Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Detecting and overcoming hemihedral twinning during the MIR structure determination of Rna1p

Roman Christian Hillig[a,b]* and Louis Renault[c]

[a]Schering AG Research Laboratories, 13342 Berlin, Germany, [b]Department of Structural Biology, Max Planck Institute for Molecular Physiology, Otto-Hahn-Strasse 11, 44227 Dortmund, Germany, and [c]Laboratoire d'Enzymologie et Biochemie Structurales, CNRS, 1 Avenue de la Terrasse, 91198 Gif-sur-Yvette, France

Correspondence e-mail: roman.hillig@schering.de

Received 19 February 2006
Accepted 2 May 2006

PDB Reference: Rna1p, 2ca6, r2ca6sf.

The structure of Rna1p was originally solved to 2.7 Å resolution by MIRAS from crystals with partial hemihedral twinning in space group $I4_1$ [Hillig *et al.* (1999), *Mol. Cell*, **3**, 781–791] by finding a low-twinned native crystal (twin fraction $\alpha = 0.06$) and after twin correction of all data sets. Rna1p crystals have now been used to examine how far twinning and twin correction affect MIR phasing with a higher resolution but highly twinned native data set. Even high hemihedral twinning [α(native) = 0.39, α(derivative) = 0.24] would not have hindered heavy-atom site identification of strong derivatives using difference Patterson maps. However, a weaker derivative could have been missed and refinement would have stalled at high R values had twinning not been identified and accounted for. Twin correction improved both site identification, experimental phasing statistics and MIR map quality. Different strategies were tested for refinement against twinned data. Using uncorrected twinned data and *TWIN-CNS*, Rna1p has now been refined to 2.2 Å resolution (final twinned R and $R_{\rm free}$ were 0.165 and 0.218, respectively). The increased resolution enabled release of the NCS restraints and allowed new conclusions to be drawn on the flexibility of the two molecules in the asymmetric unit. In the case of Rna1p, twinned crystal growth was possible owing to the presence of a twofold NCS axis almost parallel to the twin operator.

1. Introduction

Rna1p is the *Schizosaccharomyces pombe* orthologue of human RanGAP1, the GTPase-activating protein for Ran. The small GTPase Ran is a key regulator in nuclear transport, mitotic spindle assembly and nuclear envelope assembly (Pemberton & Paschal, 2005; Ciciarello & Lavia, 2005). The crystal structure of Rna1p was determined from partially hemihedrally twinned crystals (space group $I4_1$) by multiple isomorphous replacement with anomalous scattering (MIRAS; Hillig *et al.*, 1999). Whilst merohedral twinning is not unusual in small-molecule crystallography and was described early on (*e.g.* Buerger, 1969), this phenomenon was thought to be more unusual in macromolecular crystallography. This has changed over the last few years and merohedral twinning in protein crystallography is now regularly described and has been the subject of several review articles and method developments (Yeates, 1997; Yeates & Fam, 1999; Chandra *et al.*, 1999; Parsons, 2003; Dauter, 2003; Terwisscha van Scheltinga *et al.*, 2003; Padilla & Yeates, 2003; Lebedev *et al.*, 2006). In parallel, software tools to analyze data for twinning and to

account for it in refinement have become available. In addition to *SHELX* (Herbst-Irmer & Sheldrick, 1998), the *CCP*4 suite (*DETWIN*; Collaborative Computational Project, Number 4, 1994; Taylor & Leslie, 1998) and *CNS* (Brünger *et al.*, 1998) have been extended accordingly.

In most published cases, structures from merohedrally twinned crystals have been solved by molecular replacement (Redinbo & Yeates, 1993; Ito *et al.*, 1995; Luecke *et al.*, 1998; Breyer *et al.*, 1999; Ban *et al.*, 1999; Cox *et al.*, 1999; Weber *et al.*, 1999; Song *et al.*, 2001; Contreras-Martel *et al.*, 2001; Ko *et al.*, 2001; Abrescia & Subirana, 2002; Xu *et al.*, 2003; Ramadan *et al.*, 2002; Harris *et al.*, 2002; Yuan *et al.*, 2003; Lee *et al.*, 2003; Barends & Dijkstra, 2003; Rosendal *et al.*, 2004; Shimamura *et al.*, 2004; Govindasamy *et al.*, 2004; Golinelli-Pimpaneau, 2005; Schuermann *et al.*, 2005). Here, the diffraction data need not to be twin-corrected as the refinement software can refine the artificially twinned model against the original twinned data. In contrast, for an MIR structure determination partial merohedral twinning introduces an additional difference between the native and derivative crystals. A native and several derivative data sets have to be twin-corrected and the isomorphous differences between the resulting data sets analyzed. As twin correction is never perfect, this will inevitably introduce new errors. Therefore, it was not clear at the beginning of the structure determination of Rna1p whether the small isomorphous differences between the native and derivative data sets would still be detectable after twin correction. At this point, we were only aware of cases where combinations of isomorphous replacement and molecular replacement had been successful (Rees & Lipscomb, 1980; Reynolds *et al.*, 1985; Goldman *et al.*, 1987). It turned out that in our case MIRAS structure determination worked after finding a low-twinned native crystal (twin fraction 0.06) and after twin-correcting all data sets. In addition, a series of further novel structures have now been reported which have been determined from merohedrally twinned data using either isomorphous replacement methods (Igarashi, Moriyama, Fujiwara *et al.*, 1997; Forst *et al.*, 1998; Valegard *et al.*, 1998; Terwisscha van Scheltinga *et al.*, 2001; Declercq & Evrard, 2001; Hamburger *et al.*, 2004), multiple anomalous dispersion (Yang *et al.*, 2000; Rudolph *et al.*, 2003; Barends *et al.*, 2005) or *ab initio* methods (Schneider *et al.*, 2000).

1.1. Hemihedral twinning in tetragonal space groups

Twinning is a crystal-growth defect in which a crystalline specimen consists of more than one lattice and the lattices of the different domains are related by a defined symmetry operation. In non-merohedral twinning, the different domains of the crystal can be recognized from the crystal morphology or the twinning becomes obvious from the diffraction pattern, which shows reflections from more than one lattice. In merohedral twinning, the lattices of the separate twin domains superimpose perfectly in all three dimensions. Therefore, the corresponding reciprocal lattices also overlay exactly and each collected reflection is a superimposition of contributions from the two or more twin domains (two in the tetragonal system). In the tetragonal space group $I4_1$ (fourfold crystallographic axis along the c axis), the twin domains are related by a 180° rotation around the a or b axis or around the diagonal between these two axes. The resulting unit-cell orientation of the second twin domain fits into the same lattice and thus the diffraction patterns from both orientations also superimpose. This is expressed by the central equation of merohedral twinning, where any observed intensity ($I_{\rm obs}$) is a weighted sum of the true intensities of two reflections h_1 and h_2 of the untwinned crystals; the reflections h_1 and h_2 are related by the twinning operation but not by crystallographic symmetry (Yeates, 1997),

$$I_{\rm obs}(h_1) = (1-\alpha)I_{\rm true}(h_1) + \alpha I_{\rm true}(h_2), \quad (1)$$

$$I_{\rm obs}(h_2) = \alpha I_{\rm true}(h_1) + (1-\alpha)I_{\rm true}(h_2). \quad (2)$$

The twin fraction α represents the percentage of the less populated twin domain. A twin fraction of $\alpha = 0.50$ defines perfect merohedral twinning, whereas cases with $0 < \alpha < 0.50$ represent partial merohedral twinning. When there are only two possible unit-cell orientations as in the tetragonal lattice, this specific form of merohedral twinning is called hemihedral twinning.

As outlined, merohedral twinning can only occur when the orientation of the unit cell generated by the twin operation still fits into the lattice of the original unit cell and when at the same time the orientation of the molecules in the two unit-cell orientations are no longer equivalent. This holds true for only a limited group of tetragonal, trigonal, hexagonal and cubic space groups where the space-group symmetry is lower than the lattice symmetry (Yeates, 1997). This definition also describes the set of polar space groups where alternative non-equivalent indexing schemes exist. Finally, merohedral twinning can also occur in less symmetric space groups if under unfortunate circumstances the unit-cell geometry is very close or identical to a more highly symmetric one (*e.g.* an orthorhombic space group with $a = b$). Structure determinations from such pseudo-merohedrally twinned crystals have been described (Ito *et al.*, 1995; Ban *et al.*, 1999; Mueller *et al.*, 1999; Frazão *et al.*, 1999; Weber *et al.*, 1999; Yang *et al.*, 2000; Declercq & Evrard, 2001; Larsen *et al.*, 2002; Barends & Dijkstra, 2003; Houborg *et al.*, 2003; Dauter, 2003; Rudolph *et al.*, 2004; Hamburger *et al.*, 2004; Golinelli-Pimpaneau, 2005; Lehtiö *et al.*, 2005).

1.2. Symptoms of merohedral twinning

Merohedral twinning cannot be directly recognized in the diffraction pattern. When the data are processed in the correct space group, data-collection statistics such as $R_{\rm sym}$ are not affected and cannot help in identifying the problem, as all symmetry-related reflections suffer from exactly the same disturbance. However, correct space-group assignment is often difficult with merohedrally twinned crystals, which in itself may be an initial hint: in the case of perfect merohedral twinning ($\alpha = 0.50$), the two observed twin-related intensities $I_{\rm obs}(h_1)$ and $I_{\rm obs}(h_2)$ become equal, which increases the apparent symmetry of the crystal. A self-rotation function

Table 1
Data-collection and refinement statistics (space group $I4_1$, No. 80).

Soaking conditions: $HgCl_2$, 5 m*M*, 21 h; $HgAc_2$, 5 m*M*, 24 h; K_2PtCl_4, 10 m*M*, 13 h; $KAu(CN)_2$, 2 m*M*, 30 h; MeHgBr, 5 m*M*, 38 h; thiomersal, 5 m*M*, 17 h; EtHgCl, 1 m*M*, 6 h. Values in parentheses refer to the highest resolution shell.

