Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Disorder and twin refinement of RNA heptamer double helices

Uwe Mueller,[a] Yves A. Muller,[a] Regine Herbst-Irmer,[b] Mathias Sprinzl[b] and Udo Heinemann[a,d]*

[a]Forschungsgruppe Kristallographie, Max-Delbrück-Centrum für Molekulare Medizin, Robert-Rössle-Strasse 10, D-13092 Berlin, Germany, [b]Institut für Anorganische Chemie, Georg-August-Universität Göttingen, Tammannstrasse. 4, D-37077 Göttingen, Germany, [c]Laboratorium für Biochemie, Universität Bayreuth, Universitätsstrasse 30, D-95440 Bayreuth, Germany, and [d]Institut für Kristallographie, Freie Universität, Takustrasse 6, D-14195 Berlin, Germany

Correspondence e-mail: heinemann@mdc-berlin.de

Received 6 April 1999
Accepted 28 May 1999

NDB References: tALA^{wt}, ar0009; tALA^{C70}, ar0010; dALA^{wt}, ar0013; tALA^{I}, ar0014.

An RNA helix with seven base pairs which was derived from the acceptor stem of *Escherichia coli* tRNA^{Ala}, rGGGGCUA·rUAGCUCC (ALA^{wt}), as well as a variant, rGGGGCUA·rUAGCCCC (ALA^{C70}), in which the single G·U wobble base pair of ALA^{wt} was replaced by G·C, crystallize in space group *C*2. Both non-isomorphic crystal forms display a complex packing pattern, which can be described alternatively as disorder or pseudo-merohedral twinning. The structure of ALA^{wt} was determined by SIRAS phasing using an isomorphous iodine derivative, rGGGGCi^{5}UA·rUAGCUCC (ALA^{I}). All three RNA structures were subsequently subjected to twin refinement in space group *P*1, using anisotropic thermal displacement parameters at resolutions of 1.16, 1.23 and 1.4 Å for ALA^{wt}, ALA^{I} and ALA^{C70}, respectively. Alternatively, the structure of ALA^{wt} was refined in space group *C*2 assuming twofold disorder of the molecular orientation. The refined structures are of reasonable quality according to all available indicators. There are no systematic differences between the molecular models resulting from twin refinement and disorder refinement.

1. Introduction

Microhelices derived from the acceptor stem of *Escherichia coli* tRNA^{Ala} containing seven base pairs and a full A–C–C–A end are specifically aminoacylated by the cognate tRNA synthetase (Francklyn & Schimmel, 1989; Schimmel, 1993). Therefore, these small RNA fragments provide a unique opportunity to study at high resolution the structural features of an RNA double helix which are specifically recognized by a protein. The significance of these studies is underscored by *in vivo* and *in vitro* studies showing that the determinants for specific binding by alanyl-tRNA synthetase are located primarily in the acceptor stem of tRNA^{Ala} (McClain & Foss, 1988; Hou & Schimmel, 1988) and that the G3·U70 wobble base pair is of prime importance for the recognition process (Musier-Forsyth *et al.*, 1991, 1995; Beuning *et al.*, 1997).

Two independent NMR studies of acceptor stem-derived microhelices (Limmer *et al.*, 1996; Ramos & Varani, 1997) have attempted to identify structural irregularities which might serve as identity elements for tRNA^{Ala}. More recently, we have determined the crystal structures of the synthetic acceptor stem which lacks the A–C–C–A end, rGGGGCUA·rUAGCUCC (ALA^{wt}) and its G3·C70 variant, rGGGGCUA·rUAGCCCC (ALA^{C70}), at near-atomic resolutions of 1.16 and 1.4 Å, respectively (Mueller *et al.*, 1999). This analysis showed that the G3·U70 wobble base pair does not induce a significant global conformational change in the acceptor helix.

The crystal structures of ALA^{wt} and ALA^{C70} were refined as pseudo-merohedral twins in space group *P*1 (Mueller *et al.*, 1999). Pseudo-merohedral twinning may occur in crystals with unusual unit-cell geometry which produces a lattice with a pseudosymmetry higher than the point-group symmetry (Yeates, 1997), *e.g.* a triclinic cell with two edges and two angles being approximately equal. Merohedral twinning and twin refinement is not uncommon in protein structures (Carr *et al.*, 1996; Luecke *et al.*, 1998; Valegard *et al.*, 1998) but, to the best of our knowledge, has never been described for crystal structures of nucleic acid fragments where, however, cases of disorder have been reported (*e.g.* Brennan *et al.*, 1986; Shah & Brünger, 1999).

