Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Direct-methods determination of an RNA/DNA hybrid decamer at 1.15 Å resolution

Gye Won Han

Molecular Biology Institute, University of California at Los Angles, Los Angeles, CA 90095-1570, USA

Correspondence e-mail: gyewon@mbi.ucla.edu

Received 10 July 2000
Accepted 14 November 2000

PDB and NDB References: RNA/DNA hybrid, 1g4q, AH0012.

For the first time, *ab initio* direct methods have been used to solve the crystal structure of an RNA/DNA hybrid decamer. The RNA/DNA sequence corresponds to the leftmost two-thirds of the polypurine tract (PPT), the primer for second-strand DNA synthesis by HIV-1 reverse transcriptase (RT). Direct methods using *Shake-and-Bake* (*SnB*) yielded solutions for the RNA/DNA decamer molecule using 1.15 Å data, which is just on the resolution edge of what might work with direct methods. Atomic positions for 96% of the entire molecule, containing 514 non-H atoms including three Ca^{2+} ions, were easily interpreted from a Fourier map based on the '*Shake-and-Bake*' minimal function and *CROQUE* phase-refinement program. Only six atoms, primarily in the sugar linkage, were missing in this Fourier map. At present, the *R* factor of the model is 0.143 ($R_{\rm free}$ = 0.186) for the 562 non-H atom sites located. The conformation of the RNA/DNA helix is A-form, with a typical A-helix minor-groove width. This paper presents the methodology used in solving this structure.

1. Introduction

As a human immunodeficiency virus (HIV) infected cell copies the RNA of the infecting virus into the first or minus strand DNA, a special domain of reverse transcriptase or RT, known as RNase H, digests away the now unnecessary RNA strand. However, digestion stops when the enzyme reaches a particular stretch of the RNA/DNA hybrid known as the polypurine tract or PPT, 5′-aaaagaaaaaggggggg-3′/5′-CCC-CCCTTTTCTTTT-3′ (lower case for RNA and upper case for DNA). Prevention of further digestion is essential, as the PPT is to serve as a primer for the second or plus strand DNA synthesis. If PPT RNA were digested along with all the other RNA, then the second DNA strand could not be made and the virus could not infect the host. Hence, we would like to elucidate the special structural features of the PPT that prevent it from being digested by the RNase H domain of RT. The direct methods used to solve the crystal structure of this RNA/DNA hybrid are presented here.

A DNA/DNA crystal structure of the leftmost two-thirds of the PPT, substituting C for the initial A in order to prevent fraying of the end of the helix, was solved previously by MIR plus anomalous data using bromo and iodo derivatives (Han *et al.*, 1997). The DNA/DNA decamer was a classical B-helix, but one that exhibited the straight helix axis and extra-narrow minor groove expected of A-tracts or runs of adenine bases. Groove narrowing is fully developed in the A-A-A-A region, but not in the A-A-A region, which previous investigators have proposed as being too short to exhibit typical A-tract properties (Koo *et al.*, 1986). It was expected that crystals

containing the same sequence of RNA/DNA hybrid, caaagaaaag/CTTTTCTTTG, would be different from the A-like or heteromerous helices that are ordinarily seen with RNA/DNA hybrids. If so, then this distinctive structure of the PPT might explain its indigestability by RNase H.

It is the crystallographer's dream to solve a macromolecular crystal structure by direct methods without any *a priori* knowledge of the structure or heavy-atom isomorphous derivatives. The tremendous increases in computer speed along with the high resolution and accuracy of data available from high-energy synchrotron radiation in recent years have facilitated progress towards using direct methods to solve macromolecular structures. It is now possible to determine structures that are more than ten times more complex than were routinely possible to solve merely five years ago (Weeks *et al.*, 1994; Miller *et al.*, 1994; Schäfer *et al.*, 1996). This paper presents the method which worked in solving this RNA/DNA hybrid decamer.

