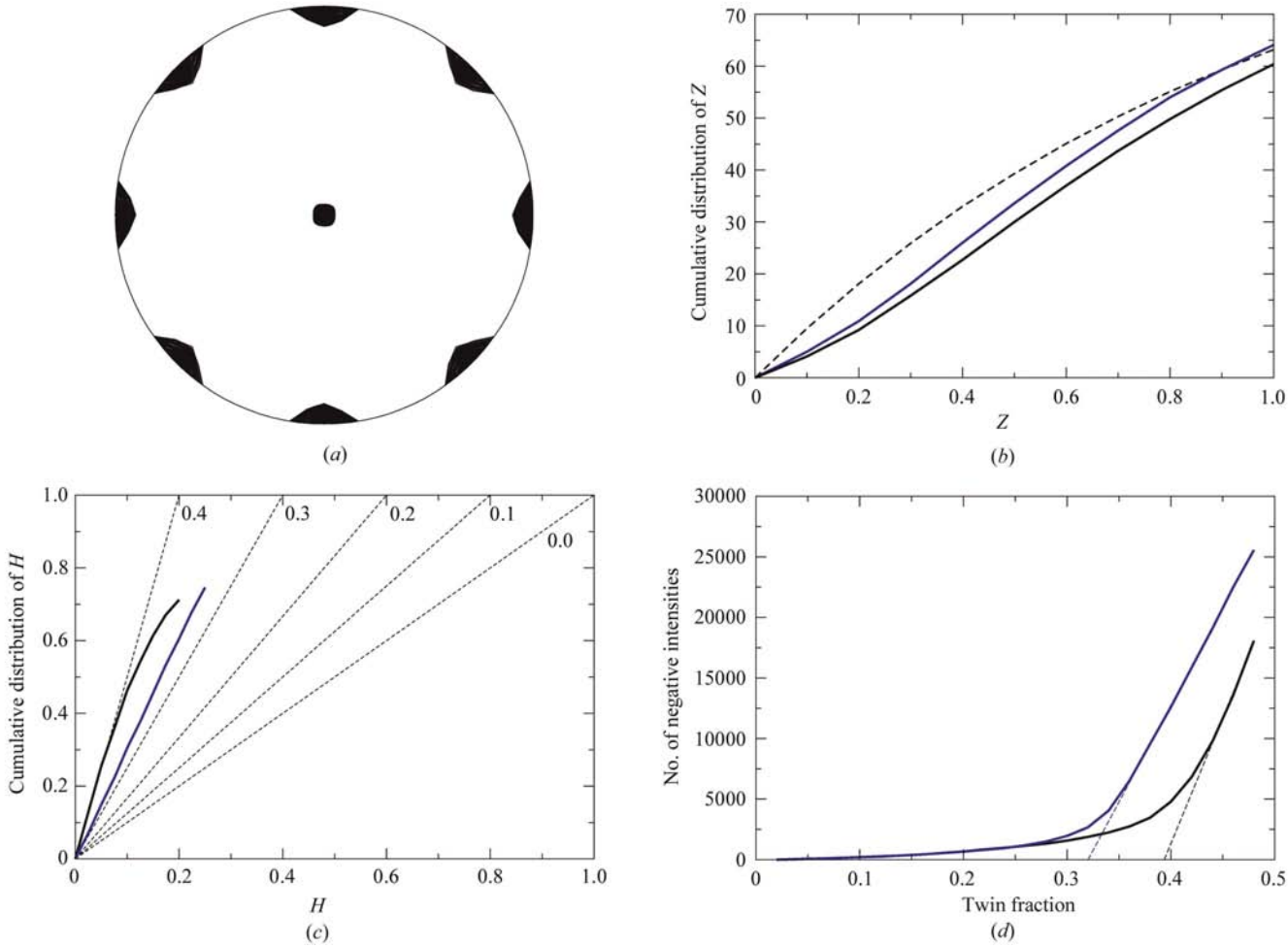

Figure 1
Detection and analysis of twinning. (*a*) Stereographic projection plot of the $\kappa = 180^\circ$ section of the self-rotation function of the native data set. The function was calculated with resolution limits of 15 and 3 Å and a Patterson integration radius of 15 Å and was contoured at <60% of the maximum peak height using *POLARRFN* (Collaborative Computational Project, Number 4, 1994). The data were reduced in $P4$ but the plot shows 422 symmetry, with the fourfold axis ($\omega = 0$, at the centre) and the twofold axes ($\omega = 90^\circ$, at the perimeter) having equal heights. After detwinning of the data, the peaks for the twofold axes are absent (not shown). (*b*) Cumulative distribution of $Z = I/\langle I \rangle$, where I is the intensity, for the acentric native (black) and anomalous (blue) data. The theoretical distribution for untwinned data is shown as a dotted line. The sigmoidal shape of the distribution indicates potential twinning. (*c*) Estimation of the twin fraction α by plotting the cumulative fractional intensity difference of acentric twin-related intensities, $H = |I_1 - I_2|/(I_1 + I_2)$, as a function of H (Yeates, 1988). The initial slope of the distribution is a measure of α. The dotted lines represent the expected slopes for the indicated twin fractions. (*d*) Estimation of α by Britton plots (Britton, 1972; Fisher & Sweet, 1980). The number of negative intensities after detwinning is plotted as a function of the assumed value of α. An overestimation of α will increase the number of negative intensities and the actual value of α is extrapolated from this increase (dotted lines).

hedral twinning in the crystals of the IL-1β F42W/W120F mutant.

Further analysis of the diffraction data using *XPREP* revealed lower mean $|E^2 - 1|$ values than expected for non-centrosymmetric space groups for the native and MAD data sets (Table 1), which is another hallmark of twinning (Herbst-Irmer & Sheldrick, 1998). The presence of pseudo-422 symmetry is apparent from a self-rotation function analysis of the data that were reduced in space group $P4$ (Fig. 1*a*). The $\kappa = 180^\circ$ section shows twofold axes perpendicular to the crystallographic fourfold axis. Interestingly, the peak heights of the twofold and fourfold axes were the same and did not correlate (as