Data set	Native-A	Native-B†	$HgCl_2$	$HgAc_2$†	K_2PtCl_4	$KAu(CN)_2$	MeHgBr	Thiomersal	EtHgCl†
Twin fraction α‡	0.34 [0.35]	0.06 [0.07]	0.12	0.18	0.19	0.21	0.23	0.24	0.24
Source	BW7A, DESY	BM14, ESRF	BM14, ESRF	BW6, DESY	X11, DESY	BM14, ESRF	BM14, ESRF	BW6, DESY	BW6, DESY
Unit-cell parameters (Å)									
a	175.2	175.5	176.0	175.1	175.6	175.7	175.5	175.0	175.3
c	55.9	55.8	56.1	56.0	55.9	56.0	56.1	56.1	56.0
Mosaicity (°)	0.50	0.69	0.39	0.69	0.75	0.85	0.87	0.75	0.64
Resolution (Å)	31.0–2.2	20.0–2.6	25.0–3.2	40.0–3.0	19.7–3.0	25.0–3.2	22.0–3.2	29.7–3.2	29.7–3.0
Highest shell (Å)	2.25–2.20	2.69–2.64							
Unique observations	42106 (2493)	23730	13970	17194	16829	14337	13729	14112	16912
Multiplicity	3.8 (2.0)	6.1	5.6	3.1	2.9	4.0	5.0	2.9	2.6
Completeness	0.97 (0.87)	0.98 (0.59)	0.99	0.99	0.96	0.99	0.99	0.98	0.97
$I/\sigma(I)$	8.0 (2.1)	18.2 (4.2)	16.2	13.7	9.4	19.3	12.4	10.6	13.1
R_{sym}§	0.137 (0.369)	0.084 (0.196)	0.093	0.077	0.086	0.061	0.103	0.095	0.072
R_{iso}¶	0.207	—	0.188	0.229	0.139	0.139	0.212	0.232	0.169
Refinement	Strategy 2††	Strategy 1†,††							
PDB code	2ca6	1yrg							
Twin fraction used	0.39	0.06							
Resolution (Å)	31.0–2.20	20.0–2.66							
Highest shell (Å)	2.24– 2.20	2.83–2.66							
Reflections (work/test)	37258/3957	22515/1215							
R_{work}	0.277 (0.338)	n.d.							
R_{free}‡‡	0.312 (0.378)	n.d.							
Detwinned R_{work}§§	0.257 (0.379)	0.228 (0.299)							
Detwinned R_{free}‡‡	0.301 (0.475)	0.277 (0.373)							
Twinned R_{work}¶¶	0.165 (0.231)	n.d.							
Twinned R_{free}‡‡	0.218 (0.279)	n.d.							
Wilson *B* factor††† (Å^2)	14.4	34.0							
No. non-H atoms									
Total	5673	5399							
Molecule *A*	2690	2689							
Molecule *B*	2690	2689							
Protein	5380	5378							
Water molecules	273	21							
No. sulfate molecules	4	—							
Average *B* factors (Å^2)									
All atoms	22.7	19.9							
Molecule *A*	20.2	18.8							
Molecule *B*	24.7	21.1							
Protein	22.4	19.9							
Water molecules	24.1	13.2							
Sulfates	68.6	—							
R.m.s.d. from ideal geometry									
Bond lengths (Å)	0.005	0.007							
Angles (°)	1.23	1.31							
Ramachandran plot, residues in (%)									
Most favoured regions	494	490							
Additionally allowed regions	119	124							
Generously allowed regions	5	4							
Disallowed regions	2	2							

† Taken from Hillig *et al.* (1999). ‡ Twin fractions determined using the program described in §2.3. For native-A and native-B, the additional value in square brackets was determined using *CNX* (`twin_fraction.inp`). § $R_{sym} = \sum_h \sum_i |I_{h,i} - \langle I_h \rangle| / \sum_h \sum_i I_{h,i}$. ¶ Calculated with native-B as reference using *CCP4*/*SCALEIT* and data to 2.7 Å. †† See §3.4. ‡‡ R_{free} is the same as the respective *R* factor, but calculated on 10% (native-A) and 5% (native-B) of data excluded from refinement. §§ Detwinned *R* is the conventional *R* factor, $R = \sum_h |F_o - F_c| / \sum F_o$, but calculated from the twin-corrected experimental structure factors F_o and the calculated model structure factors F_c. ¶¶ Twinned *R* is the *R* factor calculated from the original (not twin-corrected) experimental structure factors and the artificially twinned model structure factors. ††† Determined by *CCP4*/*TRUNCATE*.

analysis will show additional symmetry elements and the data set can be processed in a higher yet wrong space group with excellent statistics. In the case of partial merohedral twinning, depending on the twin fraction α, $I_{obs}(h_1)$ and $I_{obs}(h_2)$ become more similar than expected for a non-twinned data set. Processing in the wrong higher symmetric space group does not result in excellent R_{sym} values, but they are also not as poor as one would expect for a false space group. Often, merohedral twinning can also be recognized from packing considerations. The Matthews parameter may suggest that the higher apparent symmetry caused by the twinning requires a number of molecules per unit cell that cannot reasonably be accommodated (*e.g.* Luecke *et al.*, 1998).

Apart from these symmetry considerations, merohedral twinning can be detected by analyzing whether the cumulative intensity distribution follows Wilson statistics (Wilson, 1949;

Table 2
Twinning analysis.

		Before detwinning			After detwinning			
Data set	Twin fraction α†	$\langle I^2\rangle/\langle I\rangle^2$‡	φ (self-rotation)§ (°)	R_{iso}¶ (%)	$\langle I^2\rangle/\langle I\rangle^2$‡	φ (self-rotation)§ (°)	R_{iso}¶ (%)	Heavy-atom bound?
Native-B	0.06	2.08	−2.0	n.a.	2.13	−2.5	n.a.	n.a.
Native-A	0.34	1.73	0.0	20.7	2.00	−1.5	15.4	n.a.
$HgCl_2$	0.12	2.13	−2.0	18.8	2.17	−3.0	18.0	Yes
$HgAc_2$	0.18	1.95	−1.5	20.1	2.18	−2.5	21.2	Yes
K_2PtCl_4	0.19	1.81	−1.0	13.9	2.05	−1.0	12.6	No
$KAu(CN)_2$	0.21	1.84	−1.5	13.9	2.15	−2.0	11.7	No
MeHgBr	0.23	1.91	−1.5	21.2	2.16	−2.5	21.1	Yes
Thiomersal	0.24	1.90	−1.5	23.2	2.04	−3.0	23.3	Yes
EtHgCl	0.24	1.77	−0.5	18.0	2.07	−1.0	16.2	Yes

† As determined by the twin-correction program described in §2. ‡ The ratio $\langle I^2\rangle/\langle I\rangle^2$ is expected to be 2.0 for a non-twinned crystal ($\alpha = 0.0$) and 1.5 for a perfectly merohedrally twinned crystal ($\alpha = 0.5$). § Polar angle φ of the NCS operator in the self-rotation function ($\omega = 90$, $\kappa = 180°$). After model building and refinement, φ was −2.1° for the final model. ¶ R_{iso} calculated with reference to native-B. For native-A it was calculated to a resolution limit of 2.7 Å and for all others to the maximum resolution (see Table 1).

Stanley, 1972). In every diffraction data set, a certain number of reflections with weak, medium and strong intensities are expected. (1) and (2) show that in the presence of merohedral twinning a very weak intensity $I_{true}(h_1)$ is superimposed with a second usually stronger intensity $I_{true}(h_2)$. This leads to a reduced number of very weak (and very strong) intensities in the collected data set. In the cumulative intensity distribution plot, this triggers a characteristic sigmoidal deviation of the experimental curve from the theoretical curve. This effect can be used to identify twinning *via* calculation of $\langle I^2\rangle/\langle I\rangle^2$ in thin resolution shells (Stanley, 1972; Yeates, 1997). For acentric data, the expected value is 2.0 in the absence of merohedral twinning and 1.5 for perfect merohedral twinning ($\alpha = 0.5$). It should be mentioned that in unfortunate cases even analysis of the intensity distribution may not reveal merohedral twinning. Under certain conditions, the presence of NCS translational symmetry may systematically increase the number of very weak reflections and thus reverse the effect of twinning (Lee *et al.*, 2003; Padilla & Yeates, 2003; Dauter *et al.*, 2005).

Here, we present the symptoms which led to the identification of hemihedral twinning during the MIRAS structure determination of Rna1p. Different strategies during phasing and crystallographic refinement and their influence on the localization of heavy-atom sites, MIR map calculation and refinement are compared, now using a higher resolution but highly twinned native data set ($\alpha = 0.39$). Rna1p was finally refined to 2.2 Å resolution using this data set and the higher resolution structure is discussed. This case study may help to optimize other MIR structure determinations with highly twinned data.

2. Experimental

2.1. Data collection

Expression, purification and crystallization of full-length *S. pombe* Rna1p (43.1 kDa, 386 residues), as well as the MIRAS structure determination (using a 2.7 Å native data set as well as SeMet, EtHgCl and $HgAc_2$ derivatives), have been described previously (Hillig *et al.*, 1999). The native data set used in this original structure determination will hereafter be called native-B. Here, we report the use of a higher resolution but highly twinned native data set originally collected earlier (named native-A) with a resolution limit of 2.2 Å and five further data sets collected from crystals soaked in heavy-atom solutions. These latter data sets did not contribute to the MIRAS phasing, but were of help for detecting the hemihedral twinning and validating the twin-correction procedure. Data-collection statistics are given in Table 1. Because of the small size of the needle-shaped crystals, all data had to be collected using synchrotron radiation, at various beamlines at ESRF, EMBL Outstation Grenoble and at DESY, EMBL Outstation Hamburg. Crystals were frozen as described in Hillig *et al.* (1999). All data were collected on MAR image-plate detectors.

2.2. Data analysis

Data were processed using the *DENZO*/*SCALEPACK* package (Otwinowski & Minor, 1997), subjected to *SCALEPACK2MTZ*, *TRUNCATE* and *CAD* and analyzed with *XLOGGRAPH* (Collaborative Computational Project, Number 4, 1994). The values for $\langle I^2\rangle/\langle I\rangle^2$ per resolution shell and the cumulative intensity distributions were extracted from the output of *TRUNCATE*. Values for $\langle I^2\rangle/\langle I\rangle^2$ for the individual data sets are listed in Table 2. Self-rotation functions were calculated using *POLARRFN* (Collaborative Computational Project, Number 4, 1994).

2.3. Twin correction

(1) and (2) allow the calculation of the true intensities from the observed intensities if the twin fraction α is known. For correction, both twin-related intensities $I_{obs}(h_1)$ and $I_{obs}(h_2)$ have to be collected and α has to be determined. In space group $I4_1$, the symmetry operation relating the two possible orientations of the unit cell is a 180° rotation around the *a* or *b* axes or the diagonal between them. A reflection (h, k, l) is twin-related to $(h, -k, -l)$ and to its symmetry mates $(-h, k, -l)$, $(k, h, -l)$ and $(-k, -h, -l)$. When processing without anomalous signal, both *XDS* (Kabsch, 1993) and *SCALEPACK* (Otwinowski & Minor, 1997) list (h, k, l) and $(h, -k, l)$ in their output. In this case, the reflection $(h, -k, l)$ is

a Friedel mate of one of the twin-related reflections and the two observed intensities $I_{obs}(h_1)$ and $I_{obs}(h_2)$ correspond to the intensities of the pair of (h, k, l) and $(h, -k, l)$. According to Yeates (1988), the twin fraction α can be determined by first calculating the fractional difference H with

$$H = [I_{obs}(h_1) - I_{obs}(h_2)]/[I_{obs}(h_1) + I_{obs}(h_2)] \quad (3)$$

for each pair of twin-related reflections. For acentric reflections α is then calculated from H *via*

$$\langle H \rangle = \tfrac{1}{2} - \alpha, \quad (4)$$

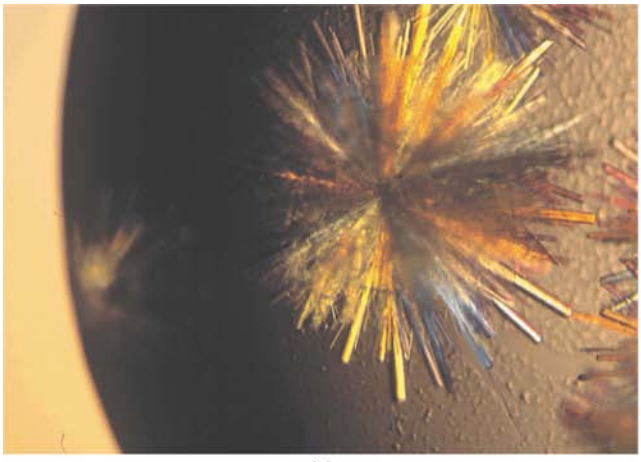
$$\langle H^2 \rangle = (1 - 2\alpha)^2/3. \quad (5)$$

For centric reflections, a similar relationship exists (see Yeates, 1988). Since no program for correcting twinned data was available at the beginning of the structure determination of Rna1p, we wrote a program[1] that first sorts out the intensities in space group $I4_1$ which are not affected (reflections with $k = 0$, $h = 0$ or $h = k$), then identifies the collected pairs $I_{obs}(h, k, l)$ and $I_{obs}(h, -k, l)$, calculates α from $\langle H \rangle$, subsequently recalculates the true intensities $I_{true}(h, k, l)$ and $I_{true}(h, -k, l)$ and outputs the corrected and the unaffected intensities as the twin-corrected data set. To handle the anomalous signal in one of the SeMet data sets described in Hillig *et al.* (1999), the program was extended. Here, $I(+)(h, k, l)$ and $I(-)(h, -k, l)$ represent one twin-related pair of intensities and $I(-)(h, k, l)$ and $I(+)(h, -k, l)$ represent a second one. Each of these pairs was then used to calculate a value for H and was afterwards twin-corrected with the twin fraction derived from $\langle H \rangle$.