The *Shake-and-Bake* (*SnB*) program (Weeks & Miller, 1999) has provided a powerful new tool for direct-methods X-ray crystal structure determination. In this procedure, the minimal function

$$R(\Phi) = \sum_{H,K} A_{H,K}\left[\cos(\varphi_H + \varphi_K + \varphi_{-H-K}) - \frac{I_1(A_{HK})}{I_0(A_{HK})}\right] / \sum_{H,K} A_{HK}$$

serves as an extremely useful figure of merit for selecting trials that have converged to solution. It expresses a relationship among phases related by triplet invariants, $\Phi = \varphi_H + \varphi_K + \varphi_{H-K}$, that have the associated weights

$$A_{HK} = 2|E_H E_K E_{H+K}|/N^{1/2},$$

where the $|E|$s are the normalized structure-factor magnitudes and N is the number of atoms, assumed identical, in the primitive unit cell.

In *SnB*, trial structures are generated which are comprised of randomly positioned atoms, and each of the trials then is refined alternately in reciprocal and direct space. It is known that the percentage of trial structures that converge to solution for *SnB* is a function of, among other things, the size and complexity of the structure, the space group and the limiting resolution and measured accuracy of data. Generally speaking, structures will become more difficult to determine as N becomes larger and as the resolution decreases to the point of not being able to resolve individual atoms from Fourier maps, *i.e.* $\geq$1.2 Å.

SnB has been successfully applied to a variety of known structures in which $N \geq 300$ non-H atoms. Examples are gramicidin A ($P2_12_12_1$, N = 317, 0.86 Å resolution; Langs, 1988), crambin ($P2_1$, N = 400, 0.83 Å; Weeks *et al.*, 1995), rubredoxin ($P2_1$, N = 497, 1.0 Å; Dauter *et al.*, 1992), toxII ($P2_12_12_1$, N = 624 atoms, 0.96 Å; Smith *et al.*, 1997), lysozyme ($P1$, N = 1200, 0.9 Å; Deacon *et al.*, 1998) and one DNA structure of GCGTA^{M}TACGC at 0.83 Å where ^{M}T is 2-methoxy-3′-methylenephosphonate-thymidine (Egli *et al.*, 2000). Examples of unknown structures solved by direct methods are two forms of vancomycin ($P4_32_12$, N = 258, 0.90 Å; Loll *et al.*, 1997; $P1$, N = 547, 1.0 Å; Loll *et al.*, 1998), α-conotoxin EpI ($I4$, N = 289, 1.10 Å; Hu *et al.*, 1998), E*r*-1 pheromone ($C2$, N = 328, 1.0 Å; Anderson *et al.*, 1998) and α-1 peptide ($P1$, N = 471, 0.90 Å; Privé *et al.*, 1999). To our knowledge, no determination of an macromolecular hybrid structure by direct methods has been reported to date. This structure is also the first macromolecular structure solved with 1.15 Å data just on the resolution edge of what might work with direct methods.

The *CROQUE*[1] phase-refinement program (Langs *et al.*, 2000) uses peak-list optimization (Sheldrick & Gould, 1995), a procedure whereby false peaks in a list from a Fourier map may be detected by a decrease in the value of the correlation coefficient (CC),

$$\mathrm{CC} = \frac{\langle|E_o|^2|E_c|^2\rangle - \langle|E_o|^2\rangle\langle|E_c|^2\rangle}{[(\langle|E_o|^4\rangle - \langle|E_o|^2\rangle^2)(\langle|E_c|^4\rangle - \langle|E_c|^2\rangle^2)]^{1/2}},$$

between the observed and calculated E magnitudes when incorrect sites are included in the calculation of E_c from the list of atoms being tested. This program may allow one to reduce the phase error to 10 or 20° for all data prior to actually inspecting any map to see how well it may fit a structural model.

The Fourier map after *CROQUE* phase refinement of the initial *SnB* solution showed the entire RNA/DNA hybrid structure containing 514 fully occupied non-H atoms including three calcium ions and 103 water molecules.

