Article No. jmbi.1998.2385 available online at http://www.idealibrary.com on IDEAL J. Mol. Biol. (1999) 285, 1577–1588

JMB

The 1.8 Å Crystal Structure of a Statically Disordered 17 Base-pair RNA Duplex: Principles of RNA Crystal Packing and its Effect on Nucleic Acid Structure

Sapan A. Shah[1] and Axel T. Brunger[1,2]*

[1]*Department of Molecular Biophysics and Biochemistry and* [2]*The Howard Hughes Medical Institute, Yale University, New Haven CT 06511, USA*

**Corresponding author*

The crystal structure of a 17 base RNA oligomer, r(CACCGGAUG GUUCGGUG), has been solved to a resolution of 1.8 Å through a combination of molecular replacement, multiple isomorphous replacement phasing, and analysis of observed intensity distributions. The oligomer, which forms a stem-loop in solution, crystallized as a pseudo-infinite duplex in spacegroup *P*321. The asymmetric unit of the crystal contains four superimposed orientations of the duplex that are out of register, such that backbones superimpose, but base identity differs. This static disorder was initially discovered by brominating a single residue per strand in the sequence, and observing four peaks per strand in difference maps phased with a native molecular replacement solution. The presence of four superimposed duplex ''motifs'' related by non-crystallographic hypersymmetry was detected by computing $\langle I^2 \rangle / \langle I \rangle^2$ and Wilson ratios for the observed intensities. The observed ratios matched those produced from calculated intensities of a 4-fold statically disordered model. Multi-conformer simulated annealing refinement against a maximum-likelihood target incorporating experimental phase information was used to refine the 4-fold disordered model to an R_{free} and R of 29.35 % and 25.5 %, respectively. The resulting structure reveals four distinct conformations of the duplex, with an average pairwise backbone rmsd of 2.35 Å. The structural differences between the four conformations, which can be attributed to differences in packing environment, highlight the possible influence of crystal packing forces on nucleic acid X-ray structures. Analysis of inter-helical packing between symmetry-related molecules reveals an RNA ''zipper'' that mediates direct phosphate oxygen-2′ hydroxyl interactions between close-packed phosphate-sugar backbones. This may be a general mode for RNA tertiary interaction that does not depend on metal ions or primary sequence.

Keywords: X-ray crystallography; RNA; disorder; intensity statistics; Wilson ratio

Introduction

The crystal structures of several RNA duplexes (Bayans *et al.*, 1995, 1996; Leonard *et al.*, 1994; Holbrook *et al.*, 1991) have reinforced the notion that base-stacking and packing play an important role in driving RNA crystal formation. In these structures, small RNA oligomers that can exist as stem-loops in solution form mis-paired duplexes in the crystalline state. These duplexes typically stack end-to-end to form pseudo-infinite helices in the lattice. Such a packing scheme maximizes the

Present address: S. A. Shah, McKinsey & Company, 600 Campus Drive, Florham Park, NJ 07932, USA.

Abbreviations used: rmsd, root-mean-square deviation; TAD, torsion angle dynamics; MIR, multiple isomorphous replacement; CNS, Crystallography & NMR system; PEG, polyethylene glycol; NCS, non-crystallographic symmetry.

E-mail address of the corresponding author: brunger@laplace.csb.yale.edu

0022-2836/99/041577–12 $30.00/0

favorable energy associated with π orbital stacking of the bases, and more than compensates for the free energy required to convert a stem-loop conformation to duplex. We have determined the crystal structure of a 17 base-pair RNA duplex in which these same principles apply, but in this case it is at the expense of crystallographic order.

The self-complementary RNA sequence ($RNAI_{inv}$) studied here exists as a stem-loop in solution (Eguchi & Tomizawa, 1991), but has crystallized as hexagonally packed pseudo-infinite duplexes. In addition, the sequence register of adjacently packed helices in the crystalline lattice is variable, with of a total of four such registers distributed non-periodically throughout the crystal. The result, when trigonal symmetry is applied, is an asymmetric unit that is the average of four duplexes whose backbones approximately superimpose but base identity differs. Each conformation, in principle, is related to another by a unique non-crystallographic rotation and translation along the helical axis.

Here, we describe the identification of this type of static disorder through a combination of molecular replacement, site-specific bromination of selected residues, and analysis of $\langle I^2\rangle / \langle I\rangle^2$ (Stanley, 1972) and Wilson ratios (Wilson, 1949, 1951) for the observed intensities. The intensity ratios calculated here can be used generally to screen RNA crystals for the presence of crystallographic hypersymmetry. Structure determination of the $RNAI_{inv}$ crystal involved multi-conformer simulated annealing refinement against a maximum-likelihood target incorporating experimental phase information (Pannu *et al.*, 1998) using the Crystallography & NMR system (CNS) (Brunger *et al.*, 1998). The four conformations of the duplex, although related by a non-crystallographic rotation and translation along the helical axis, are effectively in different packing environments with respect to the phosphate backbone of adjacent helices. These differences are reflected in the 2.35 Å pairwise backbone rmsd between the four conformations when aligned with respect to sequence. The final model obtained after refinement provides four distinct views of the same duplex in different crystal packing environments. These differences are the result of very close packing of RNA helices mediated by direct phosphate oxygen-2′ hydroxyl ''zipper'' interactions.

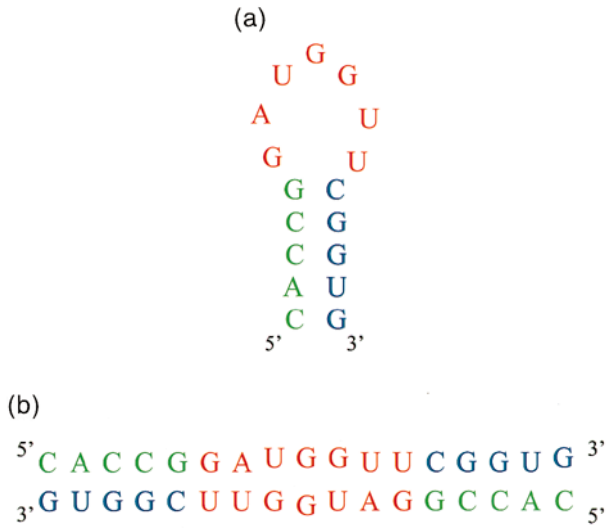

Figure 1. Secondary structure conformations of $RNAI_{inv}$. (a) The $RNAI_{inv}$ sequence in a stem loop conformation. The sequence consists of a five base-pair stem region (green and blue) and a seven residue loop region (red). Two stem loops can form a self duplex (b) by base-pairing in the self-complementary stem regions (green and blue) leaving a mis-paired bulge region (red) that contains a G·G mismatch. Although the $RNAI_{inv}$ oligomer was in conformation (a) prior to crystallization the stem loop was converted to (b) during the crystallization process.

Methodology and Results

RNA crystallization

The RNA sequence crystallized here was based on an antisense RNA transcript involved in the regulation of plasmid copy number in the ColE1 family (Tomizawa, 1990). The construct, known as $RNAI_{inv}$, forms a stem-loop in solution (Eguchi & Tomizawa, 1991) and is shown in Figure 1(a). In an attempt to obtain a crystal structure of the 17 bp stem-loop, milligram quantities of the RNA were chemically synthesized and purified by reverse-phase chromatography, and crystals were obtained at 28 °C using PEG (polyethylene glycol) 300 and NaCl as precipitants. Despite the stability of the stem-loop in solution, native gel analysis (results not shown) indicated that the conformation present in the crystal was duplex, and not stem-loop (Figure 1(b)). Crystal packing forces in combination with the high crystallization temperature presumably drove conversion of the stem-loop into a self-duplex conformation. This type of conformational variability is not unique to the sequence presented here, as evidenced by the self-complementarity of several recent RNA duplex structures (Bayans *et al.*, 1995, 1996; Leonard *et al.*, 1994; Holbrook *et al.*, 1991). Because of the high resolution diffraction limit (>2.3 Å on a CuK_α rotating anode source) and its unusual length (nearly all RNA duplex crystal structures solved to date are only one turn in length), the structure determination of this duplex was pursued.