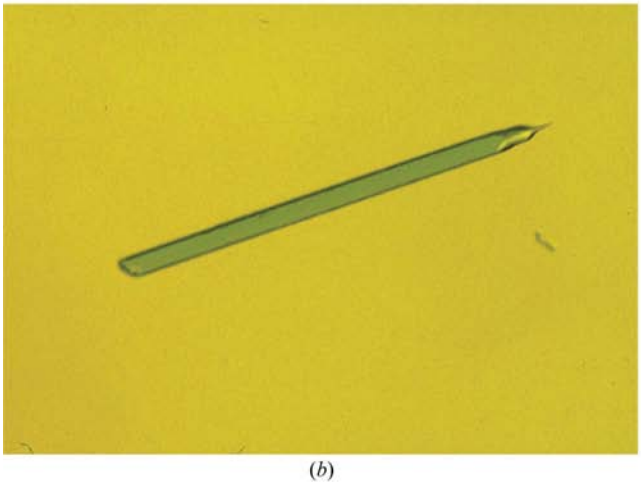
If there is an NCS axis closely parallel to one of the crystallographic axes, the statistical basis for the determination of α from H is no longer valid (Yeates, 1988; Padilla & Yeates, 2003). For this case, Forst and coworkers suggested a different way to determine α (Forst *et al.*, 1998). This method was initially followed up in parallel, but was dropped when it became clear that with Rna1p the NCS is almost parallel to, but translated in space from, the crystallographic axes a and b.

After most of the twin-correction work described above had been completed, a number of crystallographic packages were extended to handle merohedral twinning: *DETWIN* from the

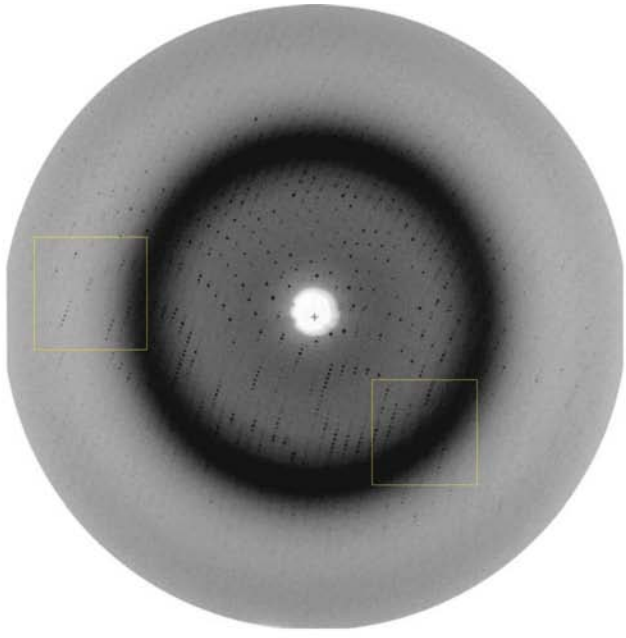
[1] The program (Fortran code) can be obtained from the authors on request. However, newer programs such as *CNS* and *CCP*4 *DETWIN* may be more useful as they are not restricted to space group $I4_1$.

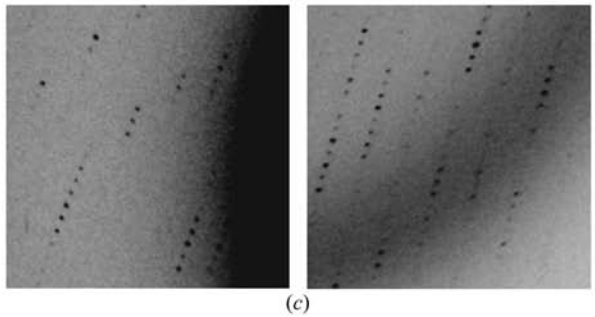


(*a*)



(*b*)





(*c*)

Figure 1
Crystals and diffraction pattern of *S. pombe* Rna1p. (*a*) Typical sea-urchin-like crystal clusters. (*b*) Manually separated crystals of about 600 × 40 × 40 µm. (*c*) Diffraction pattern from native-A (high-resolution sweep, $\Delta\varphi = 0.5^\circ$, exposure time 8 min, detector edge 2 Å). Yellow boxes denote the enlarged regions. The reflections show no signs of splitting.

*CCP*4 package can perform the twin correction (Taylor & Leslie, 1998). A number of subroutines of *CNS* (v.0.9 and higher) can determine the twin fraction, twin-correct the data or take the twinning into account during refinement (Brünger *et al.*, 1998). *SHELXL*, as a program developed for small-molecule crystallography, can handle merohedral twinning during refinement (Herbst-Irmer & Sheldrick, 1998) and a web-based server helps in identifying twinning (http://www.doe-mbi.ucla.edu/Services/Twinning/).

2.4. MIR phasing with twin-corrected data

Heavy-atom sites were originally identified using *SOLVE* (Terwilliger & Berendzen, 1999) with twin-corrected data and confirmed in difference Patterson and difference Fourier maps as described in Hillig *et al.* (1999). Since several recent reports have described the use of twinned data with an anomalous signal (Yang *et al.*, 2000; Rudolph *et al.*, 2003; Barends *et al.*, 2005), we focused here on the analysis of experimental phasing by MIR, which is expected to require the determination of both accurate native and derivative structure factors. For analysis of the influence of twin correction on site identification and phasing, difference Patterson maps were calculated using the original and the twin-corrected data (*FFT* from the *CCP*4 package). For both cases, the sites were refined using *MLPHARE* (Collaborative Computational Project, Number 4, 1994) and density-modified NCS-averaged MIR phases were calculated using *DM* (Collaborative Computational Project, Number 4, 1994).

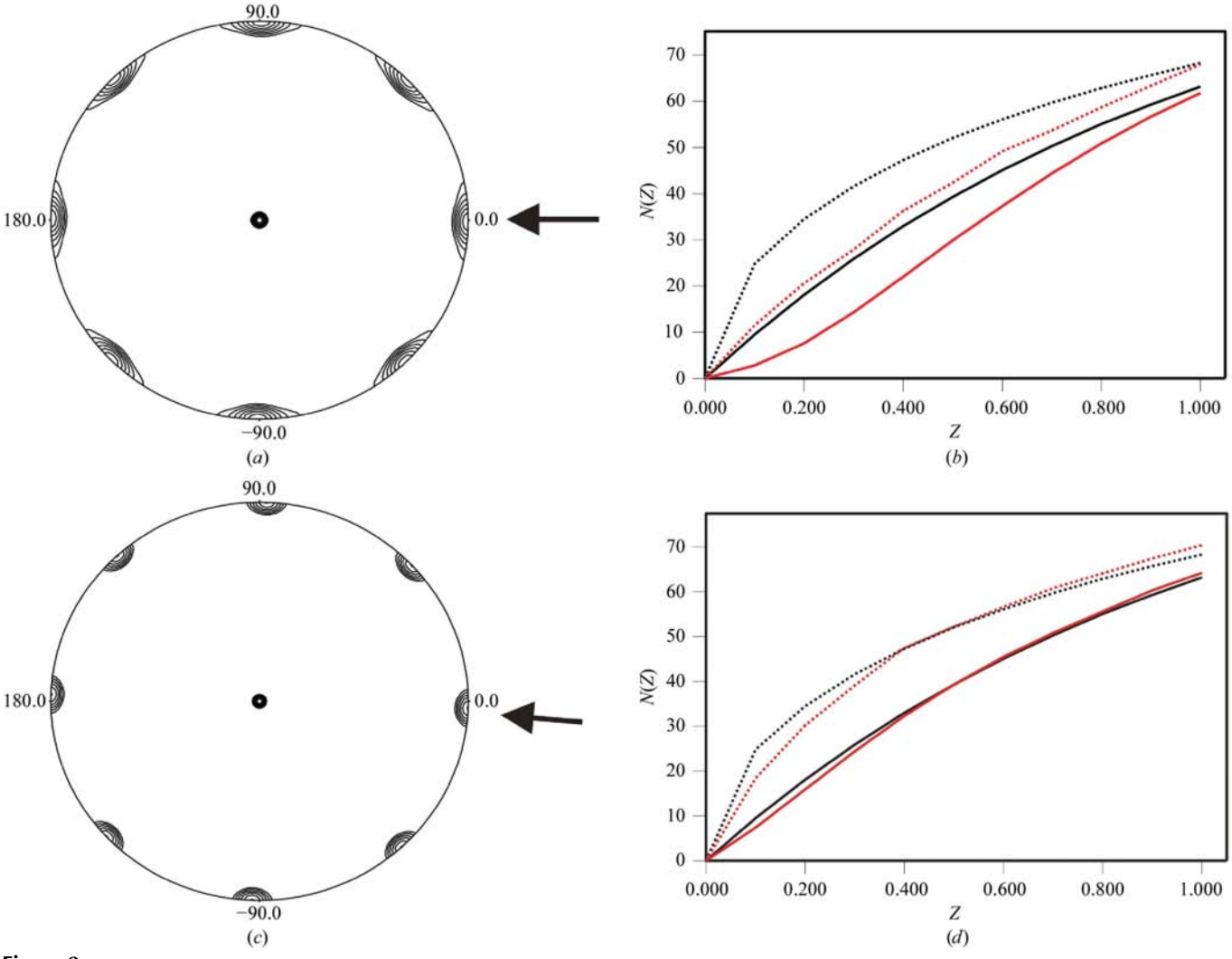


Figure 2
Self-rotation functions and cumulative intensity distributions of native-A and native-B (both processed in *I*4 and cut at 2.7 Å resolution for comparison). (*a*) and (*b*) show the self-rotation function and the intensity distribution for native-A. Self-rotation function: $\kappa = 180^\circ$ section, φ plotted on the circumference, ω along the radius (centre, $\omega = 0^\circ$; edge, $\omega = 90^\circ$), resolution 8–4.3 Å. Note that in addition to the central peak representing the crystallographic fourfold axis, there is a peak (and its symmetry mates) at $\varphi = 0^\circ$. The cumulative intensity distribution for native-A (theoretically expected curves are shown as a black line for acentric reflections and a dotted black line for centric reflections and experimentally observed distributions are shown as a red line for acentric reflections and a dotted red line for centric reflections) shows a pronounced sigmoidal deviation from the expected values, which is typical of merohedral twinning (see text). (*c*) and (*d*) show the self-rotation function and intensity distribution for the less twinned native-B. Here, the self-rotation peak is at $\varphi = -2^\circ$, revealing a twofold NCS axis, and the intensity distributions shows only a very small sigmoidal deviation.