2. Materials and methods

It is very difficult to grow RNA/DNA hybrid crystals that diffract to high resolution. Initial attempts to crystallize the RNA/DNA hybrid PPT produced only poorly diffracting crystals (4.5 Å). For a long time we were unable to overcome this problem. Gel electrophoresis implicated the purity of the RNA strand as a possible cause of the weakly diffracting crystals. Further purification of this strand was the most important step in obtaining crystals that diffract to high resolution. Crystals of the RNA/DNA hybrid were grown in calcium acetate, spermidine, *n*-octyl-β-D-glucoside, 2,4-methyl-pentanediol (MPD) and sodium cacodylate at pH 6.8 (Kopka *et al.*, 2001, in preparation). They are not isomorphous with DNA/DNA crystals of the corresponding sequence. The DNA/DNA crystals belong to monoclinic space group $C2$, with unit-cell parameters a = 59.82, b = 28.28, c = 72.49 Å, β = 103.88° (Han *et al.*, 1997). However, the RNA/DNA hybrid PPT crystallized in a different space group, orthorhombic $P2_12_12_1$, with unit-cell parameters a = 25.699, b = 41.067, c = 46.113 Å. Data were collected to 1.10 Å at the BNL synchrotron-radiation source facility. Data beyond 1.15 Å with weaker counting statistics were omitted from the *SnB* calculation.

The high-resolution data set collected on this RNA/DNA hybrid allowed us to solve the structure by direct methods.

[1] The original program name *CRUNCH* was changed to *CROQUE* to avoid confusion with another crystallographic program which had the same name.

Table 1 shows crystal parameters and X-ray data-collection statistics. The resolution of the data range from 9.59 to 1.10 Å. The counting statistics of the data for each shell are shown in Table 2.

3. Results

3.1. *Ab initio* phase determination

In all, 17 309 non-redundant diffraction intensities to a minimum interplanar spacing of 1.15 Å were recorded. Normalized structure-factor magnitudes ($|E|$s) were computed using the program *DREAR* (Blessing & Smith, 1999) and were ordered by decreasing magnitude of $|E|$. The 4140 reflections with the largest $|E|$s then were used to generate the corresponding 41 400 triplets having the largest values of A_{HK}.

1000 trial structures were generated and subjected to 414 cycles of the *SnB* procedure. The PPT RNA/DNA molecule contains 414 unique non-H atoms. High-resolution A-DNA and B-DNA structures of this size typically have about 150 and 200 solvent molecules, respectively (Egli *et al.*, 2000). This suggests that about 150 fully occupied water molecules would also be present in the asymmetric unit. Therefore, a total of approximately 600 peaks were selected from the E maps for real-space recycling.

$R(\Phi)$ range	Number of trials	
0.461–0.462	3	*
0.462–0.464	7	**
0.464–0.466	3	*
0.466–0.468	3	*
0.468–0.470	8	**
0.470–0.472	7	**
0.472–0.473	2	*
0.473–0.475	20	*****
0.475–0.477	45	**********
0.477–0.479	48	**********
0.479–0.481	133	**********
0.481–0.483	189	**********
0.483–0.484	211	**********
0.484–0.486	170	**********
0.486–0.488	96	**********
0.488–0.490	22	******
0.490–0.492	20	*****
0.492–0.494	6	**
0.494–0.495	5	**
0.495–0.497	2	*

(*a*)

$R(\Phi)$ range	Number of trials	
0.460–0.462	3	*
0.462–0.463	1	*
0.463–0.465	13	***
0.465–0.467	5	**
0.467–0.468	1	*
0.468–0.470	1	*
0.470–0.472	2	*
0.472–0.474	0	
0.474–0.475	1	*
0.475–0.477	2	*
0.477–0.479	4	*
0.479–0.481	19	*****
0.481–0.482	81	**********
0.482–0.484	142	**********
0.484–0.486	109	**********
0.486–0.488	222	**********
0.488–0.490	183	**********
0.490–0.491	126	**********
0.491–0.493	58	**********
0.493–0.495	13	***
0.495–0.497	8	**
0.497–0.498	5	**

(*b*)

Figure 1
Histograms of the final $R(\Phi)$ values from *SnB* for 1000 trials of the RNA/DNA PPT. (*a*) Using data to 1.15 Å resolution. The distribution separates about 30 possible solutions from the non-solutions. (*b*) Using all data to 1.10 Å resolution. The bimodal distributions separates 26 solutions from non-solutions.