Molecular replacement

The space group symmetry and unit cell geometry obtained for the $RNAI_{inv}$ crystals is shown schematically in Figure 2. The *c*-axis of the unit cell is long enough to accommodate two 17 bp duplexes stacked end-to-end, assuming a rise per base-pair of approximately 2.7 Å. As a result, an asymmetric unit consisting of 17 bp is sufficient to generate the

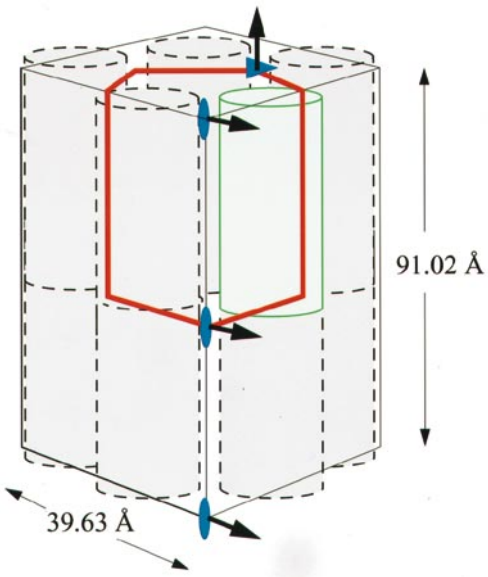

Figure 2. Unit cell diagram. Shown is the observed *P*321 unit cell for the $RNAI_{inv}$, crystal with an asymmetric unit shown in red. The crystallographic 3-fold is parallel to the *c*-axis which is 91.02 Å in length. The 2-fold axes are located parallel with the *a* and *b*-axis, which are both 39.63 Å in length. Superimposed on the observed unit cell is a hexagonal arrangement of a 17 bp canonical *A*-form RNA with a 2.7 Å rise per base-pair, and a helical diameter of approximately 22 Å (green cylinder). Assuming close packing of adjacent cylinders, and one cylinder per asymmetric unit, a unit cell of dimensions similar to the observed cell is obtained.

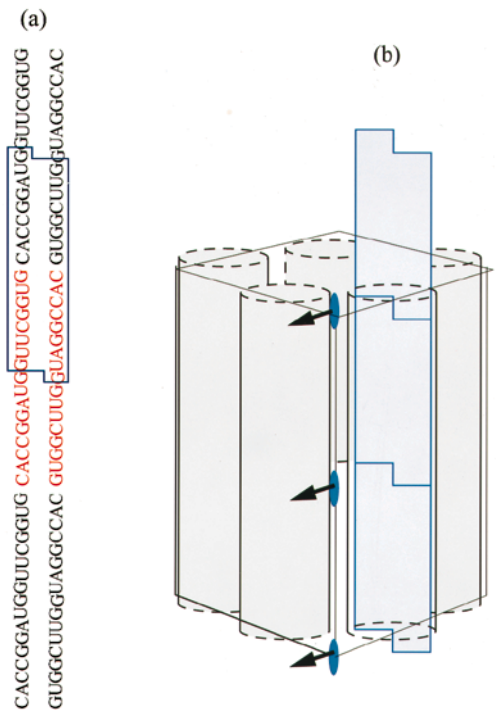

Figure 3. Summary of search model and asymmetric unit. (a) Two-dimensional view of multiple 17 bp $RNAI_{inv}$ sequences, stacked end-to-end in a continuous fashion. Initial molecular replacement trials with a full 17 bp sequence were unsuccessful. A rotation and translation solution was obtained by searching with a 16 bp sequence with single base overhangs, as depicted with the blue outline in (a). An asymmetric unit consisting of this 16 bp duplex plus overhangs, generates a continuous pseudo-infinite helix along the *c*-axis of the crystal when the 2-fold symmetry operators are applied. (b) The crystallographic 2-folds occur directly between adjacent overhangs, generating a continous helix. Crystallographic 3-folds fill the rest of the unit cell (grey cylinders).

entire unit cell when crystallographic operators are applied (Figure 2). Initial rotation searches identified the *c*-axis of the unit cell as the helical axis. However, translation searches using a full 17 bp duplex of arbitrary sequence produced symmetry mates that clashed when crystallographic 2 and 3-fold operators were applied. The problem was solved by searching with a 16 bp duplex with single base overhangs at each end. This construct produced a rotation and translation solution that was ''locked'' into place by the overhangs as shown in Figure 3. The overhang search model generated a pseudo-infinite helix along the *c*-axis of the crystal once crystallographic 2-folds were applied. Packing of the pseudo-infinite helices in the *a* and *b* directions is mediated by the crystallographic 3-folds, which produce a hexagonal packing arrangement. The rotation and translation solution did not depend on the sequence identity of the *A*-form RNA search model, leading us to believe that while the phosphate-sugar backbone was accurately placed, the actual sequence register was still unknown. Initial rigid-body refinements with this model of arbitrary sequence produced an R_{free} of 49 %.

Site-specific bromination of RNA

In order to establish the exact sequence register of the duplex within the backbone framework that molecular replacement had identified, crystals were obtained in which a single uracil in the RNA had been replaced with a bromo-uracil analogue. Data collection statistics for the various brominated sequences are shown in Table 1. The positions of the bromo-uracils were then obtained from ($|F_{Br}| - |F_{nat}|$, Φ_{calc}) difference Fourier maps. The phases were calculated from a duplex of arbitrary sequence that was positioned by molecular replacement.

Only one peak per strand in the asymmetric unit was expected in these difference Fourier maps. However, the difference maps for the BrU11 derivative revealed four sites per strand (Figure 4(a)). Similar results were obtained with the BrU12 (Figure 4(b)) and BrU16 (not shown) modified sequences. The four bromine peaks per strand were located where one would expect them to be, given an *A*-form RNA model in which the base closest to each of the sites was a bromo-uracil. However, in reality, only a single bromo-uracil per

Table 1. Data collection statistics

Construct	$RNAI_{inv}$			
Precipitant	40% PEG 300, 150 mM NaCl, 28 °C			
Space group	*P*321			
Unit cell *a*, *b*, *c* (Å)	39.63, 39.63, 91.02			
Crystal	Native	BrU12	BrU11	BrU16
Resolution (Å)	60-2.5 (2.57-2.5)	50-1.8 (1.85-1.8)	50-2.5 (2.57-2.5)	50-2.5 (2.59-2.5)
Unique reflections	3017	8052	3147	2963
Redundancy	4.20	4.70	5.71	4.53
Completeness (%)	95.2 (68.3)	98.3 (94.7)	98.6 (92.0)	97.6 (93.5)
$I/\sigma(I)$	17.87 (2.04)	19.36 (4.73)	18.78 (4.31)	20.3 (8.41)
R_{merge}[a] (%)	9.2 (27.6)	8.4 (31.7)	9.9 (34.1)	9.7 (26.9)

[a] $R_{merge} = \Sigma_{hkl}\ \Sigma_i|\langle I_{(hkl)}\rangle - I_{(hkl)_i}|/\Sigma_{hkl}\langle I_{(hkl)}\rangle$, where $\langle I_{(hkl)_i}\rangle$ is the measured diffraction intensity and $\langle I_{(hkl)}\rangle$ is the mean of the intensity. Values indicated in parentheses are for the highest resolution shell.

strand was actually introduced. MIR (multiple isomorphous replacement) phasing using all 24 sites obtained from the three derivatives (Table 2) produced a solvent-flattened experimental electron density map (Figure 5) in which the RNA backbone was very well defined, but the majority of bases were disordered. Based on the positions of the bromine peaks for each respective derivative, we concluded that the asymmetric unit in this crystal was an average of four orientations of the duplex that were out of register, the sequences of which are shown in Figure 6. The electron density in the asymmetric unit is the average of these four continuous helices whose backbones superimpose, but base identity differs.