2.5. Refinement against data set native-A

The model of Rna1p (PDB code 1yrg; originally built and refined to 2.7 Å resolution using the twin-corrected data set native-B) was refined against the original (not twin-corrected) structure factors of native-A using *CNS* in normal or in TWIN mode and employing various resolution cutoffs. $2F_o - F_c$ difference maps were inspected for quality of the electron density. For these refinement runs a new test set for R_{free} calculation was introduced using the `make_cv_twin` routine of *CNS*, which ensures that twin-related pairs of reflections are included either both in the work or both in the test set. Coordinates of the initial model 1yrg (without water molecules) were subjected to a random shift of 0.5 Å using *ARP*/*wARP* (Collaborative Computational Project, Number 4, 1994) followed by a simulated-annealing refinement run (using data to 2.2 Å resolution and NCS restraints) to remove model bias from the new R_{free} test set. The resulting model was subjected to iterative cycles of positional refinement, individual *B*-factor refinement and manual rebuilding in $3F_o - 2F_c$ and $F_o - F_c$ maps using *O* (Jones & Kjeldgaard, 1997) and *QUANTA* (Accelrys). The final stages of refinement were carried out using *CNX* (Accelrys), the commercial version of *CNS*. (For simplicity, we will refer to both as '*CNS*' in the following.) The final model contains 684 residues (two Rna1p molecules; NCS restraints completely released), 273 water molecules and four sulfate ions. It was refined to 2.2 Å resolution to a twinned *R* value of 0.165 (twinned R_{free} = 0.218). All residues except Ser257 are in the allowed regions of the Ramachandran plot (*PROCHECK*; Collaborative Computational Project, Number 4, 1994). Ser257 shows well defined electron density in both molecules. It is located in the leucine-rich repeat LRR9 just before the beginning of α-helix H9 (Hillig *et al.*, 1999) and is engaged in a series of hydrogen bonds.

3. Results and discussion

3.1. Identification of hemihedral twinning

S. pombe Rna1p crystallizes in sea-urchin-like clusters of tetragonal needles, with individual needles having a length of up to 1 mm but a thickness of only 20–40 µm, extending in very rare cases up to 60 µm (Fig. 1). Even with synchrotron radiation, only needles of more than 40 µm thickness diffracted to a resolution of better than 3.5 Å. Separated crystals did not feature any growth defects indicative of twinning and none of the crystals showed split reflections (Fig. 1). Hemihedral twinning was recognized when a high-resolution native data set and a number of potential derivative data sets had already been collected. Up to this point, the space-group determination had been uncertain, indicating either $I4_1$ or $I4_122$. Twinning could not be detected from packing considerations as the large asymmetric unit allowed the presence of two NCS-related protein molecules for space group $I4_1$ or one molecule for space group $I4_122$, both resulting in identical and reasonable Matthews parameters.

Structure determination was initiated with highly twinned native data sets such as native-A (Table 1). Data processing of native-A resulted in an R_{sym} value of 0.069 for $I4_1$ and 0.139 for $I4_122$ (calculated for data to 2.7 Å). This difference should have implied that the higher symmetric space group was not correct and may have hinted at the presence of twinning. In particular, the *R* factors for the lowest resolution shell (31–5.8 Å) were unusually high (0.096 for $I4_122$ compared with 0.041 for $I4_1$). At the time, however, the higher value for $I4_122$ appeared acceptable given the small crystal size, the weak diffraction and an anisotropic diffraction pattern arising from the needle shape of the crystals. The self-rotation function of such highly twinned native data sets of Rna1p gave ambiguous results. Fig. 2(*a*) shows the $\kappa = 180^\circ$ section for native-A. The peak at $\omega = 90$, $\kappa = 180$, $\varphi = 0^\circ$ and the symmetry-related peaks at φ = 45, 90, 135° *etc*. could either have represented the crystallographic twofold axes of $I4_122$ or twofold NCS axes (parallel or almost parallel to the crystallographic *a* and *b* axes) in $I4_1$.

Hemihedral twinning was identified by analysis of $\langle I^2\rangle/\langle I\rangle^2$ and the cumulative intensity distribution plots of all data sets. Fig. 2(*b*) shows this distribution for native-A. The sigmoidal deviation from the expected curves, a characteristic of merohedral twinning, was present in all native and derivative data sets, though sometimes more and sometimes less pronounced. Its variation indicated partial hemihedral twinning with twin fractions changing from crystal to crystal. Similarly, the average value of $\langle I^2\rangle/\langle I\rangle^2$ over all resolution shells was below 2.0 for all data sets but one (see Table 2). After realising the presence of hemihedral twinning, we screened for a less twinned crystal and finally collected native-B. This data set showed a much better intensity distribution (Fig. 2*d*) and turned out to be only very slightly twinned.

Analysis of the self-rotation function of native-B (Fig. 2*c*) revealed that the observed peaks in the $\kappa = 180^\circ$ section were clearly not at $\varphi = 0^\circ$ but at -2° and thus represented a twofold NCS axis rather than crystallographic symmetry. In a more highly twinned crystal, the NCS peaks at $\varphi = -2^\circ$ of the first twin domain and at $+2^\circ$ of the second twin domain would average, resulting in a peak at $\varphi = 0^\circ$, which indeed was observed with native-A (Fig. 2*a*). In addition, the symmetry introduced by the twin operation itself gives rise to a peak at $\varphi = 0^\circ$ for high twin fractions. Comparison of the peak shapes in the self-rotation functions of native-A and native-B (Figs. 2*a* and 2*c*) shows that the peak at $\varphi = 0^\circ$ (introduced by twinning in combination with the NCS axis) is broadened, whereas the 'pure' NCS peak at $\varphi = -2^\circ$ is much sharper. This is consistent with the idea that the peak of the highly twinned native-A is a superimposition of the twin symmetry (peak at $\varphi = 0^\circ$) and of the two NCS peaks at -2 and $+2^\circ$. This broadening of the self-rotation peak requires the presence of an NCS axis almost parallel to the twinning operator. It is therefore not a general feature of merohedrally twinned crystals. The identification of heavy-atom derivatives and the calculation of interpretable MIR electron-density maps finally confirmed that the true space group was $I4_1$, with two NCS-related molecules per asymmetric unit. The polar angle of the NCS operation

relating the two independently refined molecules in the final model (see §3.4) is $\varphi = -2.1^\circ$.

3.2. Correction of the twinned data

An intensity $I_{obs}(h_1)$ can only be corrected if both $I_{obs}(h_1)$ and the twin-related intensity $I_{obs}(h_2)$ have been collected. Missing one of the two during data collection leads to the loss of the other during twin correction. For a data set from merohedrally twinned crystals, extra emphasis should therefore be put on high completeness. In addition, completeness is further decreased during twin correction as some of the twin-corrected intensities become negative. In the work presented here, both the reflection with the negative intensity and its twin-related reflection were discarded in such cases. The twin fractions α of all data sets were determined (Table 2, derivative data sets sorted according to α) and the intensities were twin-corrected using the set of programs described in §2.3. For each data set, Table 2 lists the values of three parameters ($\langle I^2\rangle/\langle I\rangle^2$, the angle φ of the NCS peak in the self-rotation function and R_{iso} against native-B) for the situation before and after twin correction. All crystals were partially twinned with α ranging from 0.06 to 0.43 (the latter was for a poorly diffracting crystal not reported here). The value of α correlated well with the values of $\langle I^2\rangle/\langle I\rangle^2$, except for the $HgCl_2$ data set. In this case, an α value of 0.12 and the intensity distribution pointed towards a moderately high degree of twinning, while the value of $\langle I^2\rangle/\langle I\rangle^2$ (2.13) would have masked the presence of twinning, indicating the importance of checking the cumulative intensity distribution.

The success of the detwinning procedure was verified by two parameters. Firstly, the corrected data showed an improved fit to the Wilson statistics. After twin correction, the cumulative intensity distribution plots showed a reduced deviation from the expected distribution (Fig. 3). Similarly, the values of $\langle I^2\rangle/\langle I\rangle^2$ increased towards 2.0 and higher (Table 2). Secondly, the φ coordinate of the twofold NCS axis moved away from $\varphi = 0^\circ$ and towards $\varphi = -2^\circ$ or even further (Table 2).

Data sets from derivative soaks that ultimately showed heavy-atom binding featured R_{iso} values of 18–23% when compared with native-B (Table 2). The two data sets where no heavy-atom binding was found showed lower R_{iso} values (14%). This value is, however, still surprisingly high and the comparison of native-A and native-B also revealed a significant R_{iso} (21%). This was not correlated with changes in unit-cell volumes compared with native-B, which were smaller than 1.2% for all data sets. Part of the large isomorphous difference can be attributed to the difference in twin fraction, which is low only for native-B. After twin correction, R_{iso} between the two native data sets decreased by about 5% (Table 2).

3.3. MIR analysis: heavy-atom site detection, refinement and phasing

Originally, the structure was solved and refined using the twin-corrected native-B and twin-corrected derivative data sets (Hillig *et al.*, 1999). The heavy-atom binding sites used for phasing had been identified by *SOLVE* (with all native and derivative data twin-corrected and the low-twinned native-B chosen as native data set). The gold and platinum soaks showed no heavy-atom binding sites. A series of mercury compounds tested in soaking experiments showed binding, but only two of these data sets were finally used: a shorter soak with EtHgCl at a low concentration (1 m*M*, 6 h), which resulted in two bound Hg atoms, and a longer soak with $HgAc_2$ (5 m*M*, 24 h), which led to the identification of four mercury-binding sites by *SOLVE*. We now analyzed how far MIR phasing would have been hindered had the twinning not been detected.

For both these derivatives, the difference Patterson maps were re-calculated, first using the highly twinned native-A data set and without applying any twin correction (Fig. 4, left maps) and then with the low-twinned native-B and twin correction of both the native and the derivative data (Fig. 4, right maps). The shorter soak with EtHgCl resulted in very clear Harker peaks in both cases (Fig. 4*a*). Almost all major peaks could be explained by the two sites (predicted peak positions superimposed as green circles) and the background noise was very low. There was an improvement upon choosing the less twinned native-B and employing correction, but the two sites of this derivative would have also been detected had the twinning been completely ignored. The longer soak with $HgAc_2$ produced two additional weak mercury sites, but also a considerably noisier difference Patterson map (Fig. 4*b*). With uncorrected data and the highly twinned native-A, the prominent peaks in the map do not represent the peaks expected from the four known mercury sites. Choosing native-B and employing twin correction more clearly brought up at least some of the expected peaks. Owing to the very noisy maps, the $HgAc_2$ derivative would probably not have been recognized as a useful derivative if twinning had been overlooked and the data had not been corrected.

In order to test whether state-of-the art programs for heavy-atom site identification would have retrieved all sites in EtHgCl and $HgAc_2$ even without twin correction, we tested *SHELXD* with the twin-corrected and the original twinned data sets. *SHELXD* identifies heavy-atom sites using dual-space direct methods (Sheldrick *et al.*, 2001; Schneider & Sheldrick, 2002). Firstly, we retrospectively identified all heavy-atom sites by inspection of $F_o - F_c$ maps which were calculated using the final model of Rna1p rigid-body refined against the derivative structure factors (*CNS* in TWIN mode, see §3.5). For EtHgCl, this revealed the two sites at Cys93 of both Rna1p molecules, which *SOLVE* had also identified correctly. For $HgAc_2$, the two sites at Cys93 and the two at the N-termini used for MIR phasing were confirmed. However, the $F_o - F_c$ maps of $HgAc_2$ showed two further more weakly occupied sites at both Cys166 residues and two even weaker sites at His*B*126 and Cys*B*215. Using twin-corrected data, *SHELXD* clearly retrieved all the heavy-atom sites [correlation coefficients between E_c calculated from the substructure and E_o from the data of 28.4 and 17.4% for CC and CC(weak) values, respectively, and data to 3.9 Å]. When twinning was ignored and the original not twin-corrected data sets native-A

and a weak derivative such as $HgAc_2$ were analysed, *SHELXD* produced heavy-atom substructure solutions which were incomplete and much more noisy [CC and CC(weak) values of only 19.7 and 9.6%, respectively].