might have been expected) with the twin fraction, but were more dependent on the resolution limits and Patterson radius used for the self-rotation function calculation. A further indication for the presence of hemihedral twinning is an abnormal distribution of $Z = I/\langle I\rangle$. The cumulative intensity distributions of the IL-1β data, as calculated by *TRUNCATE* (Collaborative Computational Project, Number 4, 1994), are sigmoidal in shape compared with the theoretical distribution of untwinned data (Fig. 1*b*), further pointing to twinned crystals.

Table 2
Fractional coordinates, apparent occupancies, B values and height of the Se—Se Harker vectors.

Se site	x	y	z	Occupancy	B value (Å^2)	Peak height (σ)
1	0.863	0.215	0.163	0.48	29.4	14.9
2	0.701	0.064	0.228	0.55	36.5	14.1
3	0.576	0.319	0.120	0.45	35.4	11.0
4	0.901	0.282	0.153	0.50	43.3	12.5
5	0.338	0.419	0.157	0.43	39.3	11.4
6	0.781	0.543	0.094	0.41	46.2	9.0

To assess the twin fraction of the IL-1β native and MAD data, different twinning tests were employed (Kahlenberg, 1999) using the programs *CNS* (Brünger *et al.*, 1998), *DETWIN* (Collaborative Computational Project, Number 4, 1994) and the twinning server (Yeates, 1997). For acentric untwinned data, the expected ratio of the average squared intensity to the square of the average intensities ($\langle I^2\rangle/\langle I\rangle^2$) is 2; for perfectly twinned data, it is 1.5. The native data set value was 1.68, while the values for the MAD data were ~1.64 for all three wavelengths, thus eliminating perfect twinning but confirming the presence of partial twinning with a high twin fraction. A test for partial twinning (Yeates, 1988) using the relations $\langle H\rangle = 0.5 - \alpha$ and $\langle H^2\rangle = (1 - 2\alpha)^2/3$, where $H = |I_{\rm obs}(\mathbf{h}_1) - I_{\rm obs}(\mathbf{h}_2)|/[I_{\rm obs}(\mathbf{h}_1) + I_{\rm obs}(\mathbf{h}_2)]$, yielded twin-fraction estimates of 0.35 and 0.42 for the native and MAD data, respectively (Table 1). These values are slightly larger than the twin fractions of 0.32 and 0.40 estimated from both the cumulative distribution of H (Yeates, 1988) and the Britton plot (Fisher & Sweet, 1980; Figs. 1*c* and 1*d*). For the sake of clarity, twin fractions of 0.35 and 0.40 for the native and MAD data, respectively, were used in all subsequent calculations unless otherwise noted.