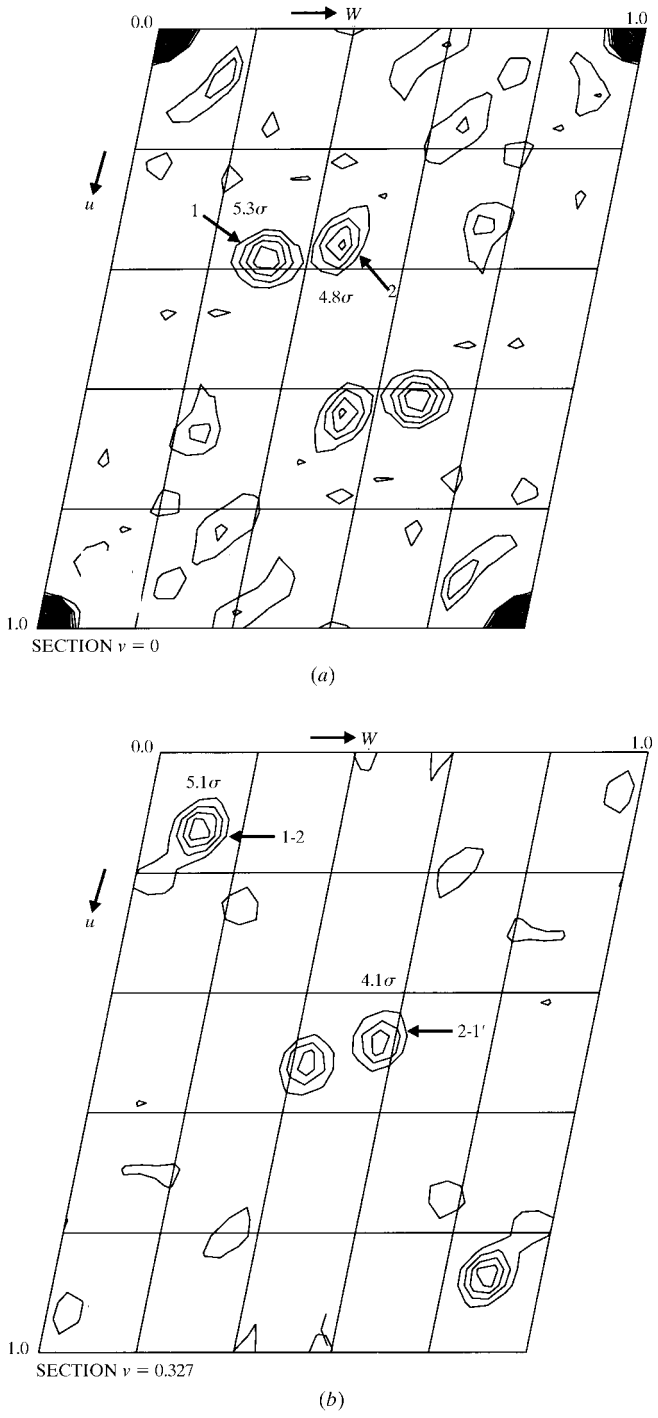

Figure 1
Harker section $v = 0$ and section $v = 0.327$ of the isomorphous difference Patterson map calculated from ALA^{wt} and ALA^{I} in space group *C*2. The two Harker peaks at 5.3σ and 4.8σ and the cross peaks result from the I atoms with fractional coordinates (−0.192, 0.000, −0.159) and (−0.320, −0.326, 0.272), respectively.

Table 1
Diffraction data in space group *C*2, phasing and disorder refinement statistics.

	ALA^{wt}	ALA^{I}
Diffraction data		
Indexed space group	*C*2	*C*2
Unit-cell dimensions (Å)	32.9, 47.5, 26.2	32.9, 47.9, 26.3
Unit-cell angles (°)	90, 101.9, 90	90, 101.6, 90
Radiation source	DESY X11	ESRF BM14
Wavelength (Å)	0.907	1.569
Resolution range (Å)	15–1.16	15–2.19
Unique observations	13258	1925
Completeness† (%)	96.7	93.0
R_{sym}†‡ (%)	6.9/56.4	4.9/7.4
Phasing statistics		
Resolution range (Å)		10.0–3.0
R_{deriv}§ (%)		25.1
R_{Cullis}¶, acentric (%)		57.0
Phasing power††, acentric		2.3
Mean FOM‡‡, acentric (%)		0.736
Refinement statistics		
Final R/R_{free}§§ (%)	15.0/24.0	
RNA atoms	588	
Water O atoms	85	
R.m.s.d.¶¶ bond lengths (Å)	0.022	
R.m.s.d.¶¶ bond-angle distances (Å)	0.017	
R.m.s.d.¶¶ base planes (Å)	0.017	
Mean positional e.s.d.††† (Å)	0.039	

† Resolution shells are 15–1.16/1.17–1.16 Å for ALA^{wt} and 24–2.19/2.25–2.19 Å for ALA^{I}. ‡ $R_{sym} = \sum_{h,i} |I_{h,i} - I_h| / \sum_{h,i} I_{h,i}$, where the summation is over all observations $I_{h,i}$ contributing to the reflection intensities I_h. § $R_{deriv} = \sum |F_{PH} - F_P| / \sum F_P$, where F_{PH} and F_P are the heavy-atom derivative and native structure-factor amplitudes, respectively. ¶ $R_{Cullis} = E / \sum |F_{PH} \pm F_P|$, where E is the residual lack-of-closure error. †† Phasing power = $\langle F_H \rangle / E$, where $\langle F_H \rangle$ is the average heavy-atom structure-factor amplitude. ‡‡ Figure of merit after five cycles of density modification with *DM*. §§ R_{free} (Brünger, 1992) based on 5% of the data selected from thin resolution shells with *DATAMAN* (Kleywegt & Jones, 1996). ¶¶ Root-mean-square deviation, here with respect to the RNA library obtained from http://ndb-mirror-2.rutgers.edu/NDB/archives/proj/standards.html. †††Estimated standard deviation for RNA atoms from least-squares matrix.

Here, we describe the X-ray analysis of the ALA helices by SIRAS phasing using the disordered single-site iodine derivative rGGGGCi^{5}UA·rUAGCUCC (ALA^{I}) and compare alternative refinements of the RNA structure using disorder and twin models. The physical significance of the two models and relevant aspects of crystal packing in ALA^{wt} and ALA^{C70} are discussed. Since synthetic nucleic acid fragments particularly lend themselves to the type of lattice disorder found here, this discussion may be of more general interest.

2. Materials and methods

2.1. Crystallization

ALA^{wt} and ALA^{C70} were crystallized as described elsewhere (Mueller *et al.*, 1999). The RNA heptamer rGGGGCi^{5}UA containing 5-iodouracil at sequence position 6

was synthesized, purified and hybridized with rUAGCUCC according to a published protocol (Ott *et al.*, 1996) to yield ALA^{I}. In order to prevent photolytic cleavage, the iodinated sample was stored in the dark. For crystallization using the hanging-drop vapour-diffusion method at 293 K, 1.5 µl of 0.5 m*M* ALA^{I} duplex was mixed with 1.5 µl of a crystallization buffer consisting of 40 m*M* sodium cacodylate pH 6.0, 12 m*M* spermine tetra-HCl, 80 m*M* $SrCl_2$, 40 m*M* LiCl and 10%(*v*/*v*) 2-methyl 2,4-pentanediol (MPD), exploiting the screen described by Berger *et al.* (1996). Equilibrating the droplet against a reservoir of 35%(*v*/*v*) MPD gave small crystals with dimensions of 150 × 80 × 50 µm after 48 h. Similar to the large crystals of ALA^{wt}, these crystals have a clear morphology without any indication of disorder or twinning. The high MPD concentration permitted flash-freezing of the crystals to 100 K for data collection.

2.2. Data collection

Diffraction data of ALA^{I} were collected at beamline BM14 of the ESRF (Grenoble) at 100 K on a MAR Research (Hamburg) imaging-plate detector. A first data set was collected at a wavelength of 1.569 Å in order to record anomalous differences at the very far remote energy of the L_{III} absorption edge of iodine, which is at 2.72 Å (Table 1). Therefore, two φ scans separated from each other by 180° were performed at a detector distance of 160 mm and a φ increment of 3°. Data reduction was performed with *DENZO*/*SCALEPACK* (Otwinowski & Minor, 1997). The two subsets were merged and scaled with *SCALA*, *AGROVATA* and *TRUNCATE* (Collaborative Computational Project, Number 4, 1994).

A second diffraction experiment at the same synchrotron beamline aimed at extending the resolution as far as possible. This data set was collected at a wavelength of 0.905 Å with a crystal-to-detector distance of 160 mm and a φ increment of 3°, and was supplemented with a low-resolution pass at a detector distance of 200 mm and one-third of the exposure time. Even at the resolution limit, all reflection profiles are clean, showing no indication whatsoever of splitting which might arise from non-merohedral twinning. When reducing the high-resolution data in space group *P*1 (see below), the completeness is 83.1% with an overall R_{sym} of 8.7% (resolution shell 1.24–1.23 Å, R_{sym} = 25.0%). In Table 2, the high-resolution ALA^{I} data set is compared with those for ALA^{wt} and ALA^{C70} (Mueller *et al.*, 1999).