Assuming that the sites would have been identified successfully without twin correction, either from difference Patterson maps or using dedicated software such as *SHELXD*, we analyzed what influence the correction

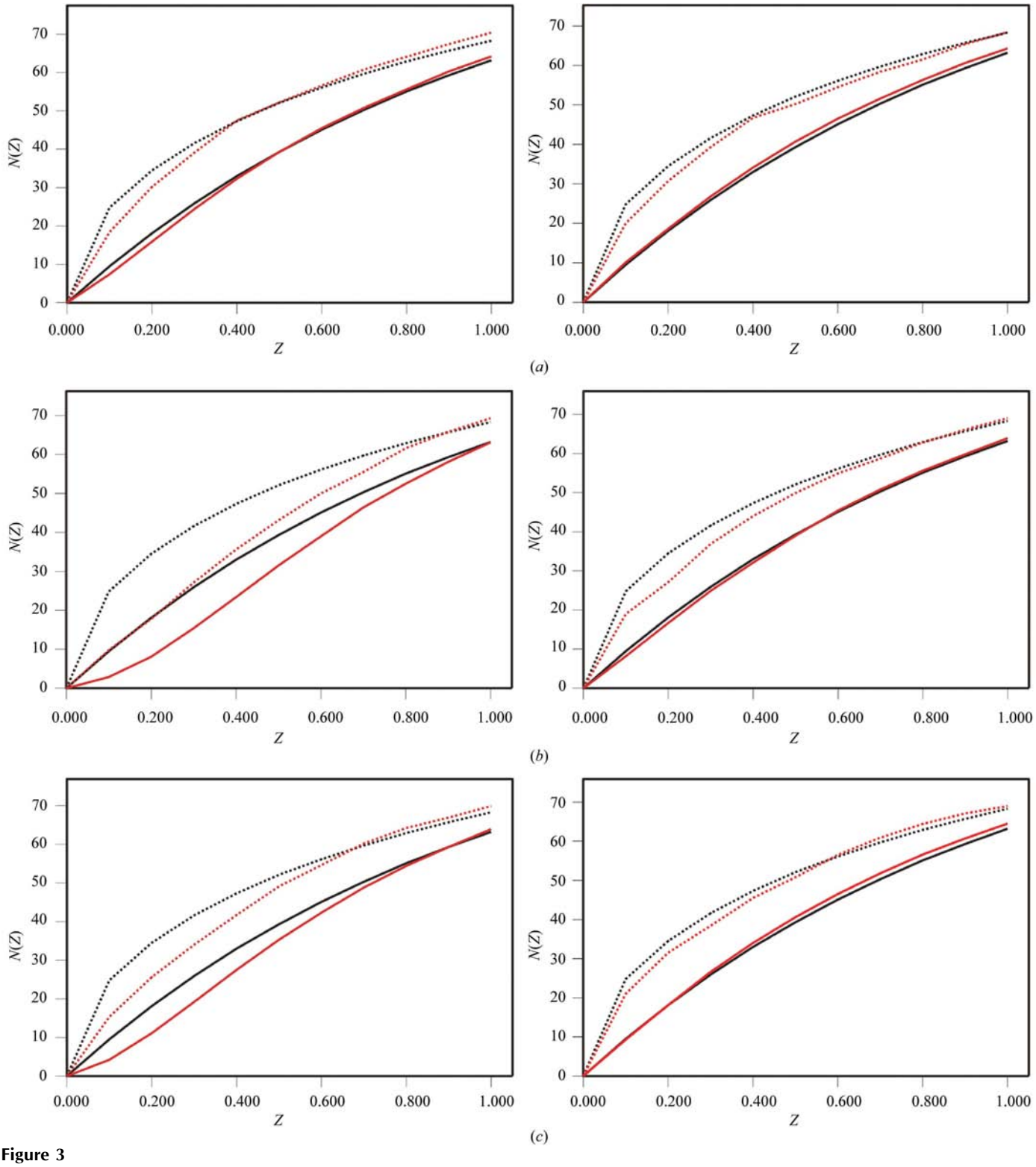


Figure 3
Cumulative intensity distributions of (*a*) native-B, (*b*) EtHgCl and (*c*) $HgAc_2$ before (left) and after (right) twin correction, illustrating the success of the twin-correction procedure (theoretically expected curves are shown as a black line for acentric and a dotted black line for centric reflections; experimental distributions are shown as a red line for acentric and a dotted red line for centric reflections).

had on site refinement and phasing. We determined the refined occupancies and *B* factors as well as the resulting phasing power for the four sites of the $HgAc_2$ derivative and the two sites of the EtHgCl derivative under three scenarios. Firstly, the highly twinned native-A was used as native data set and neither native nor derivative data were twin-corrected, representing the worst-case scenario where twinning is not noticed and by coincidence a highly twinned native data set has been collected. Secondly, the low-twinned native-B was used, but still no twin correction was introduced. This represents the case when twinning has been identified and as a remedy screening for a less twinned crystal was successful. Thirdly, the low-twinned native-B was used and additionally both native and derivative data were twin-corrected. The results (Table 3) clearly indicate that the choice of a less twinned native data set increased the occupancies and reduced the *B* factors of the heavy-atom sites and resulted in increased phasing power. This positive trend was further continued when both data sets were additionally twin-corrected.

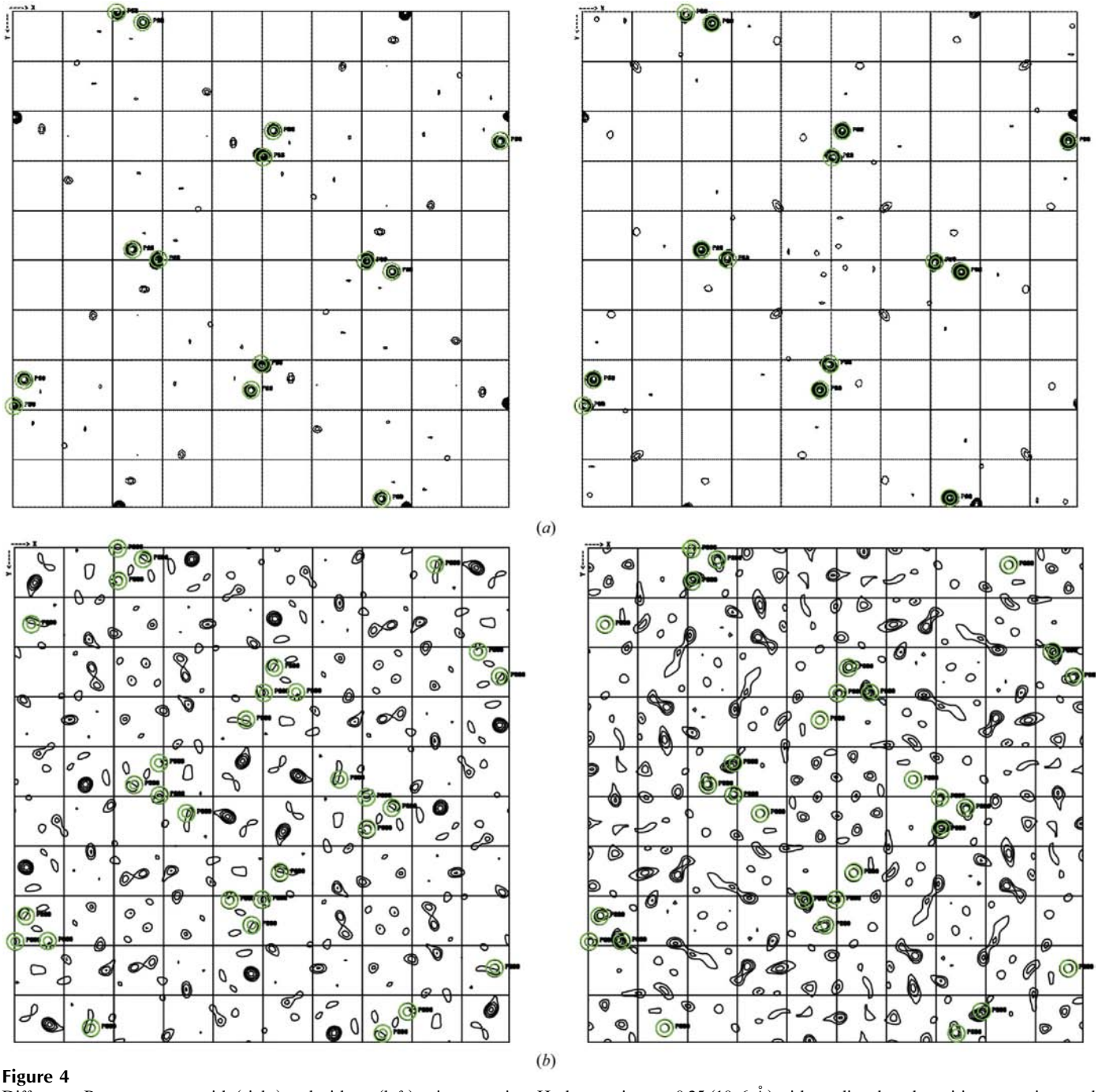


Figure 4
Difference Patterson maps with (right) and without (left) twin correction. Harker section $z = 0.25$ (10–6 Å) with predicted peak positions superimposed as green circles. Twin fractions were 0.24 for EtHgCl, 0.17 for $HgAc_2$, 0.34 for native-A and 0.06 for native-B. (*a*) for EtHgCl (contour level 2σ, step 0.5σ). (*b*) for $HgAc_2$ (contour level 1σ, step 0.5σ).

Table 3
Comparison of phasing statistics with and without twin correction.

Derivative	$HgAc_2$			EtHgCl		
Twin fraction α†	0.18			0.24		
No. of sites	4			2		
Native data set used for phasing (twin fraction†)	Native-A (0.34)	Native-B (0.06)	Native-B (0.06)	Native-A (0.34)	Native-B (0.06)	Native-B (0.06)
Twin correction	No	No	Yes	No	No	Yes
Occupancy of sites (%)	27/26/21/16	27/27/22/19	40/39/31/28	23/23	28/27	41/39
B factors of sites (Å^2)	56/67/91/96	40/45/60/75	23/30/50/60	44/46	43/45	25/32
FOM	0.25	0.34	0.36	0.31	0.33	0.34
$R_{\rm cullis}$ (acentric/centric)	0.87/0.82	0.77/0.73	0.73/0.70	0.81/0.74	0.78/0.71	0.63/0.59
Phasing power (acentric/centric)	0.87/0.68	1.34/1.04	1.49/1.14	1.12/0.87	1.30/0.96	1.96/1.49

† As determined by the twin-correction program described in §2.3.

In order to verify whether these improvements ultimately resulted in improved experimental electron-density maps, MIR maps were calculated based on phases from the EtHgCl and the $HgAc_2$ derivatives. This was again performed using native-A ($\alpha = 0.34$) without twin correction, native-B ($\alpha = 0.06$) without twin correction or native-B with additional twin correction of both the derivatives and the native data set. Fig. 5 shows a representative section of the electron-density maps, together with the refined low-resolution structure of Rna1p (PDB code 1yrg). The choice of a less twinned native data set already drastically reduced the ghost peaks in the map. The map quality was then further increased when all data sets were additionally twin-corrected. However, the most significant map improvement came from the use of a less twinned native. It would probably have been possible to build an initial model based on the MIR maps just by screening for a less-twinned native and without twin-correcting any of the data.