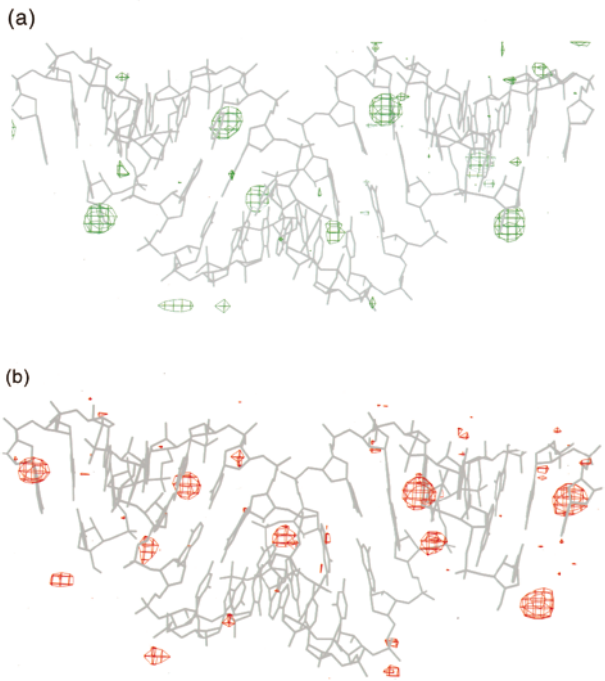

Figure 4. Difference Fourier maps. Only two bromine sites are expected after specifically brominating uracil 12 in the $RNAI_{inv}$ sequence, if only a single conformation is present. A ($F_{Br} - F_{native}$) difference map contoured at 3.0σ (shown in green) phased with a molecular replacement model of arbitrary but non-symmetric sequence (shown in grey) is shown. Eight bromine sites were observed per asymmetric unit, all of which are located exactly where one would expect if the adjacent base was brominated. A similar result is shown in (b), which was obtained after brominating the adjacent residue (uracil 11) in the $RNAI_{inv}$ sequence.

Analysis of intensity statistics

From the difference Fourier maps made with the various bromine derivatives, it was clear that the $RNAI_{inv}$ crystals were disordered. However, based on the difference maps alone, it was impossible to distinguish between twinning or random static disorder as the cause of the multiple bromine sites. Specifically, one must choose between multiple twinned crystals each with a different sequence register (Figure 7(a)), or randomly distributed pseudo-infinite helices that are out of register (Figure 7(b)). Analysis of intensity statistics, based on the work of Wilson & Stanley, allow these two cases to be resolved by computing $\langle I^2\rangle/\langle I\rangle^2$ (Stanley, 1972) and $\langle|F|\rangle^2/\langle I\rangle$ (Wilson) ratios (Wilson, 1949; Wilson, 1951) for the observed intensities. For a non-centrosymmetric crystal $\langle I^2\rangle/\langle I\rangle^2$ equals 2, and $\langle|F|\rangle^2/\langle I\rangle$ equals 0.785 for acentric data. Twinning results in a decrease in the observed $\langle I^2\rangle/\langle I\rangle^2$ ratio, with a theoretical value of 1.5 for a perfectly hemihedrally twinned crystal, and a value of 1.25 for a quadrahedrally twinned crystal. Similarly, twinning results in an increase in the Wilson ratio. The observed values for the $\langle I^2\rangle/\langle I\rangle^2$ and Wilson ratios for the $RNAI_{inv}$ crystals are given in Table 3. The observed intensities for this crystal exhibit a markedly increased $\langle I^2\rangle/\langle I\rangle^2$ ratio and decreased Wilson ratio in comparison with theoretical values expected for a non-centrosymmetric crystal. Thus, a simple twinning operation as the cause of the observed disorder was ruled out. Work by Rogers & Wilson (1953) has shown that parallel repetition of a motif by non-crystallographic translation results in a modulation of the Fourier transform of the motif, giving rise to increased $\langle I^2\rangle/\langle I\rangle^2$ and decreased Wilson ratios for observed intensities. Because the disorder in the $RNAI_{inv}$ crystal can be described as a non-crystallographic rotation and translation of a motif along a parallel helical axis, we suspected that this would also give rise to increased variance of intensities. Table 3 shows the $\langle I^2\rangle/\langle I\rangle^2$ and Wilson ratios based

Table 2. Phasing statistics

Crystal	Native	BrU12[a]	BrU11	BrU16	Overall
Resolution (Å)	60-2.5	50-1.8	50-2.5	50-2.5	50-2.5
Number of sites	0	8	8	8	24
Phasing power[b]	1.3	-	1.56	1.12	-
Figure of merit (%)	-	-	-	-	53.3

[a] Experimental phases were obtained with respect to the BrU12 data set which was treated as native.
[b] Phasing power $= \sqrt{\Sigma_{hkl} F_H^2 / \Sigma_{hkl} (F_{\mathrm{PH,obs}} - F_{\mathrm{PH,calc}})^2}$.

on calculated intensities derived from the $RNAI_{inv}$ unit cell containing one, two, three or four RNA duplex ''motifs''. The observed values for the $\langle I^2 \rangle / \langle I \rangle^2$ and Wilson ratios are in good agreement with those calculated for four motifs, further supporting the presence of 4-fold static disorder in the $RNAI_{inv}$ crystal. It should be noted that the four register shifts do not coincide with alternate origin choices in space group *P*321.

Multi-conformer refinement

Based on the intensity statistics and the positions of bromine peaks for each respective derivative, we were able to determine the arrangement of the four duplexes which are averaged in the asymmetric unit. In space group *P*321, the crystallographic 2-folds perpendicular to the helical axis intersect each of the four conformations at points that do not coincide with a molecular 2-fold (Figure 6(a)). When all four superimposed sequences are present, however, *P*321 symmetry is upheld. Unfortunately, when considered individually, each duplex cannot be refined in space group *P*321 without introducing breaks in the model. In order to include the biologically relevant duplex without introducing breaks in the model, we chose to refine the structure in space group *P*3. This was accomplished by expanding the *P*321 amplitudes, experimental phases and test set into space group *P*3, and then imposing strict non-crystallographic symmetry (NCS) constraints between appropriate conformers to generate the 2-fold crystallographic symmetry. All four duplexes, representing four different sequence registers, were refined as described in Materials and Methods. Final statistics for the model, which includes 180 water molecules and 8 bromine atoms, are given in Table 4. Figure 8 shows a portion of the final multi-conformer model superimposed on a σ_A weighted (Read, 1986) composite annealed omit map (Shah *et al.*, 1997; Hodel *et al.*, 1992) to 1.8 Å. The phosphate-sugar backbones for the four conformations are tightly clustered and well ordered. The differing sequence register between conformations results in disordered density for many of the base-pairs in the asymmetric unit. At several positions along the duplex, however, the four superimposed base-pairs are all oriented as purine-pyrimidine. For example, Figure 9 shows one such position in which the static disorder results in the superposition of two G·U, one A·U, and one G·C base-pairs. In this case, the base density is fairly well ordered, and shows the presence of multiple types of purine-pyrimidine base-pairs.