3. Results

3.1. Phasing of ALA^{wt} and ALA^{I}

The structure of ALA^{wt} was solved in space group *C*2 with the SIRAS method, taking advantage of the isomorphous and anomalous differences produced by ALA^{I}. Previous attempts to solve the structure by molecular replacement using a canonical A-form heptamer as a search model had failed. From packing considerations, the asymmetric unit was likely to contain one duplex of heptameric RNA with a specific volume (Heinemann, 1991) of 1438 $Å^3$ per base pair. In apparent contradiction to this result, both the isomorphous as well as the anomalous difference Patterson map of ALA^{I} clearly revealed two distinct heavy-atom sites (Fig. 1). The Patterson maps could be readily interpreted with the program *HEAVY* (Terwilliger *et al.*, 1987) and refined with *MLPHARE* (Collaborative Computational Project, Number 4, 1994). Initially, only the isomorphous differences were used in refinement. The introduction of the anomalous differences as well as phase improvement through solvent flattening with the program *DM* (Collaborative Computational Project, Number 4, 1994) at 3 Å allowed for the determination of the correct hand and produced a map which, upon inspection with program *O* (Jones *et al.*, 1991), clearly revealed the base-pair ladder and the locations of the P atoms (Fig. 2).

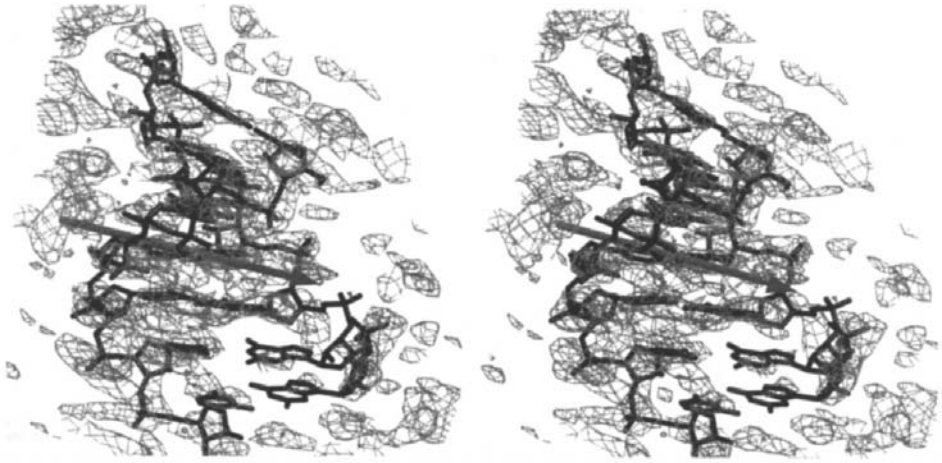

Figure 2
Portion of the solvent-flattened SIRAS electron-density map at 3 Å resolution superimposed with the final RNA model (one duplex only). Most base-pair planes are clearly revealed as are many phosphorus positions. The twofold-symmetry axis of space group *C*2 (grey arrow) passes through the central base pair, *i.e.* the duplex occupies a special position in the unit cell.

To our surprise, the map revealed that two heptamers are arranged such that within each RNA heptamer the molecular pseudo-dyad relating the two non-identical strands in the RNA duplex coincides with a crystallographic twofold axis (see Fig. 8). That is, the ALA^{wt} (and ALA^{I}) duplexes do not occupy the four general positions of space group *C*2, but instead we find two independent copies of the RNA on the two special positions. Because of the absence of an exact molecular twofold axis passing through each duplex RNA heptamer, the crystal symmetry can only be maintained if we allow for local disorder, namely that at each duplex position the RNA molecules can adopt two distinct orientations

Table 2
Diffraction data in space group *P*1 and refinement statistics.

	ALA^{wt}	ALA^{I}	ALA^{C70}
Space group	*P*1	*P*1	*P*1
Unit-cell dimensions (Å)	26.2, 28.9, 28.9	26.4, 29.0, 29.0	26.7, 26.7, 30.5
Unit-cell angles (°)	110.5, 96.9, 96.9	110.4, 96.6, 96.6	104.3, 104.3, 91.7
Radiation source	DESY X11	ESRF BM14	DESY BW7A
Wavelength (Å)	0.907	0.905	1.021
Resolution range (Å)	15–1.16	15–1.23	15–1.4
Crystal mosaicity (Å)	0.6	0.7	1.2
Unique observations	23758	18837	13834
Completeness† (%)	88.4/75.7	82.2/32.1	89.5/80.7
Multiplicity	2.7	3.4	1.6
$\langle I/\sigma(I)\rangle$†‡	12.6/1.5	13.4/2.1	11.4/2.3
R_{sym}†§(%)	9.9/54.0	5.0/30.8	5.7/28.8
Molecules per asymmetric unit	2 duplexes	2 duplexes	2 duplexes
Final R/R_{free}¶ (%)	12.4/18.0	12.7/19.3	15.8/21.8
RNA atoms	586	588	586
Water O atoms	149	155	134
R.m.s.d.†† bond lengths (Å)	0.007	0.029	0.028
R.m.s.d.†† bond-angle distances (Å)	0.019	0.018	0.010
R.m.s.d.†† base planes (Å)	0.032	0.032	0.020
Mean positional e.s.d.‡‡ (Å)	0.041	0.033	0.053

† Resolution shells 15–1.16/1.17–1.16 Å for ALA^{wt}, 15–1.23/1.24–1.23 Å for ALA^{I} and 15–1.40/1.45–1.40 Å for ALA^{C70}. ‡ *I*, reflection intensity. § $R_{sym} = \sum_{h,i} |I_{h,i} - I_h| / \sum_{h,i} I_{h,i}$, where the summation is over all observations $I_{h,i}$ contributing to the reflection intensities I_h. ¶ R_{free} (Brünger, 1992) based on 5% of the data selected from thin resolution shells with *DATAMAN* (Kleywegt & Jones, 1996). †† Root-mean-square deviation, here with respect to the RNA library obtained from http://ndb-mirror-2.rutgers.edu/NDB/archives/proj/standards.html. ‡‡ Estimated standard deviation for RNA atoms from least-squares matrix.

which are related by a crystallographic twofold axis. Each orientation has to be 50% occupied. This model is supported by the fact that we observe two distinct iodine positions, whereas packing considerations suggested a single duplex molecule with a single I atom in the asymmetric unit (see above).