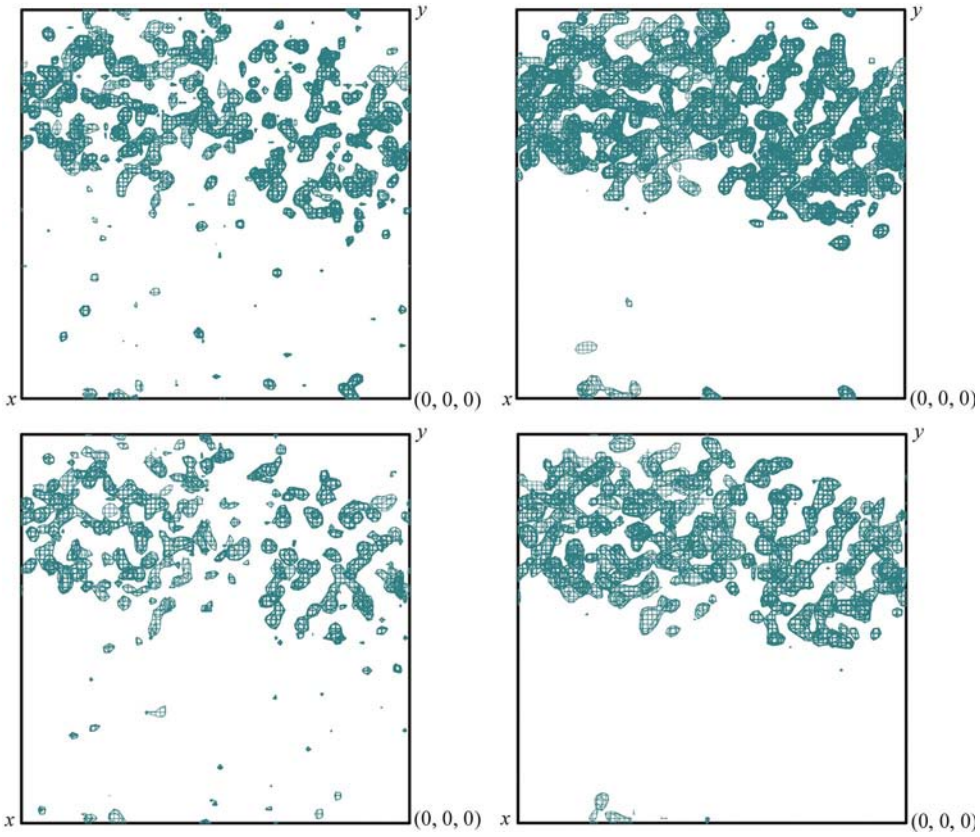

Figure 2
Comparison of the MAD-phased electron-density maps from *SOLVE* (left panels) and *RESOLVE* (right panels) using the twinned (upper panels) and detwinned (lower panels) data. A 30 Å slab of the $3F_o - 2F_c$ electron density cut at 2.5σ is shown along the c axis. The solvent boundary is already apparent from the initial *SOLVE* map, but the contrast of the map is dramatically increased after density modification. The quality of the density from the twinned data is slightly higher than that from the detwinned data.

3.2. Structure determination using twinned anomalous data

During the initial stages of structure determination, twinning was not considered. The systematic absences indicated $P4_1$ and $P4_3$ as possible space groups and consequently *SOLVE* was run with both options. Although in both cases a consistent set of six Se atoms was found, the electron-density maps were only interpretable in $P4_3$, thus establishing the correct space group for this crystal. The first electron-density map from *SOLVE* already showed a clear solvent boundary

and the contrast of the map was further enhanced by density modification in *RESOLVE* (Fig. 2, upper panel). However, the map quality was lower than expected for 1.8 Å resolution data and appeared somewhat blurred. Also, the occupancy of the Se atoms, which is output on an absolute scale by *SOLVE* (Tom Terwilliger, personal communication), was 0.4–0.5 (Table 2) and, although the Se content of the SeMet protein was not confirmed by mass spectrometry, incorporation of only 40–50% SeMet during mRNA translation seemed unlikely. Re-analysis of the data statistics then revealed the presence of hemihedral twinning with an estimated twin fraction of $\alpha = 0.40$ for the MAD data and of $\alpha = 0.35$ for the native data, which was taken into account during structure refinement (see below), but not during the initial phasing. It appears that the lower occupancy of the Se sites, as determined by *SOLVE*, reflects the relative occupancy of the Se atoms in the larger twin domain. Indeed, when the coordinates of two Se atoms were artificially twinned, *i.e.* transformed according to the twin law, and used as seeds for *SOLVE*, the remaining four Se positions of the less populated twin domain were found. However, a search for 12 Se sites was unsuccessful, because there are no cross-peaks in the anomalous difference Patterson map between Se atoms from the two twin domains.