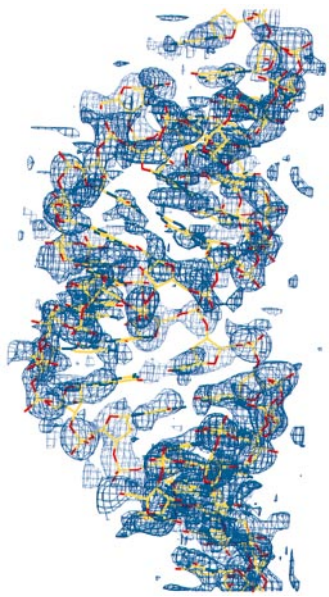

Figure 5. Experimental electron density map. Shown is a solvent-flattened MIR map to 2.5 Å contoured at 1.2 σ, superimposed on a partially refined molecular replacement solution of arbitrary sequence.

Examining the model with respect to electron density, as done above, shows how well the backbones of each static duplex conformation superimpose. It is important to note, however, that the electron density reflects the average of sections of the four duplexes that are not identical in sequence. Alternatively, the four models obtained after multi-conformer refinement can be rotated and translated in order to align sequences, thereby revealing the conformational flexibility of the $RNAI_{inv}$ helix (Figure 10). The average pair-wise rmsd for the family of aligned sequences is 2.3 Å.

Discussion

Basis for the static disorder

When crystallized, nucleic acid oligomers tend to form a continuous helix by stacking end-to-end in the lattice in order to maximize the favorable

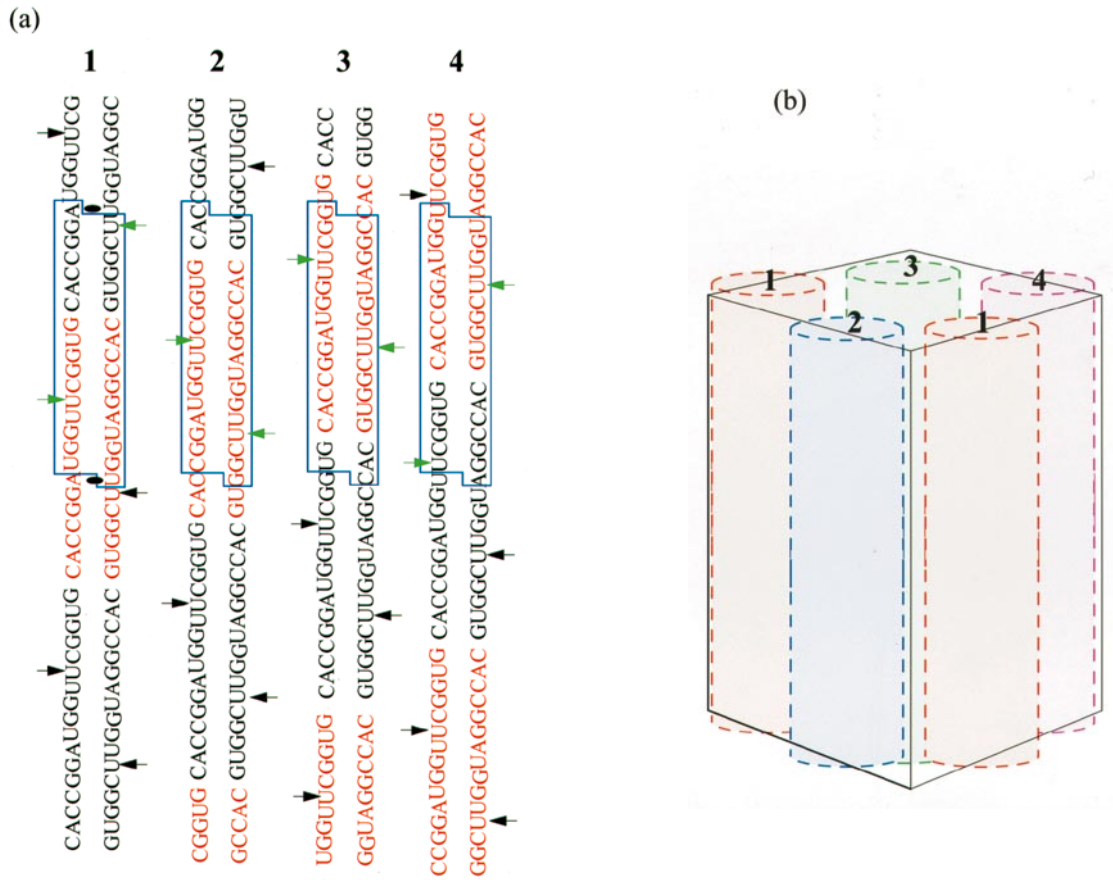

Figure 6. Summary of packing disorder. This Figure provides an explanation for the multiplicity of bromine sites observed in difference maps where only a single residue per strand was actually brominated. (a) Four continous helices made up of stacked $RNAI_{inv}$ duplexes. The pseudo-infinite helices (1,2,3,4) are out of sequence register with one another. Shown in blue outline is the *P*321 asymmetric unit, with the crystallographic 2-folds perpendicular to the helical axis indicated on conformation 1. The true asymmetric unit is the average of all four individual conformations. Indicated with arrows are where bromine peaks would be expected when uracil 12 is brominated prior to crystallization. If an equivalent asymmetric unit is taken (blue box outline) from each continuous helix, the positions of bromine peaks from each continuous helix differ, as indicated with green arrows. The result when the pseudo-continous helices are packed as shown in (b) and *P*321 symmetry is applied, is an observation of the average asymmetric unit containing eight bromine sites.

energy associated with base stacking. It is, therefore, advantageous for the sequence and backbone to repeat at identical intervals (i.e. every turn), or multiples thereof, in order to avoid register errors. When an RNA duplex of length one turn is considered, both the sequence and backbone repeat every 11 residues, and therefore segments of the continuous helix formed by the duplex can be related by simple lattice translations. In contrast, consider a duplex of length 17 bp, like the one presented here. This oligomer will still form a pseudo-continuous helix by end-to-end stacking in the crystal. However, in this case, while the backbone repeats every 11 residues, the sequence repeats every 17. If this pseudo-continuous helix would crystallize, it would produce a disordered asymmetric unit, because inter-helix packing depends only on the phosphate backbone, not the sequence. The result is some type of disordering of the effective asymmetric unit. In the case presented here, $RNAI_{inv}$ contains a G·G mismatch which distorts the phosphate-sugar backbone away from *A*-form geometry. The observed packing disorder is modulated by this bulge, with adjacent infinite helices staggered throughout the crystal in order to avoid placing G·G bulges in contact with each other. The result is four different sequence registers, or four statically disordered conformations (Figure 6).

In support of this explanation, we have also determined the structure of a 17 bp RNA duplex based on the $RNAI_{inv}$ sequence, but with all Watson-Crick base-pairs (S.A.S. & A.T.B., unpublished results). In this crystal, the asymmetric unit contains exactly 11 bp, or one turn of the infinite helix along the *c*-axis of the crystal. The asymmetric unit is not large enough to accommodate the 17 bp duplex. When a single uracil in the sequence is labeled with a bromine atom, difference peaks are observed at every residue in the asymmetric unit, indicating a complete static disorder. Therefore, when the bulge is removed from the sequence,

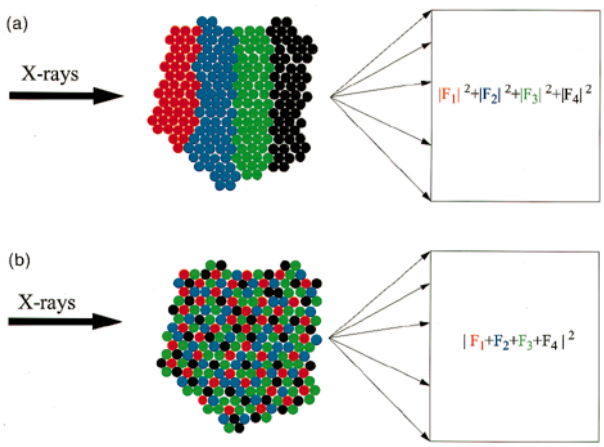

Figure 7. Diagram comparing twinning *versus* disorder. Shown in (a) and (b) are the two possible distributions for the out of register pseudo-continuous helices in the $RNAI_{inv}$ crystal. Both show a cross-section of the hexagonal packed crystal, with each of the four distinct registers indicated with a different color (red, black, blue, or green). In (a), the $RNAI_{inv}$ crystal is a quadrahedral twin, with the four psuedo-infinite helices forming microcrystals in a twinned setting. In such a case, the scattering is the sum of the intensities from each twin. In (b), the four registers are distributed evenly throughout the crystal, in a non-periodic manner. In this case, the observed intensities are derived from the complex average of the scattering factors from each conformation. The distinction between (a) and (b) can be made based on analysis of observed intensity variances and Wilson ratios.

adjacently packed helices are free to occupy any one of 17 different sequence registers, as opposed to only four in the $RNAI_{inv}$ case.