3.4. Refinement against a highly twinned native data set

Native-A was collected to higher resolution than native-B (2.2 Å compared with 2.7 Å), but was not used for refinement in the original structure determination because of its high twin fraction. We now analyzed whether this data set could still be used for refinement. In principle there are two ways to handle merohedral twinning during refinement. Either the diffraction data are twin-corrected and refinement is carried out against these 'cured' structure factors (strategy 1) or one artificially twins the calculated structure factors derived from the model and compares them with the untreated original diffraction data (strategy 2) (Pratt *et al.*, 1971; Jameson, 1982). We have shown that twin correction as needed for strategy 1 was the optimal way to obtain the best interpretable maps and an initial MIR model for Rna1p. However, it can be dangerous to use twin-corrected data during subsequent refinement. The original diffraction intensities are the central experimental quantity during refinement. Manipulation of these data during detwinning will always introduce new errors. This became obvious when for higher twinned data sets a number of intensities became negative after twin correction and had to be discarded (data not shown). This observation of negative intensities during twin correction is used for the determination of α *via* the 'Britton' plot (Britton, 1972; Fisher & Sweet, 1980). In addition, the standard deviations of the structure factors also increase upon twin correction so that the overall reliability of the data is reduced (Yeates & Fam, 1999). Therefore, strategy 2 of artificially twinning the model-derived structure factors may be the preferred way to handle twinned data in crystallographic refinement. In the original structure determination of Rna1p (Hillig *et al.*, 1999), native-B was used following strategy 1. For refinement of Rna1p against native-A following strategy 2 (using *CNS* in TWIN mode), a new test set was introduced to make sure that twin-related pairs of reflections are both either part of the test or the work set (see §2.5). This aspect was not accounted for during refinement of the published 2.7 Å model against the twin-corrected data set native-B.

For the refinement of Rna1p following strategy 2 using *CNS* in TWIN mode, the twin fractions α for both native data sets were re-calculated using the corresponding *CNS* input script (`detect_twin.inp`). For the low-twinned data set native-B, this resulted in a value of $\alpha = 0.07$, which corresponds very well to $\alpha = 0.06$ determined using our program. For native-A, our program and the *CNS* script also calculated almost identical twin fractions (0.35 in *CNS*, 0.34 with our program). *CNS* does not allow refinement of the twin fraction as an additional parameter during structure refinement. In order to compensate for this deficiency, different twin fractions were tested over a wide range to identify which one resulted in the lowest R factors. This revealed that the true twin fraction was even higher than that determined from the intensity statistics [0.41 when refinement was started with the low-resolution model (PDB code 1yrg) at 2.7 Å and converging at 0.39 when Rna1p was finally extensively rebuilt and refined to the full resolution limit of 2.2 Å]. Also, for native-A the twin correction did not result in a fully 'cured' intensity distribution using a twin fraction of either 0.34, 0.35 or 0.39 (data not shown). This suggests that very highly twinned data can no longer be fully twin-corrected and that refinement should therefore rather be carried out following strategy 2.

For a comparison of different refinement strategies with the two native data sets, positional and B-factor refinement of PDB entry 1yrg was carried out at 2.7 Å without any further manual rebuilding. For the less twinned native-B, strategy 1 resulted in an R and $R_{\rm free}$ of 0.223 and 0.276, respectively. Strategy 2 revealed that the model was already better than this conventional $R_{\rm free}$ with twin-corrected data implied. Without any manual rebuilding, this strategy resulted in a twinned R

and twinned R_{free} of 0.216 and 0.228, respectively. The detwinned R and R_{free} were 0.237 and 0.251, respectively. ('Twinned R' and 'twinned R_{free}' describe the difference between the artificially twinned model structure factors and the original untreated observed structure factors, whereas 'detwinned R' and 'detwinned R_{free}' refer to the difference between the model structure factors and the detwinned data. Both were calculated using `model_stats_twin.inp` from *TWIN-CNS*.) For this model from strategy 2, the detwinned R_{free} is 2% worse than the twinned R_{free} but still better than the conventional R_{free} of the model resulting from refinement strategy 1. This improvement clearly indicates the advantage of strategy 2 over strategy 1. Native-A was therefore subsequently refined using only strategy 2, for comparison first to 2.7 Å and subsequently (with several rounds of manual rebuilding and refinement) to the full resolution limit of 2.2 Å. Here, completely ignoring the twinning resulted in a high conventional R and R_{free} of 0.306 and 0.308, respectively, to a resolution limit of 2.7 Å. Again, this masked the good quality of the model because without any manual rebuilding, refinement using strategy 2 converged at a twinned R and R_{free} of 0.180 and 0.211, respectively, at 2.7 Å resolution, while the detwinned R and R_{free} stayed at 0.269 and 0.298, respectively. After adding the full resolution range (2.2 Å) and with manual rebuilding, the twinned R and R_{free} converged at 0.165 and 0.218, respectively (the detwinned R and R_{free} were 0.257 and 0.301, respectively).

For native-A, the detwinned R_{free} is 8% higher than the twinned R_{free}, while these values differed by only 2% for native-B. This increased discrepancy between detwinned R_{free} and twinned R_{free} for native-A further suggests that such a highly twinned crystal can no longer be accurately twin-corrected and should thus be treated according to strategy 2. The drastic drop in R_{free} also shows that despite its high twin fraction, native-A can be successfully used in refinement if the twinning is identified and the refinement strategy is adjusted accordingly.

3.5. Structure of Rna1p after twin refinement at 2.2 Å

Rna1p was rebuilt and refined to the full resolution limit of 2.2 Å using strategy 2. The final model shows a twinned R value of 0.165 (twinned R_{free} = 0.218; see also Table 1). A representative region of the electron-density map is shown in Fig. 6(*a*). Compared with the initial structure at 2.7 Å, a much larger number of water molecules was identified. In addition, four sulfate ions were found in or close to three crystal contacts, which explains the requirement for sulfate ions in the crystallization condition (200 m*M* Li_2SO_4).

Rna1p folds into 11 leucine-rich repeats (LRRs), resulting in a crescent-shaped overall structure (Fig. 6*b*). The two Rna1p molecules in the asymmetric unit were refined without NCS restraints. Fig. 6(*c*) shows a superimposition of the two molecules. The new high-resolution structure confirms the 2.7 Å model in its general features. Rna1p contains 11 LRRs, each consisting of a β-strand, a loop region, an α-helix and a loop connecting to the next LRR. The two molecules deviate slightly in the backbone conformation of LRR1 (Fig. 6*c*), which is still less well defined and probably flexible in the crystal. The C-termini of both molecules (residues 345–386) are disordered and not visible. In addition to LRR1, the region comprising residues 71–82 of LRR3 and residues 106–111 of LRR4 shows deviations between the two molecules. This implication of inherent flexibility was masked in the low-resolution structure of Rna1p owing to the NCS restraints. This second flexible region includes the unusually long loop in LRR3 which protrudes from the surface of the crescent and was initially suggested to provide a catalytic arginine residue (Arg74) to the active site of Ran (Hillig *et al.*, 1999). Subsequently, the structures of Ran-GMPPNP as well as Ran-GDP in complex with Rna1p and RanBP1 (Seewald *et al.*, 2002) revealed that Rna1p does not contribute a catalytic arginine, but rather stabilizes the switch II region of Ran and orients a catalytically crucial glutamine residue of Ran (Gln69).

A superimposition of Rna1p and the Rna1p–Ran-GDP–RanBP1 complex (Seewald *et al.*, 2002) revealed that the flexible region of LRR3/LRR4 coincides with part of the interaction interface between Rna1p and its target Ran-GTP. In the complex with Ran, both LRR3 and LRR4 have moved away from each other so that the side chain of Ran Leu43 can be accommodated between the two LRRs (Fig. 6*d*). Ran Leu43 is part of the switch I region and immediately adjacent to Thr42, a residue involved in Mg coordination in the GTP-bound form. The flexibility of LRRs 3 and 4 identified between the two molecules in the high-resolution structure of Rna1p may therefore indicate an inherent mobility designed to allow an induced fit when binding to Ran-GTP.

3.6. Packing analysis: influence of NCS on twinning

We examined the crystal packing of Rna1p in order to determine why this crystal lattice enabled hemihedral twinning to occur. In the crystal packing (left section of Fig. 7*a*) the Rna1p molecules form two superhelices along two of the four fourfold screw axes. These superhelices run in parallel to axis *c* through the complete crystal. There are intensive crystal contacts between the molecules within each superhelix. Interactions between the two superhelices are weaker. This may explain why Rna1p crystals grew as thin needles. Hemihedral twinning suggests that this same unit cell can also be incorporated in an upside-down orientation into the lattice. The unit-cell content of the second twin domain (right section of Fig. 7*a*) was generated by rotating the original cell by 180° around the diagonal between axis *a* and axis *b*. The orientation of the molecules in the two twin domains is very similar. The contents of both unit-cell orientations do not superimpose perfectly, but do approximately.

The lattice of Rna1p with dense crystal contacts within the superhelices and loose packing between the superhelices suggests that the individual twin domains in crystals of Rna1p also consist of complete superhelices in the two possible orientations. For twinning to occur, the inter-twin-domain crystal contacts between a superhelix of one twin domain and a superhelix of a neighboring second twin domain would have

to be similar to the intra-twin-domain contacts. As the superimposition between the molecules in the two twin-related unit cells is not perfect, such an inter-domain crystal contact will be slightly different from the intra-domain one. However, there are no clashes and in both cases similar surface patches of the outer convex surface of the crescent-shaped molecules are involved. This crystal contact between two convex surface patches features a flat interaction surface

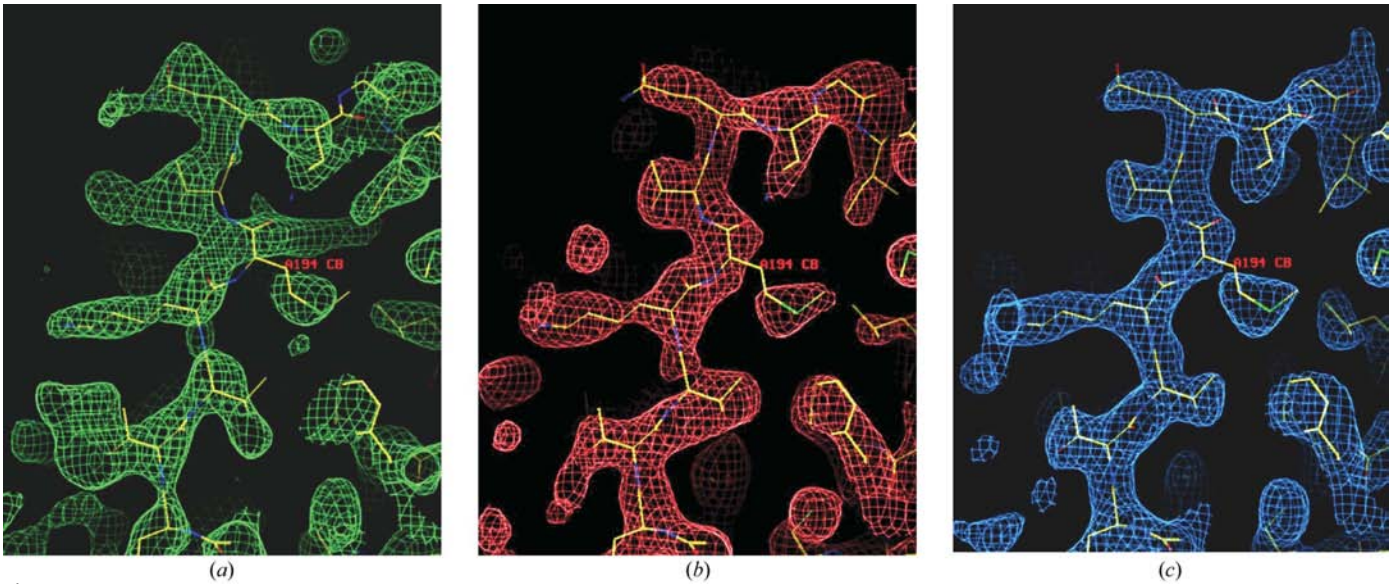


Figure 5
Influence of twin correction on MIR map quality. Map comparison with and without twin correction and while using a low-twinned or a highly twinned native data set. Shown are 2.7 Å MIR maps calculated using the derivative data sets EtHgCl (α = 0.24) and $HgAc_2$ (α = 0.17) after solvent flattening, histogram matching and twofold NCS averaging, contoured at 1.5σ. (*a*) Calculated using native-A (α = 0.34) and neither native nor derivatives twin-corrected. (*b*) Calculated using the less twinned native-B (α = 0.06), but still with neither native nor derivatives twin-corrected. (*c*) Calculated using native-B and both derivatives and native data twin-corrected. Overlaid is the refined low-resolution structure (PDB code 1yrg).