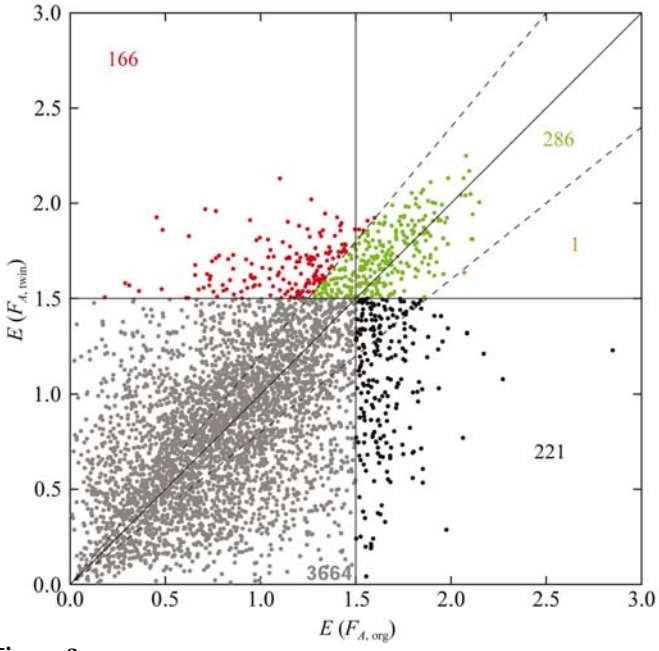

Figure 3
Scatter plot of 4338 E values derived as anomalous differences from generated normal (along horizontal direction) and twinned (along vertical direction) MAD data for the resolution range 10.0–3.0 Å. The diffraction intensities for the refined model of SeMet-substituted IL-1β in space group $P4_3$ were produced using *XPREP*, employing f' and f'' values as measured for the original diffraction data. The diffraction intensities were artificially twinned by calculating new intensities using equations (1) and (2) with a twinning fraction $\alpha = 0.4$ and a twinning operator of $k, h, -l$. F_A values were determined using *XPREP* following a standard protocol for both the original and the twinned data. A modified version of *SHELXD* was used to convert the F_A values to the E values that would be used by *SHELXD* in the 'half-baked' procedure. The regions above and below the pair of dashed lines correspond to reflections for which the deviation introduced by twinning is larger than ±20%. Reflections whose intensity diminished from $E > 1.5$ to $E \leq 1.5$ as a consequence of the twinning are shown as black dots; reflections for which both the untwinned and the twinned E values are ≤1.5 are shown in grey. Reflections for which $E > 1.5$ after twinning are shown in red, green and orange depending on whether the deviation is smaller (green) or larger (red and orange) than ±20% of the initial E value.

Similar to *SOLVE*, the *ab initio* program *SHELXD* (Schneider & Sheldrick, 2002) was able to identify all six Se positions when run against substructure structure factors (F_A values) derived from the twinned data using standard procedures in *XPREP* (Bruker AXS). The success of both *SOLVE* and *SHELXD* in determining the substructure can in part be explained by the fact that the starting models are derived by Patterson superposition methods. As the cross-vectors between the twin-related substructures are not present in the Patterson map, these methods deliver a single copy of the substructure. Furthermore, the fact that the dual-space recycling part of *SHELXD* uses only the largest E values (typically anomalous differences with $E > 1.5$, corresponding approximately to the largest 10% of all reflections), seems to be an advantage when working with twinned data.