A disordering similar to that observed here with RNA has been seen with *Z*-DNA, where a superposition of a helical and crystallographic axes occurred (Brennen *et al.*, 1986). In the *Z*-DNA case, oligomers of less than one turn in length were found to form pseudo-continuous infinite helices along a crystallographic screw axis. When the length of the oligomer was such that an integral number of oligomers could not form one turn (12 residues), disordering occurred.

Analysis of RNA duplex structure

At low resolution, the $RNAI_{inv}$ structure is close to *A*-form in conformation. The number of residues per turn is approximately 11, with an average rise per base-pair of 2.7 Å. The G·G mismatch in the sequence distorts the backbone slightly, resulting in an undertwisting of the duplex at that position. Because of the disorder present, any high-resolution analysis of the $RNAI_{inv}$ structure must be made conservatively. The multi-conformer refinement scheme used here increases the number of parameters involved by a factor of 4. However, use of high-resolution data as well as experimental phase information throughout refinement makes this a well-determined system. Therefore, analysis

Table 3. Intensity ratios for observed and calculated data

	Number of motifs				
	1	2	3	4	Observed
$\langle I^2\rangle/\langle I\rangle^2$	2.32	2.93	3.5	4.08	4.04
$\langle \|F\| \rangle^2/\langle I\rangle$	0.755	0.706	0.690	0.676	0.698

Table 4. Refinement statistics

Resolution range (Å)	15.0-1.8 (1.86-1.83)
Bulk solvent parameters	
k_{sol} (electrons/Å^3)	1.303
B_{sol} (Å^2)	36.45
R_{free}[a] (%)	29.35
R_{cryst}[a] (%)	25.56
Number of atoms (non-hydrogen)	2944
Number of atoms (bromine)	8
Number of water molecules	180
Rmsd bond length (Å)	0.0058
Rmsd bond angles (°)	1.03
Average *B*-factor (Å^2)	
Water atoms	8.63
RNA atoms	12.13

[a] $R_{cryst} = \Sigma_{hkl}||F_{obs}(hkl)| - k|F_{calc}(hkl)||/\Sigma_{hkl}|F_{obs}(hkl)|$. $R_{free} = R_{cryst}$ for a test set of reflections (10 %) not used during refinement. Values indicated in parentheses are for the highest resolution shell.

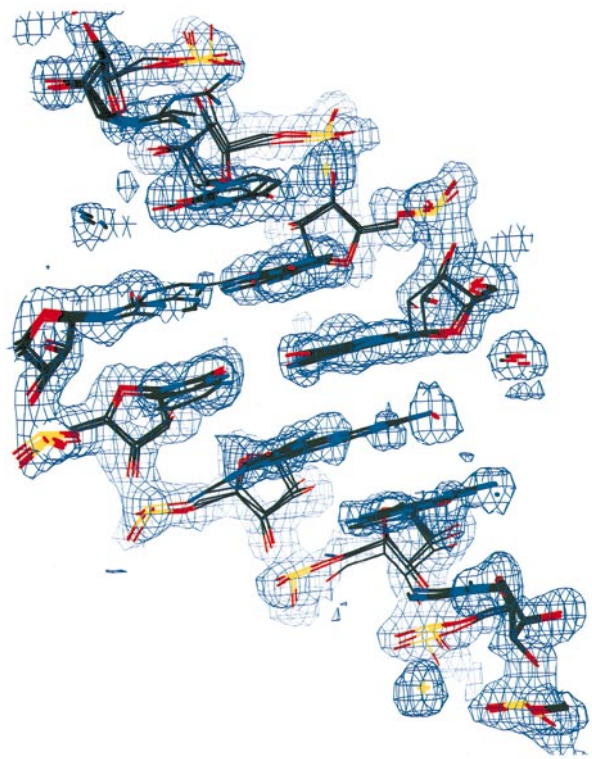

Figure 8. An σ_A weighted omit electron density map. Shown is a portion of the final model superimposed on a composite annealed omit map to 1.8 Å, contoured at 1.0σ.

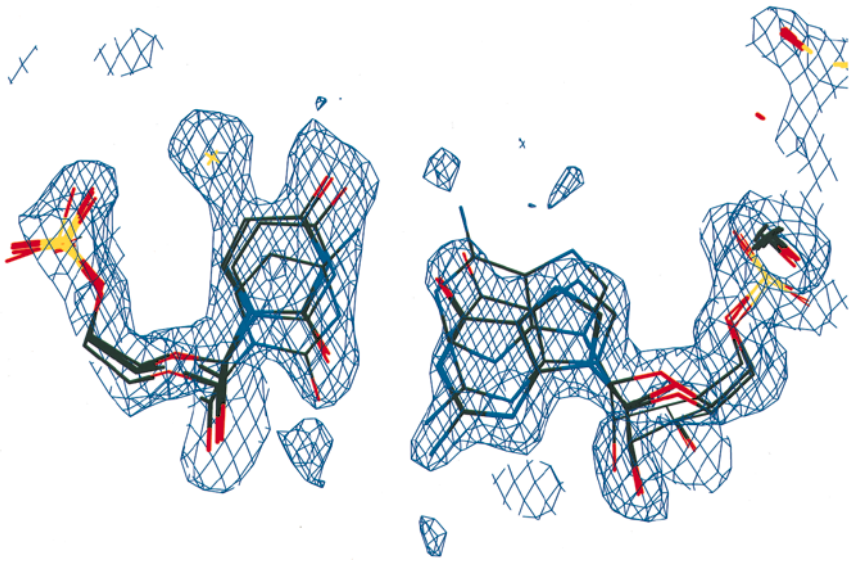

Figure 9. Static base disorder. Shown is a section of the final model superimposed on a composite omit map to 1.8 Å, contoured at 1.0σ. At the particular position shown, the electron density in the asymmetric unit consists of the average of four base-pairs (G·U, G·U, A·U, and G·C).

of the differences between phosphate backbone structure between conformers is significant. Furthermore, the observed static disorder provides the unique opportunity to view the $RNAI_{inv}$ duplex in four different conformations, each in a different packing environment.