3.2. Disorder refinement of ALA^{wt}

The refinement of ALA^{wt} was performed with *SHELXL*97 (Sheldrick & Schneider, 1997) with the disorder model, using the 'PART -1' option to turn off restraints on the symmetry-related contacts of the superpositioned RNA duplexes. The occupancy of the disordered duplexes was fixed at 0.5 to account for the twofold symmetry. The refinement against 12611 independent structure-factor amplitudes to 1.16 Å resolution, with subsequent modelling of 50 fully occupied solvent sites including one Sr^{2+} cation, led to a crystallographic *R* value of 22.1% (R_{free} = 27.6% for 5% of the data omitted from refinement according to Brünger, 1992). At this stage, further refinement using restrained anisotropic *B* factors was employed, with restrictions in DELU of 0.01 and SIMU of 0.1. The refinement of the disorder model of ALA^{wt} including 586 RNA atoms, 80 fully occupied and five half-occupied water sites, one Sr^{2+} and one Na^{+} converged at a crystallographic *R* factor of 14.7% (R_{free} = 24.0%).

The two RNA heptamer duplexes of ALA^{wt} superimposed by crystallographic symmetry are very clearly defined by electron density (Fig. 3). The molecules are in a standard A-type conformation with Watson–Crick pairing of all base pairs except for G3·U70, which shows the expected wobble arrangement. With the exception of the riboses of U66 of one duplex and G1 of the other, which both show the C2′-*exo* sugar pucker, all riboses are puckered C3′-*endo*, and no significant deviations from A-type helical parameters as calculated with *CURVES* (Lavery & Sklenar, 1989) are found (not shown).

3.3. Twin refinement of the ALA helices

At this stage, the possibility of a more convenient description of the crystal by twinning was investigated. Here, the two alternative orientations of the RNA duplexes would not result from local disorder but would arise from the presence of two separate twin domains with different orientations in the crystal. In the case of pseudo-merohedral twinning (Yeates, 1997; Herbst-Irmer & Sheldrick, 1998; Yeates & Fam, 1999), which is the only possibility here, the diffraction pattern of the two distinct twin domains superimposes perfectly, resulting in a diffraction pattern of higher Laue symmetry than the underlying crystal. Thus, the symmetry observed in space group *C*2 would result from the pseudo-merohedral twinning of a triclinic crystal with special restrictions in its unit-cell dimensions such that the twofold symmetry results from two distinct orientations of the latter. The twofold symmetry identifies the twin law (Yeates, 1997; Herbst-Irmer & Sheldrick, 1998). The relationship between the monoclinic *C*2 unit cell and the triclinic *P*1 cell is given in Fig. 4.

In order to investigate the twin model, all data sets were re-evaluated in space group *P*1 (see Table 2). Special care was taken to maintain a consistent indexing of the different images. Because two angles and two cell axes are identical in space group *P*1, mixing of the two indexing possibilities would generate an artificially twinned data set. The observed differences in the R_{sym} values (see Tables 1 and 2) are a consequence of the individual data-reduction protocols, using *SCALA*/*AGROVATA*/*TRUNCATE* (Collaborative Computational Project, Number 4, 1994) for the *C*2 data and *SCALEPACK* (Otwinowski & Minor, 1997) combined with *TRUNCATE* for the *P*1 data. For practical reasons, the structures of ALA^{wt} and ALA^{I} in space group *P*1 were solved by molecular replacement with *AMoRe* (Navaza, 1994), using the previously refined structure of ALA^{wt} as a search model. Since the twin refinement of the ALA^{wt} has been described briefly elsewhere (Mueller *et al.*, 1999), the following applies to the refinement of ALA^{I}.

Twin refinement of ALA^{I} was performed with *SHELXL*97 against intensities directly taken from the *SCALEPACK* evaluation, after selecting 5% of the data as test reflections within thin resolution shells using the program *SHELXPRO* from the *SHELX* package. After removing all water O atoms

from the coordinate set and resetting the displacement parameters to isotropic B values of 20 Å^2, the first refinement cycle using data from 15 to 2.0 Å led to R and R_{free} of 28.4 and 30.7%, respectively. Stepwise extension of the data to the resolution limit of 1.23 Å, together with the modelling of 69 water sites, let the isotropic refinement converge at R and R_{free} of 20.1 and 22.9%, respectively. Restrained anisotropic B-factor refinement of all atoms reduced R and R_{free} to 16.7 and 21.4%, respectively. Subsequent completion of the hydration model to include 155 water sites, one Sr^{2+} and one Na^+ let the twin refinement of ALA^{I} converge at R and R_{free} of 12.7 and 19.3%, respectively. The last cycle of refinement including the test set yielded an R of 12.9%. For comparison, the final model of ALA^{wt} consists of 586 RNA atoms, 149 waters, one Sr^{2+} and one Na^+, and the final crystallographic R factor is 12.4% (R_{free} = 18.0%).