To test the latter argument, artificial untwinned and twinned MAD data for IL-1β based on the final refined model were calculated in *XPREP*, anomalous difference E values were generated and their distributions were compared with each other (Fig. 3). If an E value for which the error introduced by twinning exceeds ±20% is deemed too inaccurate for substructure determination, the comparison shows that for a twin fraction of $\alpha = 0.40$, only 1707 (39.3%) of the 4338 twinned E values between 10 and 3.0 Å resolution are still sufficiently accurate (dots between the dashed lines in Fig. 3). However, for the 453 reflections with E values >1.5 after artificial twinning, 286 (63.1%) are still within the 20% error margin (green dots in Fig. 3). Thus, large E values seem to be less severely affected by twinning than small E values. To rationalize this observation, two scenarios under which an E value in a twinned data set will be strongly influenced by twinning (*i.e.* >20% of its value in the absence of twinning) have to be considered. Firstly, a twinned E value can be diminished relative to the untwinned value (222 reflections; orange and black dots in Fig. 3). For such cases, most of the resulting E values fall below the threshold of $E = 1.5$ (221 reflections; black dots in Fig. 3) and will therefore not be used in the real–reciprocal-space recycling procedure of *SHELXD*. Secondly, a large twinned E value can originate from a strong contribution from the second twin domain and thus erroneously enter the set of large E values (166 reflections; red dots in Fig. 3). Here, the latter situation amounts to 36.9% of the reflections with an inaccuracy of >20%. Interestingly, the real–reciprocal-space algorithm implemented in *SHELXD* seems to be sufficiently robust to cope with both the inaccuracy of more than a third of the normalized structure factors and with the incompleteness of the data. It should be noted that for cases with $\alpha < 0.4$, the contribution of the smaller twin domain will be smaller and hence the percentage of strongly affected (*i.e.* inaccurate or missing) E values will decrease, increasing the chances of a successful structure determination.

The substructure could also be determined using the detwinned MAD data, yielding the same subset of Se sites as in the twinned case. A key issue in detwinning data with an anomalous signal is how the indices are related for two twin-related reflections that are convoluted by the twin operator. For a pair of reflections that form a Friedel pair, the decision as to which one is labelled I^+ and which one is labelled I^- is arbitrary. This means that for two twin-related reflections, the twin law may mix I^+ with I^+ or it may mix I^+ with I^-, depending on the indices of the reflections or, more generally, on the choice of the asymmetric unit of reciprocal space. For example, if data in $P4$ are reduced so that h, k and l are all non-negative, then two reflections related by a twofold twin operation perpendicular to the c axis would be the (h, k, l) and (k, h, l). However, the real twin operation is a rotation, not a reflection. Those two reflections cannot be brought into coincidence by one of the rotation operators in the $P422$ space group; an inversion is required. Therefore, for the two reflections above (and essentially for a whole data set reduced in this asymmetric unit), I^+ from one reflection becomes mixed with I^- from another, which is the case for the IL-1β crystals.

Prior to detwinning, the data were anomalously scaled to a resolution of 1.9 Å and symmetry-equivalent reflections were merged to the asymmetric unit in *SCALEPACK*. Detwinning was performed taking into account the correct pairing of the anomalous intensities [*i.e.* $I^+(h, k, l)$ with $I^-(k, h, l)$] using in-house software (T. O. Yeates, unpublished program). A conservative estimate of $\alpha = 0.38$ rather than the estimated value of 0.40 was chosen to avoid possible overcorrection of the data, *i.e.* excessive generation of negative intensities. The resulting detwinned intensities were used for substructure determination in *SOLVE*. However, since the files output by this procedure are merged to the asymmetric unit, local scaling in *SOLVE* is different with the detwinned MAD data compared with the twinned MAD data, where the symmetry-equivalent reflections were kept separate. This may limit direct comparison of the results generated from the two differently scaled data sets. Also, although the data extended to 1.8 Å, in order to avoid generating too many reflections with negative intensities detwinning was only performed to a resolution of 1.9 Å. The resulting solutions from *SOLVE* and *RESOLVE* had FOMs for all reflections of 0.40 and 0.50, respectively. These are slightly lower than the FOMs from the twinned data (Table 1). The occupancies of all six Se sites were closer to one, as opposed to 0.4–0.5 for the untreated data (see above), confirming that the detwinning procedure was successful. Also, the B values for the six Se atoms ($\langle B \rangle = 38$ Å^2) did not change upon detwinning of the data. Counterintuitively, the Se—Se cross-peak heights using the detwinned data were in the range 13–18σ, compared with only 9–15σ for the twinned data (Table 2), although detwinning introduces additional noise into the data. Assuming identical errors of the twin-related reflections and a twin fraction of 0.4, the errors arising from detwinning increase by a factor of five (equation 5). Given this large error increase, it is surprising that the detwinned MAD data were not only still useable for phasing, but apparently yield a clearer solution for the Se sites. A possible explanation is that the increase in the random errors of the measured intensities owing to detwinning is offset by the removal of the systematic errors contributed by the scattering of the minor twin domain. However, the quality of the electron density derived from the twinned data is slightly higher than that from the detwinned data (Fig. 2). This deleterious effect of detwinning is also seen in the lower overall FOM of the final result from *SOLVE* and after density modification in *RESOLVE*, where all reflections are used. By contrast, the mean phase differences between the final model and the first phases from *SOLVE* and *RESOLVE* are very similar for the twinned and detwinned cases (Table 1).