Each conformation is related to another one by a rotation and translation along the pseudo-infinite helix parallel with the *c*-axis of the crystal. In the canonical *A*-form RNA case, the backbone of these four duplexes would be identical in structure. However, in the $RNAI_{inv}$ crystal, the close packing of helices in the *a* and *b* directions (Figure 11), places each duplex in a slightly different packing environment. We observe many direct contacts between the phosphate oxygen atoms from the backbone of one helix and the 2′-hydroxyl group of an adjacent helix. These close contacts distort the backbone geometry at various positions in each conformer. The result is an average pair-wise rmsd of 2.3 Å between the four RNA duplex conformations. Figure 12 shows a comparison of a canonical

A-form RNA helix and the final refined model for the four conformations of duplex in the $RNAI_{inv}$ crystal. The primary differences between the various conformations are distortions in phosphate-sugar backbone conformation that result in changes in the major groove width of each duplex, as calculated with CURVES 5.1 (Lavery & Sklenar, 1996). Major groove widths vary between 3 and 8 Å depending on which conformation is considered. Interestingly, minor groove widths are virtually identical between the four conformations (10 Å). These deviations away from *A*-form geometry illustrate how large an effect packing forces can have on RNA conformation, and further show that the largest deviations are present primarily in the major groove.

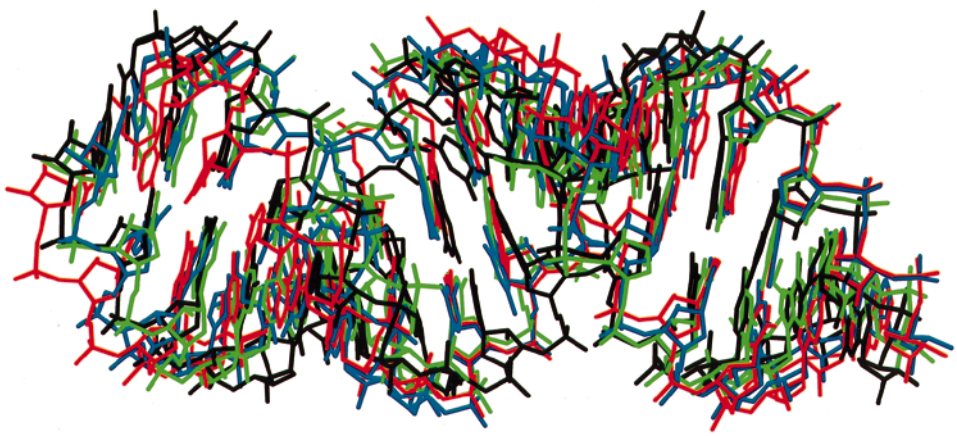

Figure 10. Superposition of conformers based on sequence. Shown in red, green, black and blue, are the final refined coordinates for the four conformations of the $RNAI_{inv}$. The conformations have been superimposed based on sequence. This required a rotation and translation away from the crystallographic position along the helical axis in the unit cell. The average pairwise rmsd between conformers is 2.3 Å.

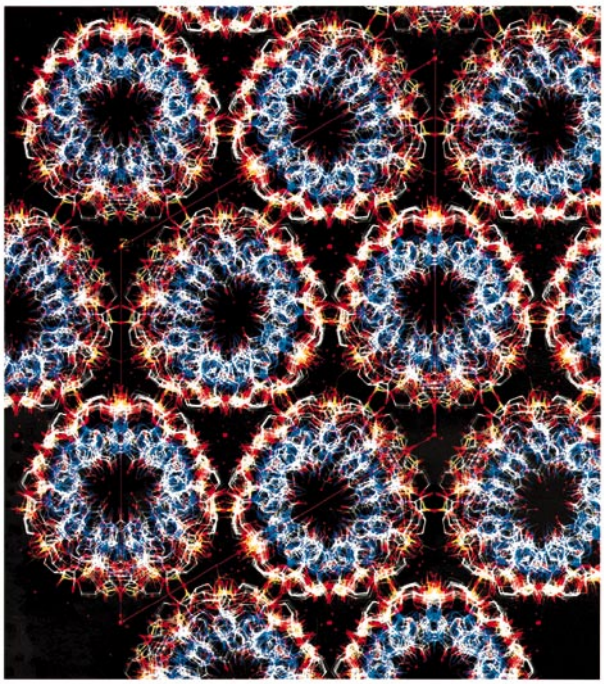

Figure 11. Inter-helical packing in the $RNAI_{inv}$ crystal. Shown is a view down the *c*-axis of the unit cell. The outline of the unit cell is in red.

Analysis of crystal packing

The pseudo-continuous infinite helices in the $RNAI_{inv}$ crystal structure are packed tightly in a close hexagonal arrangement (Figure 11), resulting in a bulk solvent content of 38 %. Such tight packing of helices was unexpected given the electrostatic penalty of placing negatively charged phosphate groups in close proximity. Detailed analysis of the packing environment reveals some important interactions that mediate the contacts between adjacent symmetry-related RNA helices. Contacts between backbone atoms are mediated by phosphate oxygen atoms and 2′-hydroxyl groups on symmetry-related RNA backbones. The symmetry-related backbone atoms intercalate by employing a zipper-like interaction (Figure 13). The zipper consists primarily of alternating phosphate oxygen-2′ hydroxyl links with additional weak 2′ hydroxyl-2′ hydroxyl interactions. Direct contacts between phosphate oxygen atoms on one strand and the 2′ hydroxyl groups on the symmetry-related strand approach 2.6 Å between heavy-atom centers. The hydroxyl-hydroxyl interactions are weaker, with a distance of 3.7 Å between heavy-atom centers. By alternating phosphate-hydroxyl interactions in this zipper-like manner, electrostatically unfavorable direct phos-

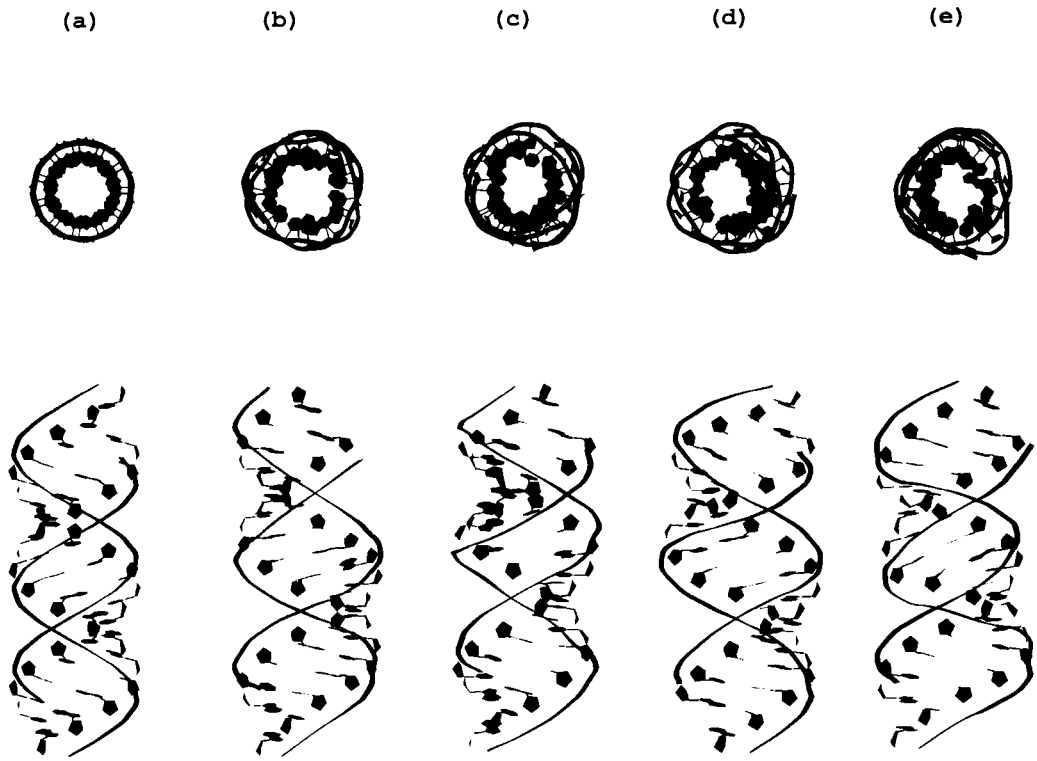

Figure 12. Comparison between the $RNAI_{inv}$ duplex and canonical *A*-form RNA. Shown are (a) two views of canonical *A*-form RNA and ((b), (c), (d), (e)) the four distinct conformations of the $RNAI_{inv}$ duplex.