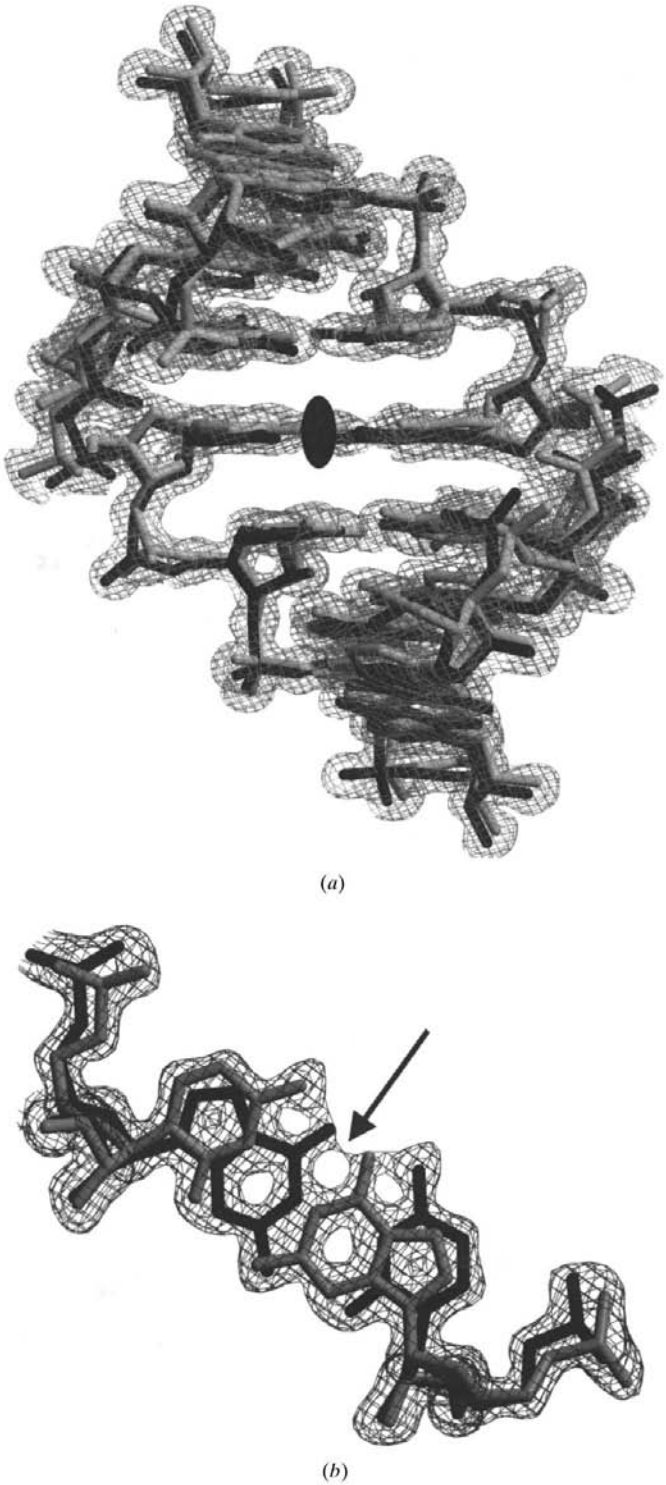

Figure 3
$2mF_o - DF_c$ difference electron density for the disorder model of ALA^{wt}. (*a*) Two copies of the RNA duplex (in black and grey) with 0.5 occupancy each are superimposed by the crystallographic dyad. Note the close match between most atoms of the sugar–phosphate backbone which is a prerequisite for this type of disorder and which is possible only in the absence of prominent sequence-dependent structural variation. (*b*) The central G·C base pair. The arrow marks the crystallographic dyad. The contour levels are 1.26σ and 1.56σ for (*a*) and (*b*), respectively.

3.4. Crystal structure of ALA^{I}

The refined structures of the two heptamer duplexes of ALA^{I} are nearly identical to those of the ALA^{wt} duplexes. After superimposing all RNA atoms in equivalent duplexes of ALA^{I} and ALA^{wt}, r.m.s.d. values of 0.169 and 0.210 Å, respectively, are obtained. Owing to different packing envir-

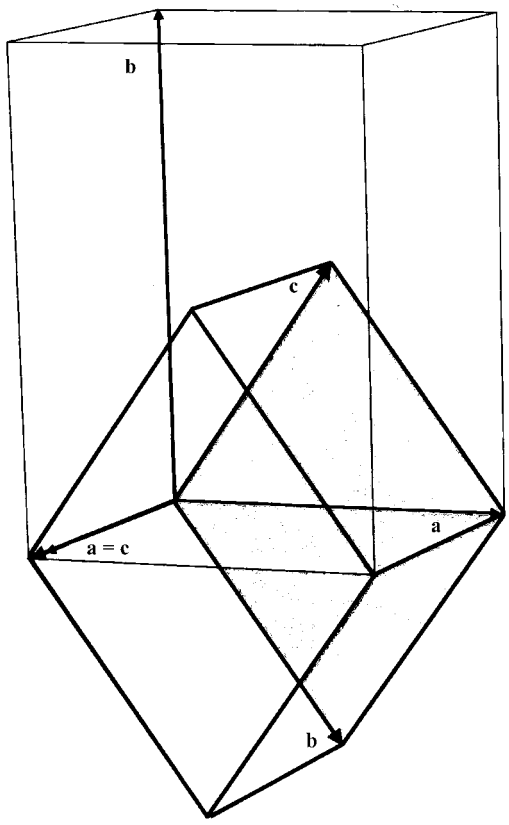

Figure 4
Relationship between the crystallographic $C2$ unit cell and the $P1$ unit cell in the cases of ALA^{wt} and ALA^{I}. The unit-cell constants are given in Tables 1 and 2. The twin law derived for space group $P1$ and expressed in reciprocal space as $I(h, k, l) \equiv I(-h, -l, -k)$ is equivalent to the crystallographic twofold axis along the b axis of the $C2$ cell. In the case of ALA^{C70}, the same relationship applies; however, the order of the $P1$ axes is permutated. Here, the monoclinic c axis is collinear with the triclinic c axis. The twin law is expressed as $I(h, k, l) \equiv I(-k, -h, -l)$.

onments, the two crystallographically independent ALA^{I} molecules differ from each other with an r.m.s.d. of 0.851 Å. In spite of this, the two duplexes show overall A-form geometry, with an average helical twist value of 32.7 or 31.5°, mean rise values of 2.85 Å for both helices, inclination values from 8.1 to 10.6° and displacement values of the base pairs between 3.6 and 4.4 Å with respect to a curved global helical axis. Additionally, all 28 riboses adopt the C3′-*endo* conformation.

Obviously, the I atom which is bound to C5 of U6 has no measurable influence on the helical geometry. The I atoms point straight into the major groove of the duplexes and are 3.6–3.8 Å away from the O2P atom of the same residue. The pyrimidine ring of the C5 residue displays a nearly complete intrastrand overlap with the I atom of U6. This interaction does not require any departure from standard A-type helix geometry.

3.5. Comparison of the disorder and twin models of ALA^{wt}

The two independent molecules of ALA^{wt} representing either the disorder model ($dALA^{wt}$-A and $dALA^{wt}$-B) or the twin model of the $tRNA^{Ala}$ microhelix ($tALA^{wt}$-A and $tALA^{wt}$-B; Mueller *et al.*, 1999) were compared in order to assess the effects of the different refinement strategies and the influence of crystal-lattice contacts on the double helical structures (Fig. 5). Least-squares superpositions of $dALA^{wt}$-A and $dALA^{wt}$-B and of $tALA^{wt}$-A and $tALA^{wt}$-B using all RNA atoms yielded r.m.s.d. values of 0.922 and 0.883 Å, respectively. Thus, the different crystal environments of the duplexes cause significant conformational differences in each refinement protocol. A comparison of the equivalent duplexes $dALA^{wt}$-A and $tALA^{wt}$-A or $dALA^{wt}$-B and $tALA^{wt}$-B yields smaller r.m.s.d. values of 0.242 or 0.234 Å, respectively. Obviously, the conformational differences arising from the different descriptions of the crystal are much smaller than those resulting from crystal packing. Considering the mean coordinate e.s.d. of 0.039 Å (see Table 2), however, these differences are still significant. They correspond to small differences in the *R* values and in the number of modelled solvent sites and may be caused by slight variations in the computational procedures employed in data processing and refinement.

Nevertheless, the atomic models are of satisfactory quality as outlined above and in Table 2 and indicated by the electron density. The G·U wobble base pair of the $tRNA^{Ala}$ acceptor stem shows a distinct hydration pattern (Fig. 6) which is indicated by clear density in all models of ALA^{wt} and ALA^{I}, whether produced by disorder or by twin refinement. Furthermore, at the resolution of 1.16 Å the density permits the unambiguous discrimination between carbon, nitrogen, oxygen and, of course, phosphorus in most parts of the ALA^{wt} double helix, as shown in Fig. 7. This figure also shows qualitatively how the electron-density levels correspond to the anisotropic displacement parameters: atoms with higher density have smaller 50% probability ellipsoids than atoms with lower density; see, for example, the pairs O2′/O3′, O4/O2 and N1/N3.