Also, as might have been expected, molecular replacement using the twinned native data and a previously determined structure of IL-1β (Yu *et al.*, 1999) as the search model resulted in two solutions of nearly equal quality (R value of 34.8% and correlation coefficient of 0.41, with values of 40.4% and 0.39, respectively, for the best false solution), one for each twin domain. The two rotational parts of the molecular-replacement solutions are related to each other by the real-space twin operator, a 180° rotation about the crystallographic a axis.

3.3. Refinement against twinned native data and model analysis

A partial model (112 of 155 residues) was generated using the 1.8 Å resolution experimental electron-density maps generated by *SOLVE* from the twinned MAD data, refined with phase extension to 1.54 Å using the native data performed in *ARP*/*wARP* (Lamzin & Wilson, 1993). While detwinning of intensities is common practice for structure-determination purposes, refinement against detwinned data is not recommended owing to the systematic increase of the intensity errors. Thus, two refinement programs, *CNS* and *SHELXL*97, were considered for refinement of the initial IL-1β model. The first refinement cycles were performed in *CNS* using a fixed twin fraction of $\alpha = 0.35$ and the protocols for hemihedrally twinned data (see §2). However, the $R_{\rm free}$ value seemed to be stalled at unacceptably high values (24.3%) for this resolution (1.54 Å). Therefore, refinement was continued using *SHELXL*97 (Sheldrick & Schneider, 1997), which simultaneously refines the coordinates, the B values and the twin fraction. This refinement strategy converged to an $R_{\rm free}$ of 19.3% and produced a model (Fig. 4) with excellent stereochemical properties (Table 1). The twin fraction, which was fixed in *CNS* at $\alpha = 0.35$, refined to a slightly larger final value of 0.37. This final value differs by 5–14% from the values estimated by the various data statistics (Fig. 1, Table 1), indicating that the errors on the twin-fraction estimates are in the same range.

3.4. Conclusions

Recently, Yang *et al.* (2000) described a MAD structure determination of the λ-phage display platform protein gpD using twinned SeMet data. The crystal was of $P2_1$ symmetry and pseudo-merohedral twinning was possible owing to the

near-equality of the *a*- and *c*-axis lengths. The SeMet MAD data extended to 1.8 Å resolution and allowed ready solution of the substructure using *SOLVE*. Similar to the case of IL-1β described here, the possibility of twinning was not considered during structure determination and the initial stages of refinement. The high twin fractions ($\alpha > 0.35$ in both cases) did not impede structure determination by MAD phasing, but efficiently hampered refinement of the gpD and IL-1β structures. Ultimately, both structures were refined with *SHELXL* using the twinned native data and the TWIN and BASF options to define the twin operator and to refine the twin fraction, respectively. In general, it seems that the presence of twinning neither necessarily prohibits a successful structure determination using SeMet (or other anomalous) data, nor renders refinement impossible. The success of programs such as *SOLVE* and *SHELXD* in determining substructures from twinned MAD data is based on the use of Patterson superposition methods, which are robust because no cross-vectors between the twin domains exist in the anomalous Patterson maps. The real–reciprocal-space recycling procedure, as implemented in *SHELXD*, is also robust owing to the use of large *E* values of the anomalous differences that are least affected by twinning. If the twin fraction is particularly high, the additional noise introduced into the data by detwinning may more than counterbalance the positive effect of removing the contribution of the minor twin domain. Thus, phasing with twinned MAD data (without detwinning) is a possibility that should not be excluded from consideration. In the present case, attempts to phase using twinned and detwinned MAD data were similarly successful. However, there may well be cases where the twin fraction is so high that detwinning is impossible, while the twinned data can still be phased. We would expect that even larger substructures (>6 sites) can be determined with twinned MAD data despite the presence of a larger number of peaks in the Patterson maps owing to the two twin domains. Finally, with the arsenal of statistical twinning tests that, among others, are available in the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994), *XPREP* (Bruker, AXS), *CNS* (Brünger *et al.*, 1998) and the twinning server (Yeates, 1997), a test for twinning can easily be implemented as part of the data analysis prior to the time-consuming refinement procedures.