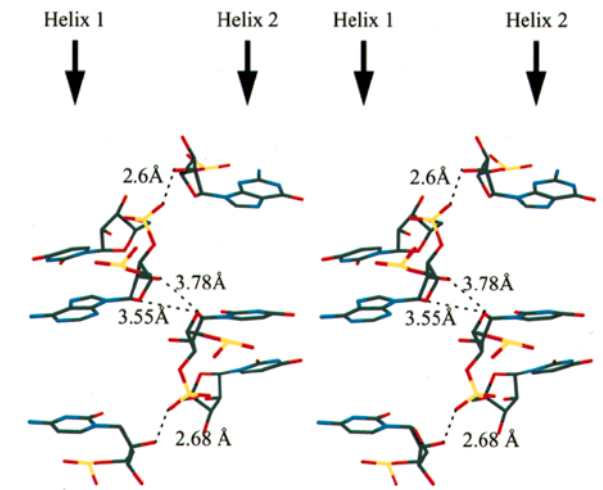

Figure 13. Shown is a stereo view of the phosphate-ribose zipper that mediates direct backbone contacts between symmetry-related helices (helix 1 and helix 2) in the $RNAI_{inv}$ crystal. Distances indicated are between heavy-atom centers.

phate oxygen-phosphate oxygen interactions are avoided. There are, however, indirect phosphate-phosphate contacts. The *c*-axis projection of the packing in the $RNAI_{inv}$ crystal (Figure 11) reveals a solvent channel down the 3-fold axis between helices. These solvent molecules mediate indirect phosphate oxygen-phosphate oxygen and phosphate oxygen-2′ hydroxyl interactions. Interestingly, most water molecules lie directly on the 3-fold axis, requiring them to be treated as special positions throughout refinement. They appear to play a critical role in stabilizing the inter-helical packing.

Summary

In conclusion, we have determined the 1.8 Å resolution crystal structure of a 17 bp RNA duplex, r(CACCGGAUGGUUCGGUG). The structure is statically disordered, with the asymmetric unit containing four superimposed conformations of the duplex, each related by a non-crystallographic rotation and translation along a helical axis parallel with the *c*-axis of the crystal. The disorder was identified by site-specific incorporation of bromo-uracil analogues, and analysis of observed intensity variances and Wilson ratios. These observed intensity ratios were found to be sensitive to the number of static conformations present, and have been shown here to be generally useful for screening crystals for the presence of non-crystallographic hypersymmetry. The structure was solved by a combination of molecular replacement and MIR phasing. Refinement, which was carried out against a maximum-likelihood target incorporating experimental phase information (Pannu *et al.*, 1998) and included four conformations of the duplex (Burling & Brunger, 1994), resulted in a R_{free} of 29 %. The duplex structure is essentially *A*-form in nature, with a pitch of 2.7 Å and 11 residues per turn. Close packing of adjacent helices in the crystal is observed. Specifically, adjacent backbones interact *via* a zipper interaction that mediates alternating direct contacts between phosphate oxygen atoms and 2′ hydroxyl groups. This may be a general mode for tertiary RNA backbone interaction that does not depend on primary sequence or metal ions. In the $RNAI_{inv}$ crystal, these interactions distort the duplex away from *A*-form geometry and differences are observed between the four conformations obtained after multi-conformer refinement, when aligned with respect to sequence. Most notably, large differences (>4 Å) in major groove width are observed between conformers. The structural differences observed between static conformers of the duplex also illustrate the significant influence of crystal packing forces on nucleic acid backbone geometry and the inherent flexibility of RNA oligomers.

Materials and Methods

RNA crystallization

The RNA oligomer r(CACCGGAUGGUUCGGUG), both native and with bromo-uracil, was chemically synthesized by the W. M. Keck Foundation Biotechnology Resource Laboratory at Yale University. Following base deprotection with tetra-butyl ammonium flouride, oligonucleotides were purified to homogeneity by HPLC on a Vydac C4 reverse-phase column using a triethylammonium acetate (pH 6.8)/acetonitrile buffer system (Webster *et al.*, 1991). Samples were snap-cooled prior to crystallization, in order to facilitate stem-loop formation. Native gel analysis (Gregorian & Crothers, 1995) indicated that the conformation prior to crystallization was indeed stem-loop. Crystals were obtained by the hanging drop vapor diffusion method at 28 °C against 40 % (v/v) PEG 300, 150 mM NaCl, and 100 mM sodium acetate (pH 4.5). Trigonal crystals of length 0.2 mm in each dimension were obtained after one day. Native gel analysis on the crystals indicated that the conformation present in the lattice was a mis-paired duplex, as opposed to a stem-loop.

Data collection

A native data set complete to 2.5 Å resolution was obtained on a RAXIS II system and a CuK_{α} rotating anode X-ray source. After data reduction and scaling using the programs DENZO and SCALEPACK (Otwinowski, 1993), the space group was determined to be *P*321 with cell parameters $a = 39.627$ Å, $b = 39.627$ Å, $c = 91.018$ Å. Derivatives were obtained by site-specifically incorporating bromo-uracil at positions U11, U12 and U16 in the 17 base sequence. An anomalous data set complete to 1.8 Å resolution on the U12 derivative was collected at beamline X12b at Brookhaven National Lab. Data collection statistics for the native and three derivatives are reported in Table 1.

Molecular replacement

Molecular replacement was carried out with CNS (Brunger *et al.*, 1998) using the direct rotation function

(Delano & Brunger, 1995) and correlation coefficient translation function (Fujinaga & Read, 1987). Canonical *A*-form RNA models of varying length and pitch were generated with the program Insight II (1995). Direct rotation searches, using diffraction data between 6.0 and 3.0 Å, identified the *c*-axis of the crystal as the helical axis. The highest correlation rotation solution, with a peak height of 5.5σ above the mean, was obtained using a 16 bp duplex with single base overhangs on each end, with a pitch of 2.67 Å rise/bp. The translation search carried out with this oriented model identified, with a peak height of 7.8σ above the mean, the location of the duplex in the asymmetric unit such that when crystallographic 2-folds are applied, a continuous helix along the *c*-axis of the crystal is generated. The molecular replacement solution did not depend on the sequence of the RNA helix.

Experimental phasing

MIR phasing was carried out with CNS (Brunger *et al.*, 1998). Difference Fourier maps calculated using phases from the molecular replacement solution were used to determine heavy-atom positions for each derivative. Experimental phases were obtained relative to the 1.8 Å resolution Br U12 derivative data, which was used as ''native'' (negative occupancies were used to subtract the BrU12 sites when computing lack-of-closure expressions between the native and ''derivative''). Final phasing statistics are listed in Table 2.

Analysis of intensity statistics

All calculations of $\langle I^2\rangle/\langle I\rangle^2$ and Wilson $(\langle|F|\rangle^2/\langle I\rangle)$ ratios (Stanley, 1972; Wilson, 1951) were carried out with the program CNS. Ratios were computed bin-wise from 2.5 to 1.8 Å in thin resolution shells (approximately 250 reflections per bin) to account for resolution-dependent intensity variation. Final values represent the average of the bin-wise ratios.