3.6. Crystal packing of ALA^{wt}, ALA^{I} and ALA^{C70}

It is remarkable that a disordered or twinned crystal lattice is observed for all ALA helices, although the packing of the ALA^{wt} and ALA^{I} duplexes in their isomorphous unit cells differs from the packing of ALA^{C70} in its non-isomorphous cell. These can be appreciated when considering the packing within the disordered *C*2 cell (Fig. 8). The crystals of ALA^{wt}, ALA^{I} and ALA^{C70} are characterized by the stacking of heptameric duplex RNA molecules forming long quasi-continuous double helices. Along the helix axes, pseudo-

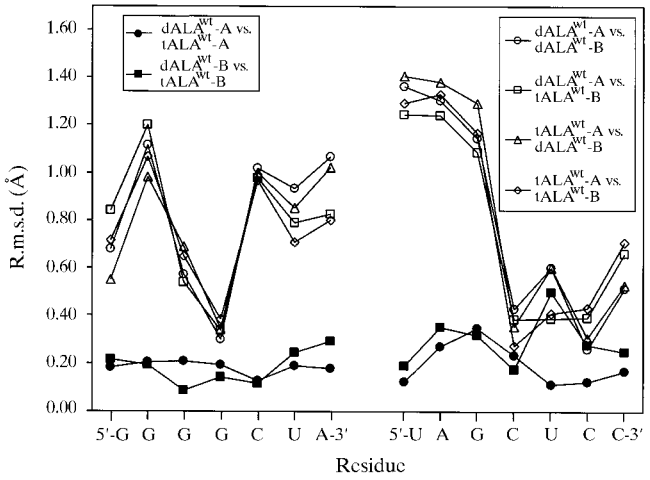

Figure 5
Root-mean-square deviation of atoms in nucleotide units of ALA^{wt} after pairwise least-squares superposition of duplex molecules. Individual ALA^{wt} duplexes are denoted as in the text. Note the close similarity of models derived from different refinement strategies and occupying equivalent positions in the unit cells (filled symbols) in comparison with models residing on different positions in the unit cells and hence experiencing different packing contacts (open symbols).

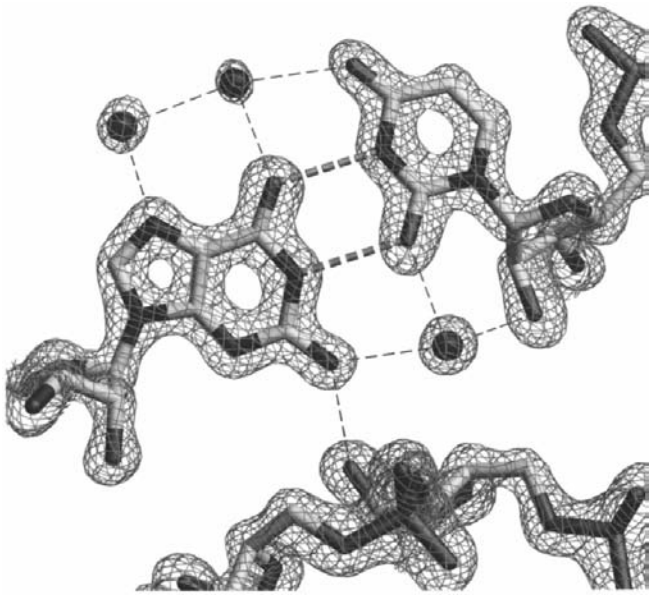

Figure 6
Wobble base pair G3·U70 of ALA^{wt} with $2mF_o - DF_c$ difference electron density contoured at 2.26σ.

twofold axes relating the two non-identical strands of the duplex alternate with twofold axes located between duplex molecules. Because these two types of twofold axes are oriented roughly perpendicular to each other, the symmetry of the quasi-continuous helices can be described as pseudo-2_122, with the 2_1 screw axis coinciding with the helix axis. Perfect 2_122 symmetry would only be observed if both strands of the RNA duplex were identical. However, for RNA heptamers this cannot be achieved without violating Watson–Crick base-pair formation.

The crystal packing of ALA^{wt} and ALA^{I} differs from that of ALA^{C70} in the arrangement of the dyad and pseudo-dyad axes with respect to the heptamer helices. In ALA^{C70}, the quasi-continuous helices are oriented such that the twofold axes which relate different duplex molecules coincide with the crystallographic twofold axes (Fig. 8*b*), but disorder (or twinning) arises from non-crystallographic dyads superimposing two copies of the duplexes which occupy general positions in the *C*2 cell. However, in the cases of ALA^{wt} and ALA^{I}, the crystallographic twofold axes coincide with the pseudo-twofold axes relating the non-identical strands within the duplex, thus requiring the description of these crystals with either a disorder or twinning model (Fig. 8*a*).

The helical repeat of the infinite RNA helices consists of two duplex molecules, thus resulting in the stacking of 14 base pairs. However, within the duplex molecules, almost ideal A-form RNA is observed (helical repeat $\simeq$ 11 bp, average twist angle $\simeq 32.7^{\circ}$). As a consequence, negative twist angles of about -16° are required for the two intermolecular base-pair stacking steps present in a single helical turn. These relatively weak stacking interactions may in part explain the propensity of heptamer duplexes for disorder/twinning. Six base-pair A-form duplexes can form more rigid quasi-continuous rods by just a slight underwinding leading to a 12 bp repeat, whereas 8 bp duplexes cannot form these rods at all and tend to pack with their terminal base pairs into the shallow groove of neighbouring helices in crystal lattices (Wahl & Sundaralingam, 1997). Our observation of further examples of disorder/twinning in RNA heptamers crystallizing in unrelated monoclinic or rhombohedral space groups (Mueller *et al.*, unpublished work) lends support to this view.

As required for local disorder, both crystal packings are characterized by the almost complete absence of sequence-specific packing contacts. Whereas in ALA^{C70} no crystal packing contacts can be observed which directly involve the bases, in ALA^{wt} and ALA^{I} two such contacts are present, one of these including the wobble base pair (see Fig. 6). However, these contacts are not expected to prevent local disorder, because identical nucleotides are present at these positions when considering the alternative orientation of the duplex.