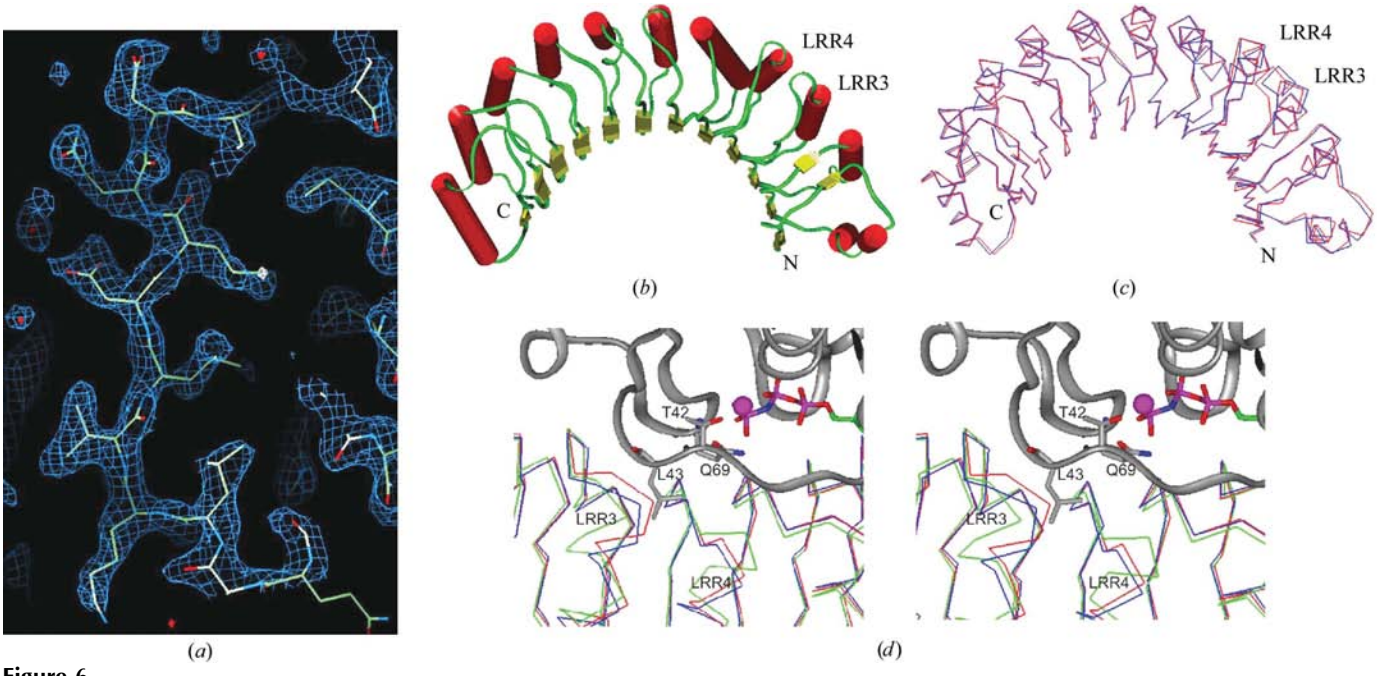


Figure 6
Structure of Rna1p refined to 2.20 Å. (*a*) Representative view of the final $3F_o - 2F_c$ electron-density map. Shown is the region around leucine-rich repeat 8 (LRR8), contoured at 1.5σ. (*b*) Ribbon representation of Rna1p. (*c*) Superimposition of the two independent molecules *A* (red) and *B* (blue) in the asymmetric unit (C^{α} backbone representation). Differences are found in the N-terminal region as well as in LRR3/LRR4. (*d*) Superimposition of Rna1p (molecules *A* and *B* in red and blue, respectively) and the complex of Rna1p (green) with Ran-GMPPNP–Mg–RanBP1 (PDB code 1k5d). The Ran backbone is shown as a grey ribbon, GMPPNP in stick representation and Mg as magenta-coloured sphere. An enlargement of the region of the flexible loops of LRR3/LRR4 is shown. This region, which differs between molecules *A* and *B* in the high-resolution structure of Rna1p, coincides with part of the interface between Rna1p and Ran-GMPPNP. The flexibility in Rna1p may indicate an inherent mobility designed to allow an induced fit.

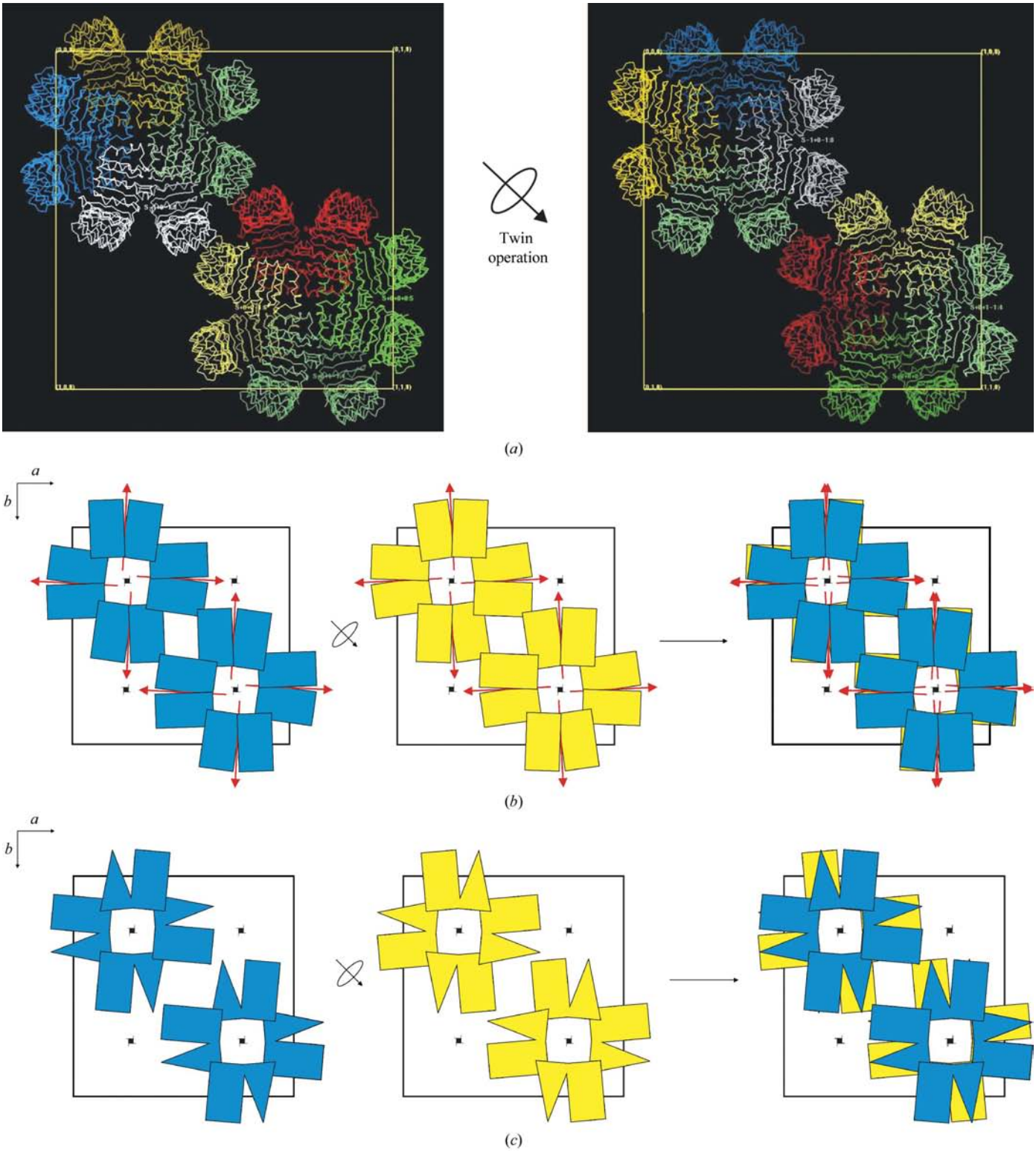


Figure 7
Crystal packing and comparison of the two twin-related orientations of the unit cell. (*a*) shows the crystal packing, looking along the *c* axis. The eight NCS-related Rna1p dimers in the unit cell are depicted in different colours. The left half shows the unit cell of twin domain 1 and the right half that of twin domain 2. Both are related by a 180° rotation around the diagonal between axes a and *b*. (*b*) Schematic representation of the packing (view along *c*) with the Rna1p molecules shown as rectangles. The twofold NCS axis (red arrow) is almost parallel to axes *a* and *b* (3° offset). The unit-cell orientations of twin domains 1 and 2 (blue and yellow) and a superimposition of both are shown. Owing to the symmetry introduced by the NCS, the inter-domain crystal packing interactions are very similar to the intra-domain interactions. (*c*) The same schematic representation as in (*b*), but with one (asymmetric) molecule instead of two identical NCS-related molecules in the asymmetric unit. Here, the inter-domain interactions in the *ab* plane differ drastically from the intra-domain interactions.

without deeper cavities. As Rna1p crystals grew in a twinned way, the changes were small enough to be accommodated in the crystal lattice when a new unit cell was added to a growing crystal in an upside-down orientation.

The schematic representation of this packing in Figs. 7(*b*) and 7(*c*) reveals the importance of the twofold NCS axis for the occurrence of hemihedral twinning with Rna1p. The inter-domain crystal packing could only be that similar to the intra-domain packing because of the presence of the twofold NCS axis almost parallel to the twin operator (Fig. 7*b*). If the asymmetric unit contained one asymmetric molecule instead of two NCS-related identical molecules (Fig. 7*c*) or if the NCS axis was not highly parallel to the twin operator twofold axis, the inter-domain interactions would differ drastically from the intra-domain crystal contacts and twinned crystal growth would not be possible. A combination of NCS almost parallel to the twin operator has been reported remarkably often for structures from merohedrally twinned protein crystals (Rees & Lipscomb, 1980; Forst *et al.*, 1998; Hillig *et al.*, 1999; Frazão *et al.*, 1999; Declercq & Evrard, 2001; Larsen *et al.*, 2002; Xu *et al.*, 2003; Lee *et al.*, 2003; Barends & Dijkstra, 2003; Yuan *et al.*, 2003; Golinelli-Pimpaneau, 2005). For crystals of Rna1p, we now see that this NCS nearly parallel to the twin operator was a necessary prerequisite for twinning to happen.

4. Conclusions

The presented case study shows that MIR phasing of Rna1p was not hindered by hemihedral twinning. Heavy-atom sites of a strong derivative data set such as EtHgCl could be detected *via* Patterson techniques even without twin correction. The sites in less strong derivatives like the $HgAc_2$ data set required state-of-the-art software such as *SHELXD* (Schneider & Sheldrick, 2002) to be at least partially identified if twinning had been ignored.