Refinement

Simulated annealing refinement was carried out with CNS. Because of the nature of the static disorder, the crystallographic 2-folds present in space group *P*321 do not coincide with the molecular 2-folds of the RNA when a single conformation is considered. As a result, refinement was carried out using *P*321 data that were expanded to *P*3 symmetry, which allowed the relevant 17 bp double helix to be refined, but doubled the number of duplexes in the asymmetric unit. Four pairs of the eight duplexes in the *P*3 cell are related by a NCS 2-fold (which was a crystallographic 2-fold in the *P*321 space-group), and were constrained with strict NCS (Weis *et al.*, 1990) throughout the refinement. A test set for cross-validation consisting of 10 % of data was picked prior to data expansion. The locations of the four statically disordered conformations of the duplex, each related by a rotation and translation along the pseudo-infinite helix parallel with the *c*-axis of the unit cell, were determined by analysing the relative positions of bromine sites from each of the three derivatives. All refinements were carried out against a maximum likelihood experimental phase target (MLHL; Pannu *et al.*, 1998), using all data, i.e. no cutoffs were employed. Each conformation, which consisted of two duplexes stacked end-to-end, was refined individually against the *P*3 data through 100 steps of constant temperature torsion angle molecular dynamics (TAD; Rice & Brunger, 1994) at 5000 K followed by 100 steps of conjugate gradient minimization. The R_{free} after individual refinement of each conformation was approximately 38 %. The four conformations, which consisted of eight duplexes in total, were then simultaneously included in the refinement by using a multi-conformer refinement scheme (Kuriyan *et al.*, 1991; Burling & Brunger, 1994), and subjected to 1000 steps of TAD at 1000 K, which resulted in a R_{free} of 33 %. Multiple cycles of grouped *B*-factor refinement and water placement, as well as inclusion of a bulk-solvent model, reduced the R_{free} to less than 30 %. Multiple rounds of water placement and inspection of composite omit maps were carried out as described (Shah *et al.*, 1997). Statistics for the final model, which contains four unique conformations of the 17 bp duplex, and 150 water molecules, are given in Table 4.

Protein Data Bank accession number

The coordinates and structure factors have been deposited in the NDB (deposit number AR0003).

Acknowledgements

We acknowledge Dushyant Pathak and Julie Carruthers for assistance with RNA crystallization, and Malcolm Capel for access to beamline X12b at BNL. We also thank Pete Klostermann and Thomas A. Steitz for stimulating discussions, Paul D. Adams for advice concerning structure refinement and Sarah Stallings for critical reading of the manuscript. Support by the National Science Foundation to A.T.B (grant DBI9514819) is gratefully acknowledged.

References

Bayens, K. J., DeBondt, H. L. & Holbrook, S. R. (1995). Structure of an RNA double helix including uracil-uracil base-pairs in an internal loop. *Nature Struct. Biol.* **2**, 56-62.

Bayens, K. J., DeBondt, H. L., Pardi, A. & Holbrook, S. R. (1996). A curved RNA helix incorporating an internal loop with G-A and A-A non-Watson-Crick base-pairing. *Proc. Natl Acad. Sci. USA*, **93**, 12851-12855.

Brennen, R. G., Westhof, E. & Sundarlingam, M. (1986). Structure of a Z-DNA with two different backbone chain conformations. Stabilization of the decadeoxyoligonucleotide d(CGTACGTACG) by $[Co(NH_3)_6]^{3+}$ binding to the guanine. *J. Biomol. Struct. Dynam.* **3**, 649-665.

Brunger, A. T., Adams, P. D., Clove, G. M., DeLano, W. L., Gros, P., Grosse-Kantsleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). Crystallogaphy & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallog. sect. D*. In the press.

Burling, F. T. & Brunger, A. T. (1994). Thermal motion and conformational disorder in protein crystal structures: comparison of multi-conformer and time-averaging models. *Israel J. Chem.* **34**, 165-175.

Delano, W. L. & Brunger, A. T. (1995). The direct rotation function: rotational Patterson correlation

search applied to molecular replacement. *Acta Crystallog. sect. D*, **51**, 740-748.
Eguchi, Y. & Tomizawa, J. (1991). Complexes formed by complementary RNA stem-loops: their formations, structures and interactions with ColE1 Rom protein. *J. Mol. Biol.* **220**, 831-842.
Fujinaga, M. & Read, R. J. (1987). Experiences with a new translation-function program. *Acta Crystallog.* **20**, 517-521.
Gregorian, R. S. & Crothers, D. M. (1995). Determinants of RNA hairpin loop-loop complex stability. *J. Mol. Biol.* **248**, 968-984.
Hodel, A., Kim, S.-H. & Brunger, A. T. (1992). Model bias in macromolecular structures. *Acta Crystallog. sect. A*, **48**, 851-859.
Holbrook, S. R., Cheong, C., Tinoco, I. & Kim, S. H. (1991). Crystal structure of an RNA double helix incorporating a track of non-Watson-Crick base-pairs. *Nature*, **353**, 579-581.
InsightII, (1995). *Molecular Modeling System, version 95.0*, BIOSYM/MSI.
Kuriyan, J., Osapay, K., Burley, S. K., Brunger, A. T., Hendrickson, W. A. & Karplus, M. (1991). Exploration of disorder in protein structures by X-ray restrained molecular dynamics. *Proteins: Struct. Funct. Genet.* **10**, 340-358.
Lavery, R. & Sklenar, H. (1996). *CURVES 5.1, Helical Analysis of Irregular Nucleic Acids*.
Leonard, G. A., McAuley-Hecht, K. E., Ebel, S., Lough, D. M., Brown, T. & Hunter, W. N. (1994). Crystal and molecular structure of r(CGCGAAUUAGCG): an RNA duplex containing two G(anti)-A(anti) base-pairs. *Structure*, **2**, 483-494.
Otwinowski, Z. (1993). Oscillation data reduction program. In *Data Collection and Processing: Proceedings of the CCP4 Study Weekend* (Sawyer, L., Isaacs, N. & Bailey, S., eds), pp. 56-62, SERC Daresbury Laboratory, Warrington, UK.
Pannu, N. S., Murshudov, G. N., Dodson, E. J., Read, R. J. (1998). Incorporation of prior phase information strengthens maximum likelihood structural refinement. *Acta Crystallog. sect. D.* In the press.
Read, R. J. (1986). Improved fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallog. sect. A*, **42**, 140-149.
Rice, L. M. & Brunger, A. T. (1994). Torsion angle dynamics: reduced variable conformational sampling enhances crystallographic structure refinement. *Proteins: Struct. Funct. Genet.* **19**, 277-290.
Rogers, D. & Wilson, A. J. C. (1953). The probability distribution of X-ray intensities. V. A note on some hypersymmetric distributions. *Acta Crystallog.* **6**, 439-449.
Shah, S. A., Shen, B. W. & Brunger, A. T. (1997). Human ornithine aminotransferase complexed with L-canaline andgabaculine: structural basis for substrate recognition. *Structure*, **5(8)**, 1067-1075.
Stanley, E. (1972). The identification of twins from intensity statistics. *J. Appl. Crystallog.* **5**, 191-194.
Tomizawa, J. (1990). Control of ColE1 plasmid replication: intermediates in the binding of RNA I and RNA II. *J. Mol. Biol.* **212**, 683-694.
Webster, K. R., Shamoo, Y., Konigsberg, W. & Spicer, E. K. (1991). A rapid method for purification of synthetic oligoribonucleotides. *Biotechniques*, **11**, 658-661.
Weiss, W. I., Brunger, A. T., Skehel, J. J. & Wiley, D. C. (1990). Refinement of the influenza virus haemagglutinin by simulated annealing. *J. Mol. Biol.* **212(4)**, 737-761.
Wilson, A. J. C. (1949). The probability distribution of X-ray intensities. *Acta Crystallog.* **2**, 318-321.
Wilson, A. J. C. (1951). Variance of X-ray intensities. *Research*, **4**, 141-142.

Edited by I. Tinoco

(Received 24 August 1998; received in revised form 3 November 1998; accepted 4 November 1998)