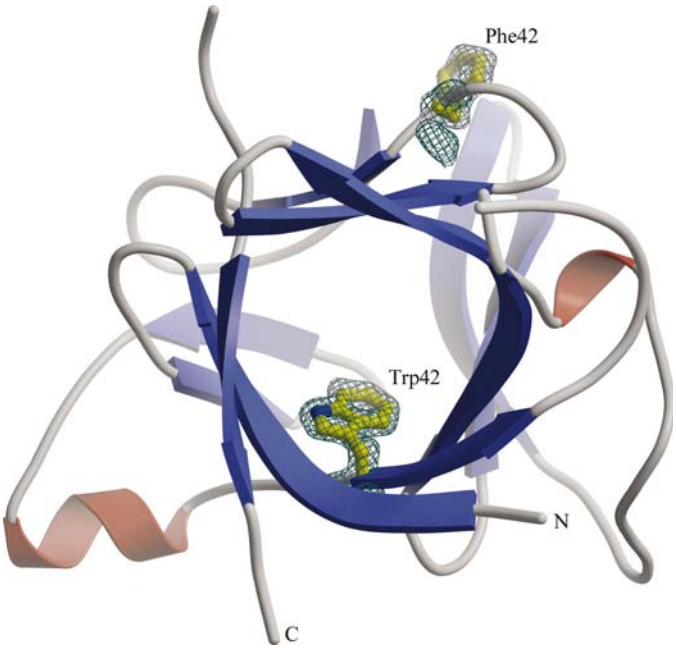

Figure 4
Ribbon representation of the final IL-1β model, with side chains drawn in ball-and-stick representation. The β-strands are coloured blue, α-helices red and loop regions grey. The termini are labelled. The final $F_o - F_c$ simulated-annealing omit maps around the F42W and W120F mutated residues are contoured at 2.5σ.

We thank the staff at APS beamline 19-ID and SSRL beamline 9-2 for support during synchrotron data collection, Michael Marino for collection of the native data set and George Sheldrick, Xiaoping Dai and Robyn Stanfield for helpful discussions. This study was supported by NIH grant CA58896 (IAW), a pre-doctoral fellowship from the Skaggs Institute (MSK) and postdoctoral fellowships from the German Academic Exchange Service and the Skaggs Institute (MGR). This is publication 15270-MB from the Scripps Research Institute.

References

Ban, N., Nissen, P., Hansen, J., Capel, M., Moore, P. B. & Steitz, T. A. (1999). *Nature (London)*, **400**, 841–847.
Breyer, W. A., Kingston, R. L., Anderson, B. F. & Baker, E. N. (1999). *Acta Cryst.* D**55**, 129–138.
Britton, D. (1972). *Acta Cryst.* A**28**, 296–297.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., 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.
Catti, M. & Ferraris, G. (1976). *Acta Cryst.* A**32**, 163–165.
Chandra, N., Acharya, K. R. & Moody, P. C. (1999). *Acta Cryst.* D**55**, 1750–1758.
Collaborative Computational Program, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Contreras-Martel, C., Martinez-Oyanedel, J., Bunster, M., Legrand, P., Piras, C., Vernede, X. & Fontecilla-Camps, J. C. (2001). *Acta Cryst.* D**57**, 52–60.
Covalt, J. C. Jr, Roy, M. & Jennings, P. A. (2001). *J. Mol. Biol.* **307**, 657–669.
Declercq, J. P. & Evrard, C. (2001). *Acta Cryst.* D**57**, 1829–1835.
Dinarello, C. A. (1996). *Blood*, **87**, 2095–2147.
Donnay, G. & Donnay, J. D. H. (1974). *Can. Mineral.* **12**, 422–425.
Epps, D. E., Yem, A. W., McGee, J. M., Tomich, C. S., Curry, K. A., Chosay, J. G. & Deibel, M. R. (1997). *Cytokine*, **9**, 149–156.
Ernst, J. A., Clubb, R. T., Zhou, H. X., Gronenborn, A. M. & Clore, G. M. (1995). *Science*, **267**, 1813–1817.
Finke, J. M., Roy, M., Zimm, B. H. & Jennings, P. A. (2000). *Biochemistry*, **39**, 575–583.
Finzel, B. C., Clancy, L. L., Holland, D. R., Muchmore, S. W., Watenpaugh, K. D. & Einspahr, H. M. (1989). *J. Mol. Biol.* **209**, 779–791.
Fisher, R. G. & Sweet, R. M. (1980). *Acta Cryst.* A**36**, 755–760.
Friedel, G. (1926). *Leçons de Cristallographie.* Nancy/Paris/Strasbourg: Berger–Leurault.
Heidary, D. K., Gross, L. A., Roy, M. & Jennings, P. A. (1997). *Nature Struct. Biol.* **4**, 725–731.
Hendrickson, W. A. & Ogata, C. M. (1997). *Methods Enzymol.* **276**, 494–523.