4. Discussion

4.1. Twinning *versus* disorder model

An apparent contradiction results from the fact that the structures of ALA^{wt} and ALA^{I} were initially solved in space group *C*2 on the basis of local disorder and that at a later stage the duplex RNA heptamers were refined to convergence assuming pseudo-merohedral twinning. Although in real space both treatments result in very similar final atomic models, the different descriptions should be distinguishable in reciprocal

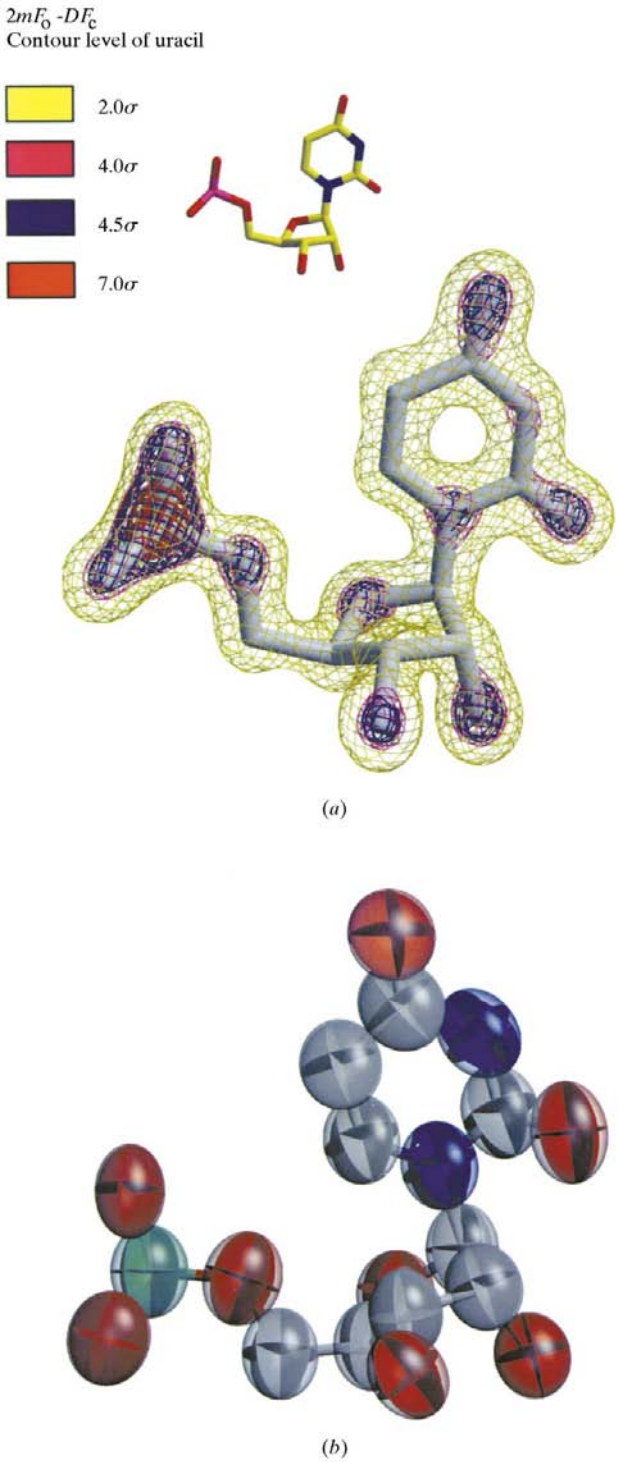

Figure 7
Nucleotide U6 from the 1.16 Å crystal structure of ALA^{wt} after twin refinement. (*a*) The $2mF_o - DF_c$ difference electron density contoured at different levels permits the unambiguous discrimination of C, N, O and P atoms. (*b*) Thermal ellipsoids for the heavy atoms of U6.

space. In a disordered lattice, all atoms contribute to the overall scattering of the crystal (although with occupancies different from unity), the resulting structure factors being the vector sum of the individual atomic structure factors. In case of pseudo-merohedral twinning, this only holds true for the individual separate twin domains. The resulting structure-factor amplitudes equal the algebraic addition of the structure-factor amplitudes from the individual twin domains. Thus, the two models generate different intensity distribution profiles, and all standard methods used in detecting merohedral twinning take advantage of this (Stanley, 1972; Rees, 1980; Yeates, 1997; Herbst-Irmer & Sheldrick, 1998).

Standard tests for ALA^{wt}, ALA^{I} and ALA^{C70} suggest that the crystals are not twinned. For ALA^{wt}, ALA^{I} and ALA^{C70}, $\langle|E^2 - 1|\rangle >> 0.736$ for acentric reflections, 0.736 being the value which should not be exceeded for twinned data (Yeates, 1997). In addition, their ratios of $\langle I^2\rangle/\langle I\rangle^2$ are 2.13, 1.98 and 2.66, respectively, for acentric reflections, again indicating the absence of twinning. The strongest argument against the presence of pseudo-merohedral twinning results from the observed isomorphous as well as anomalous Patterson maps. The Patterson maps observed can be interpreted unambiguously in space group $C2$ with two heavy-atom positions (see Fig. 1), and all expected Harker vectors and cross vectors are present. Model calculations were performed in order to determine the shape of the Patterson map assuming a pseudo-merohedral twinning model. To this end, the coordinates of the two iodine positions in the $C2$ cell of ALA^{I} were transformed into the $P1$ cell. Structure factors were calculated and twinning was introduced by averaging the amplitudes of $F(h, k, l)$ and $F(-h, -l, -k)$ according to the twin law. The resulting amplitudes were reindexed in space group $C2$ and a Patterson map was calculated. As expected, the resulting Patterson map is characterized by the absence of Harker vectors as well as the absence of some cross vectors, the reason for this being the absence of interference between the scattered waves of related I atoms in the twinning model. In the non-twinned case, these atoms are related by crystallographic symmetry, and their difference vectors, therefore, can be observed on the Harker plane. The observed Patterson map clearly suggests the absence of pseudo-merohedral twinning.