Identification of a lower twinned native crystal and subsequent twin correction of both native and derivative data sets improved heavy-atom site identification in the weaker derivatives, MIR phasing statistics and map quality. Although the twin correction is expected to introduce new errors, the twin-corrected data were still more accurate for heavy-atom site identification and phase calculation than the twinned data. Similar to our results, Igarashi and coworkers and Forst and coworkers have reported MIR structure determinations from merohedrally twinned crystals where twin correction provided cleaner difference Patterson functions and was necessary for their interpretation (Igarashi, Moriyama, Mikama *et al.*, 1997; Forst *et al.*, 1998). Together, this suggests that corrected data should be used at the initial stage of an MIR structure determination to identify all useful derivatives, to efficiently refine the heavy-atom binding sites and to obtain the best possible maps to build an initial model.

Once such a model is available, the original non-twin-corrected native data should be used for refinement. Here, the twinning is taken into account by artificially twinning the calculated model structure factors (as implemented, for example, in *CNS-TWIN* or *SHELXL*). Following this strategy, even a highly twinned native data set ($\alpha = 0.39$) was successfully used for refinement and resulted in a more detailed model. If twinning had been ignored at this point, refinement would have stalled at $R_{\rm free}$ values around 0.30, whereas identifying and accounting for the twinning revealed that without further manual rebuilding the same model refined to a twinned $R_{\rm free}$ of 0.21. It would be desirable to implement a similar approach in heavy-atom phasing programs. At present, these cannot yet take the effects of twinning into account, despite the fact that work on the theory and test cases of phasing with twinned MIR data was published many years ago (Yeates & Rees, 1987).

The presented project suffered in the beginning from difficulties in space-group assignment caused by the additional symmetry introduced by the twinning. Hemihedral twinning was identified *via* inspection of the cumulative intensity distribution plots. Careful analysis of the self-rotation functions helped to subsequently assign the correct space group. After space-group assignment and twin correction, MIR structure determination proceeded smoothly and, with the twinning taken into account during refinement, even data from a highly twinned native crystal could now be exploited to arrive at an improved higher resolution model for Rna1p. When working in space groups that allow merohedral twinning or with unit cells that allow pseudo-merohedral twinning, intensity distribution plots should therefore be checked routinely for symptoms of this kind of twinning.

5. PDB deposition

The original (not twin-corrected) structure factors of data set native-A used in refinement and the coordinates of Rna1p refined to 2.2 Å resolution have been deposited in the PDB (code 2ca6). In addition, the structure factors of data set native-B as well as of all derivative data sets listed in Table 1 have also been submitted to the PDB (same code; original data, not twin-corrected).

The authors would like to thank Alfred Wittinghofer and Jörg Becker for support and encouragement, Magarethe Schulte-Spechtel and Carolin Koerner for excellent technical assistance, Balaji Prakash for helpful discussions and Ingrid Vetter, Nicolas Nassar and the beamline scientists at EMBL Outstation Hamburg and at EMBL Outstation Grenoble for help with data collection. Also, we would like to thank Andy Thompson for initial advice on twinning and DESY Hamburg and ESRF Grenoble for access to synchrotron beam time. Data collection at ESRF was supported by the European Community Access to Research Infrastructures Action of the Improving Human Potential Programme.

References

Abrescia, N. G. & Subirana, J. A. (2002). *Acta Cryst.* D**58**, 2205–2208.
Ban, N., Nissen, P., Hansen, J., Capel, M., Moore, P. B. & Steitz, T. A. (1999). *Nature (London)*, **26**, 841–847.

Barends, T. R., de Jong, R. M., van Straaten, K. E., Thunnissen, A. M. & Dijkstra, B. W. (2005). *Acta Cryst.* D**61**, 613–621.
Barends, T. R. & Dijkstra, B. W. (2003). *Acta Cryst.* D**59**, 2237–2241.
Buerger, M. J. (1969). *Crystal Structure Analysis.* New York: John Wiley.
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.
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.
Chandra, N., Acharya, K. R. & Moody, P. C. E. (1999). *Acta Cryst.* D**55**, 1750–1758.
Ciciarello, M. & Lavia, P. (2005). *EMBO Rep.* **6**, 714–716.
Collaborative Computational Project, 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.
Cox, J. D., Kim, N. N., Traish, A. M. & Christianson, D. W. (1999). *Nature Struct. Biol.* **6**, 1043–1047.
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.
Declercq, J. P. & Evrard, C. (2001). *Acta Cryst.* D**57**, 1829–1835.
Fisher, R. G. & Sweet, R. M. (1980). *Acta Cryst.* A**36**, 755–760.
Forst, D., Welte, W., Wacker, T. & Diederichs, K. (1998). *Nature Struct. Biol.* **5**, 37–46.
Frazão, C., Sieker, L., Coelho, R., Morais, J., Pacheco, I., Chen, L., LeGall, J., Dauter, Z., Wilson, K. & Carrondo, M. A. (1999). *Acta Cryst.* D**55**, 1465–1467.
Goldman, A., Ollis, D. L. & Steitz, T. A. (1987). *J. Mol. Biol.* **194**, 143–153.
Golinelli-Pimpaneau, B. (2005). *Acta Cryst.* D**61**, 472–476.
Govindasamy, L., Reutzel, R., Agbandje-McKenna, M. & McKenna, R. (2004). *Acta Cryst.* D**60**, 1040–1047.
Hamburger, A. E., West, A. P. & Bjorkman, P. J. (2004). *Structure*, **12**, 1925–1935.
Harris, P., Poulsen, J. C., Jensen, K. F. & Larsen, S. (2002). *J. Mol. Biol.* **318**, 1019–1029.
Herbst-Irmer, R. & Sheldrick, G. M. (1998). *Acta Cryst.* B**54**, 443–449.
Hillig, R. C., Renault, L., Vetter, I. R., Drell, T., Wittinghofer, A. & Becker, J. (1999). *Mol. Cell*, **3**, 781–791.
Houborg, K., Harris, P., Petersen, J., Rowland, P., Poulsen, J. C., Schneider, P., Vind, J. & Larsen, S. (2003). *Acta Cryst.* D**59**, 989–996.
Igarashi, N., Moriyama, H., Fujiwara, T., Fukumori, Y. & Tanaka, N. (1997). *Nature Struct. Biol.* **4**, 276–284.
Igarashi, N., Moriyama, H., Mikami, T. & Tanaka, N. (1997). *J. Appl. Cryst.* **30**, 362–367.
Ito, N., Komiyama, N. H. & Fermi, G., (1995). *J. Mol. Biol.* **28**, 648–658.
Jameson, G. B. (1982). *Acta Cryst.* A**38**, 817–820.
Jones, T. A. & Kjeldgaard, M. (1997). *Methods Enzymol.* **277**, 173–208.
Kabsch, W. (1993). *J. Appl. Cryst.* **26**, 795–800.
Ko, T.-P., Kuznetsov, Y. G., Malkin, A. J., Day, J. & McPherson, A. (2001). *Acta Cryst.* D**57**, 829–839.
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.
Lebedev, A. A., Vagin, A. A. & Murshudov, G. N. (2006). *Acta Cryst.* D**62**, 83–95.
Lee, S., Sawaya, M. R. & Eisenberg, D. (2003). *Acta Cryst.* D**59**, 2191–2199.
Lehtiö, L., Fabrichniy, I., Hansen, T., Schönheit, P. & Goldman, A. (2005). *Acta Cryst.* D**61**, 350–354.
Luecke, H., Richter, H. T. & Lanyi, J. K. (1998). *Science*, **280**, 1934–1937.
Mueller, U., Muller, Y. A., Herbst-Irmer, R., Sprinzl, M. & Heinemann, U. (1999). *Acta Cryst.* D**55**, 1405–1413.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Padilla, J. E. & Yeates, T. O. (2003). *Acta Cryst.* D**59**, 1124–1130.
Parsons, S. (2003). *Acta Cryst.* D**59**, 1995–2003.
Pratt, C. S., Coyle, B. A. & Ibers, J. A. (1971). *J. Chem. Soc.*, pp. 2146–2151.
Pemberton, L. F. & Paschal, B. M. (2005). *Traffic*, **6**, 187–198.
Ramadan, M. A., Shrive, A. K., Holden, D., Myles, D. A., Volanakis, J. E., DeLucas, L. J. & Greenhough, T. J. (2002). *Acta Cryst.* D**58**, 992–1001.
Redinbo, M. R. & Yeates, T. O. (1993). *Acta Cryst.* D**49**, 375–380.
Rees, D. C. & Lipscomb, W. N. (1980). *Proc. Natl Acad. Sci. USA*, **77**, 277–280.
Reynolds, R. A., Remington, S. J., Weaver, L. H., Fisher, R. G., Anderson, W. F., Ammon, H. L. & Matthews, B. W. (1985). *Acta Cryst.* B**41**, 139–147.
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.
Rudolph, M. G., Wingren, C., Crowley, M. P., Chien, Y. H. & Wilson, I. A. (2004). *Acta Cryst.* D**60**, 656–664.
Sheldrick, G. M., Hauptman, H. A., Weeks, C. M., Miller, R. & Usón, I. (2001). *International Tables for Crystallography*, Vol. *F*, edited by E. Arnold & M. G. Rossmann, pp. 333–351. Dordrecht: Kluwer Academic Publishers.
Schneider, T. R., Karcher, J., Pohl, E., Lubini, P. & Sheldrick, G. M. (2000). *Acta Cryst.* D**56**, 705–713.
Schneider, T. R. & Sheldrick, G. M. (2002). *Acta Cryst.* D**58**, 1772–1779.
Schuermann, J. P., Prewitt, S. P., Davies, C., Deutscher, S. L. & Tanner, J. J. (2005). *J. Mol. Biol.* **347**, 965–78.
Shimamura, T., Shamotienko, O., Akhtar, S., Dolly, J. O. & Iwata, S. (2004). *Acta Cryst.* D**60**, 912–914.
Seewald, M. J., Korner, C., Wittinghofer, A. & Vetter, I. R. (2002). *Nature (London)*, **415**, 662–666.
Song, H., Hanlon, N., Brown, N. R., Noble, M. E., Johnson, L. N. & Barford, D. (2001). *Mol. Cell*, **7**, 615–626.
Stanley, E. (1972). *J. Appl. Cryst.* **5**, 191–194.
Taylor, H. O. & Leslie, A. G. W. (1998). *CCP4 Newsl. Protein Crystallogr.* **35**, 9.
Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.
Terwisscha van Scheltinga, A. C., Valegard, K., Hajdu, J. & Andersson, I. (2003). *Acta Cryst.* D**59**, 2017–2022.
Terwisscha van Scheltinga, A. C., Valegard, K., Ramaswamy, S., Hajdu, J. & Andersson, I. (2001). *Acta Cryst.* D**57**, 1776–85.
Valegard, K., Terwisscha van Scheltinga, A. C., Lloyd, M. D., Hara, T., Ramaswamy, S., Perrakis, A., Thompson, A., Lee, H. J., Baldwin, J. E., Schofield, C. J., Hajdu, J. & Andersson, I. (1998). *Nature (London)*, **394**, 805–809.
Weber, C. H., Park, Y. S., Sanker, S., Kent, C. & Ludwig, M. L. (1999). *Structure*, **7**, 1113–1124.
Wilson, A. J. C. (1949). *Acta Cryst.* **2**, 318–321.
Xu, S., Gu, L., Wang, Q., Shu, Y., Song, S. & Lin, Z. (2003). *Acta Cryst.* D**59**, 1574–1581.
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*, **7**, R25–R29.
Yeates, T. O & Rees, D. C. (1987). *Acta Cryst.* A**43**, 30–36.
Yuan, Y. R., Martsinkevich, O. & Hunt, J. F. (2003). *Acta Cryst.* D**59**, 225–238.