Herbst-Irmer, R. & Sheldrick, G. M. (1998). *Acta Cryst.* B**54**, 443–449.
Jones, T. A., Cowan, S., Zou, J. Y. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kahlenberg, V. V. (1999). *Acta Cryst.* B**55**, 745–751.
Lamzin, V. S. & Wilson, K. S. (1993). *Acta Cryst.* D**49**, 129–147.
Larsen, N. A., Heine, A., de Prada, P., Redwan, E.-R., Yeates, T. O., Landry, D. W. & Wilson, I. A. (2002). *Acta Cryst.* D**58**, 2055–2059.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Liang, A. G. W., Ealick, S., Nielsen, S., Schreiber, S. L. & Clardy, J. (1996). *Acta Cryst.* D**52**, 207–210.
Lietzke, S. E., Carperos, V. E. & Kundrot, C. E. (1996). *Acta Cryst.* D**52**, 687–692.
Luecke, H., Richter, H. T. & Lanyi, J. K. (1998). *Science*, **280**, 1934–1937.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Mueller, U., Muller, Y. A., Herbst-Irmer, R., Sprinzl, M. & Heinemann, U. (1999). *Acta Cryst.* D**55**, 1405–1413.
Mueller, U., Schubel, H., Sprinzl, M. & Heinemann, U. (1999). *RNA*, **5**, 670–677.
Oostrum, J. van, Priestle, J. P., Grutter, M. G. & Schmitz, A. (1991). *J. Struct. Biol.* **107**, 189–195.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Priestle, J. P., Schar, H. P. & Grutter, M. G. (1989). *Proc. Natl Acad. Sci. USA*, **86**, 9667–9671.
Rabijns, A., Verboven, C., Novoa de Armas, H., Van Damme, E. J., Peumans, W. J. & De Ranter, C. J. (2001). *Acta Cryst.* D**57**, 609–611.
Redinbo, M. R. & Yeates, T. O. (1993). *Acta Cryst.* D**49**, 375–380.
Schneider, T. R. & Sheldrick, G. M. (2002). *Acta Cryst.* D**58**, 1772–1779.
Shaanan, B., Gronenborn, A. M., Cohen, G. H., Gilliland, G. L., Veerapandian, B., Davies, D. R. & Clore, G. M. (1992). *Science*, **257**, 961–964.
Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **277**, 319–343.
Taylor, H. O., O'Reilly, M., Leslie, A. G. & Rhodes, D. (2000). *J. Mol. Biol.* **303**, 693–707.
Terwilliger, T. C. (2000). *Acta Cryst.* D**56**, 965–972.
Terwilliger, T. C. & Berendzen, J. (1996). *Acta Cryst.* D**52**, 749–757.
Terwisscha van Scheltinga, A. C., Valegard, K., Ramaswamy, S., Hajdu, J. & Andersson, I. (2001). *Acta Cryst.* D**57**, 1776–1785.
Trame, C. B. & McKay, D. B. (2001). *Acta Cryst.* D**57**, 1079–1090.
Treharne, A. C., Ohlendorf, D. H., Weber, P. C., Wendoloski, J. J. & Salemme, F. R. (1990). *Prog. Clin. Biol. Res.* **349**, 309–319.
Veerapandian, B., Gilliland, G. L., Raag, R., Svensson, A. L., Masui, Y., Hirai, Y. & Poulos, T. L. (1992). *Proteins*, **12**, 10–23.
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. & Fam, B. C. (1999). *Structure Fold. Des.* **7**, R25–R29.
Yem, A. W., Epps, D. E., Mathews, W. R., Guido, D. M., Richard, K. A., Staite, N. D. & Deibel, M. R. Jr (1992). *J. Biol. Chem.* **267**, 3122–3128.
Yu, B., Blaber, M., Gronenborn, A. M., Clore, G. M. & Caspar, D. L. (1999). *Proc. Natl Acad. Sci. USA*, **96**, 103–108.