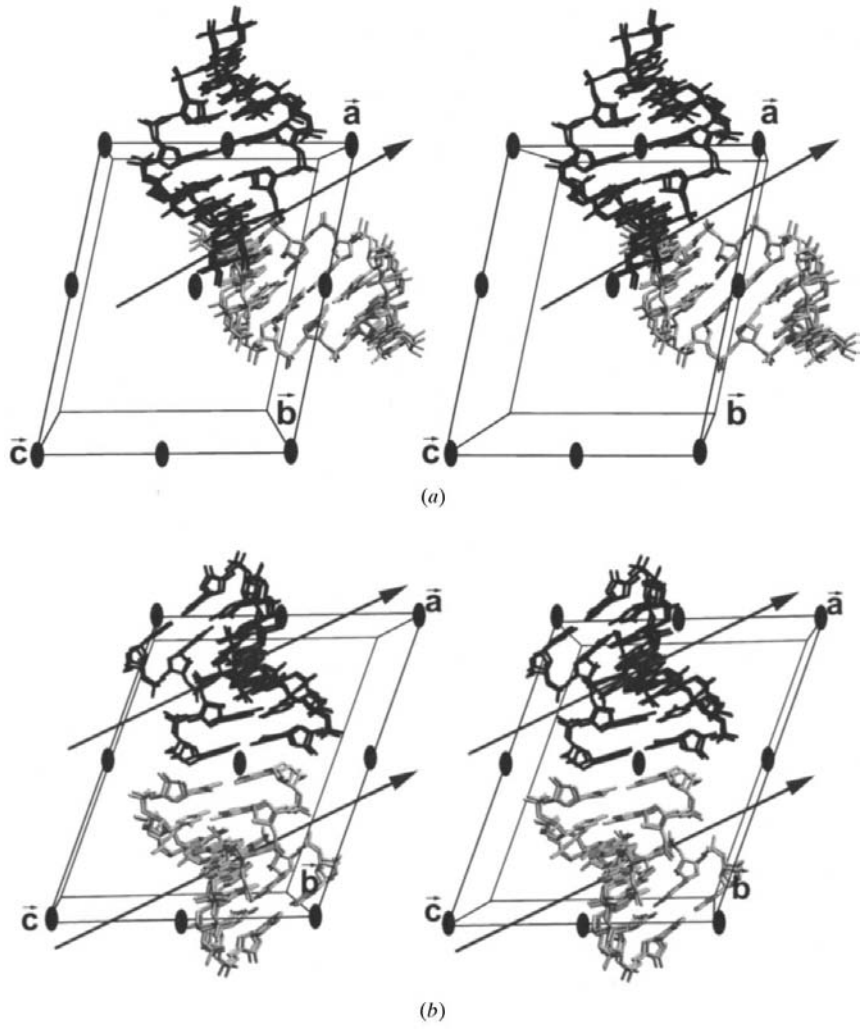

Figure 8
Crystal packing of (*a*) ALA^{wt} and ALA^{I} and (*b*) ALA^{C70} in space group $C2$. In case of ALA^{wt} and ALA^{I}, two duplex RNA heptamers are located on special positions such that their molecular pseudo-dyads which relate the two non-identical strands within a single duplex coincide with crystallographic twofold axes. The two RNA heptamers shown are related by a non-crystallographic twofold axis located in the *ac* plane (arrow). In the case of ALA^{C70}, the duplex RNA heptamer is located between two crystallographic twofold axes. Here, the crystallographic axes relate stacked RNA molecules. In this case, the molecular pseudo-dyads are non-crystallographic and located in the *ac* plane (arrows).

Nevertheless, refining the structures as pseudo-merohedral twins proved very convenient and, from the results obtained here, appears valid. The comparable R and R_{free} values obtained after twin or disorder refinement of ALA^{wt} may be a result of the molecular pseudo-symmetry of the duplexes oriented along the crystallographic twofold axis in $C2$. Thus, the calculated phases for structure factors in space group $P1$ which are related by the twin law should not be independent. In general, twin refinement has several advantages compared with disorder refinement. For example, it improves the ratio of data to parameters to be refined (see Tables 1 and 2), thus adding stability to the refinement process. This is indeed observed in the structure determinations described here, although the data in the true lower symmetry space group are of course strongly correlated according to the twin law. In practice, the distinction between twinning and disorder of the crystal may only be of marginal importance for the macromolecular structure of interest. Borderline cases between lattice disorder and twinning can be anticipated: if the twin domains become small compared with the coherence length of

the X-rays, the twinning will transform into a disorder problem and, *vice versa*, if blocks of molecules in one of the two orientations here possible assume a critical size, they will have to be regarded as twin domains.

We are grateful to the staff of EMBL/DESY (Hamburg) and ESRF (Grenoble) for help with diffraction experiments, to K. Diederichs (Konstanz) for helpful discussions and to J. J. Müller (MDC) and N. Sträter (Freie Universität Berlin) for reading the manuscript. This work was supported by the Deutsche Forschungsgemeinschaft (He 1318/12–3) and the Fonds der Chemischen Industrie.

References

Berger, I., Kang, C. H., Sinha, N., Wolters, M. & Rich, A. (1996). *Acta Cryst.* D**52**, 465–468.
Beuning, P. J., Yang, F., Schimmel, P. & Musier-Forsyth, K. (1997). *Proc. Natl Acad. Sci. USA*, **94**, 10150–10154.
Brennan, R. G., Westhof, E. & Sundaralingam, M. (1986). *J. Biomol. Struct. Dyn.* **4**, 649–665.
Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.
Carr, P. D., Cheah, E., Suffolk, P. M., Vasudevan, S. G., Dixon, N. E. & Ollis, D. L. (1996). *Acta Cryst.* D**52**, 93–104.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Francklyn, C. & Schimmel, P. (1989). *Nature (London)*, **337**, 478–481.
Heinemann, U. (1991). *J. Biomol. Struct. Dyn.* **8**, 801–811.
Herbst-Irmer, R. & Sheldrick, G. M. (1998). *Acta Cryst.* B**54**, 443–449.
Hou, Y. M. & Schimmel, P. (1988). *Nature (London)*, **333**, 140–145.
Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kleywegt, G. J. & Jones, T. A. (1996). *Acta Cryst.* D**52**, 826–828.
Lavery, R. & Sklenar, H. (1989). *J. Biomol. Struct. Dyn.* **6**, 655–667.
Limmer, S., Reif, B., Ott, G., Arnold, L. & Sprinzl, M. (1996). *FEBS Lett.* **385**, 15–20.
Luecke, H., Richter, H. T. & Lanyi, J. K. (1998). *Science*, **280**, 1934–1937.
McClain, W. H. & Foss, K. (1988). *Science*, **240**, 793–796.
Mueller, U., Schübel, H., Sprinzl, M. & Heinemann, U. (1999). *RNA*, **5**, 670–677.
Musier-Forsyth, K., Shi, J. P., Henderson, B., Bald, R., Fürste, J. P., Erdmann, V. A. & Schimmel, P. (1995). *J. Am. Chem. Soc.* **117**, 7253–7254.
Musier-Forsyth, K., Usman, N., Scaringe, S., Doudna, J., Green, R. & Schimmel, P. (1991). *Science*, **253**, 784–786.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Ott, G., Dörfler, S., Sprinzl, M., Müller, U. & Heinemann, U. (1996). *Acta Cryst.* D**52**, 871–873.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Ramos, A. & Varani, G. (1997). *Nucleic Acids Res.* **25**, 2083–2090.
Rees, D. C. (1980). *Acta Cryst.* A**36**, 578–581.
Schimmel, P. (1993). *The Translational Apparatus*, edited by K. H. Nierhaus & B. Wittmann-Liebold, pp. 13–21. New York and London: Plenum Press.
Shah, S. A. & Brünger, A. T. (1999). *J. Mol. Biol.* **285**, 1577–1588.
Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **277**, 319–343.
Stanley, E. (1972). *J. Appl. Cryst.* **5**, 191–194.
Terwilliger, T. C., Kim, S.-H. & Eisenberg, D. (1987). *Acta Cryst.* A**43**, 1–5.
Valegard, K., 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.
Wahl, M. & Sundaralingam, M. (1997). *Biopolymers*, **44**, 45–63.
Yeates, T. O. (1997). *Methods Enzymol.* **276.** 344–358.
Yeates, T. O. & Fam, B. C. (1999). *Structure*, **7**, R25–R29.