Table 1
Crystal parameters and X-ray data-collection statistics for caaagaaaag/CTTTTCTTTG.

	Crystal parameters	X-ray data statistics	
Space group	$P2_12_12_1$		
Unit-cell parameters (Å)	a = 25.699, b = 41.067, c = 46.113		
Resolution cutoff (Å)		1.15	1.10
No. of unique reflections		17309	19181
Completeness (%)		96.1	93.5
R_{sym}† (%)		8.0	8.1

† $R_{sym} = \sum|I - \langle I\rangle|/I$.

414 *SnB* cycles produced 15–30 possible solutions out of the 1000 trial structures, giving a success rate of ∼1.5–3.0% for the complete 1.15 Å PPT RNA/DNA data set. Histograms of the final $R(\Phi)$ values for 1000 PPT RNA/DNA trials for the compete 1.15 Å data set are shown in Fig. 1(*a*) and also for all data to 1.10 Å in Fig. 1(*b*). One of the most successful trials at 1.15 Å data, with a R_{min} of 0.461 in Fig. 1(*a*), was examined in detail. However, it was not possible to interpret the peaks using the *SnB* graphics display, a feature of the program which displays atoms from the strongest map peaks. It was expected that the P atoms of the phosphate backbone would show up as the 18 highest peaks. However, it was very difficult to distinguish between the atoms of the phosphate backbone and the stacked planes of the bases. Only broken fragments of nucleic acid could be recognized and the connectivity was poor.

For this reason, it was decided to examine the E map and F_o map directly at this stage using the program *O* version 7.0 (Jones *et al.*, 1991). Map phases were calculated using the top 600 peaks. The E map and F_o map at the G5·C16 base-pair region are shown in Figs. 2(*a*) and 2(*b*). The 1.0σ level of these maps looked promising, as there were too many chemically meaningful fragments for it to be incorrect.

3.2. *CROQUE* phase refinement

This *SnB* solution was then subjected to *CROQUE* phase refinement. This program allowed us to blindly take the largest peaks from the map and refine this starting model without the need to assign atom types or check the geometrical acceptability of the fragments that were present. The refinement was initiated with the top 600 *SnB* atoms in the PEAKS file. Six refinement cycles were performed during which time the correlation coefficient (CC) for each cycle increased from 0.6230 to 0.7244. Difference maps were then calculated to locate missing atoms. After several more cycles CC increased to 0.7791.

At this stage, the Fourier map was examined. The map clearly showed an A-like RNA/DNA hybrid molecule. A total of 502 fully occupied non-H atoms including 103 water molecules were successfully located. The Fourier map at the same region as in Figs. 2(*a*) and 2(*b*) is shown in Fig. 2(*c*). Three phosphate O atoms and 15 atoms (mainly from the

sugar linkage) were not located in the Fourier map at this stage, probably owing to the higher mobility of these backbone atoms compared with that of the bases. Examples of this missing region are shown in Figs. 3(*a*) and 3(*b*).

Before starting structural refinement using the peaks from the Fourier map, we went back to test whether refining individual isotropic temperature factors for the PEAKS file would improve the accuracy of the *CROQUE* phases and pick up the missing atoms. The program *CROQUE* (Langs *et al.*, 2000) was used to perform this task, *i.e.* to refine individual isotropic B values by least squares, with the constraint that 'refined' xyz coordinates could be obtained from successive maps as refinement proceeded. Peaks for which $B_{iso} > 60$ for any cycle were rejected prior to the next cycle of refinement. After four least-squares cycles CC increased from 0.7791 to 0.8732 and seven missing atoms (O5′A2, C4′G5, C3′G5, C2′A8, C5′T12, O2PT14 and O1PG20) were located in the next difference map. With one more round of *CROQUE* refinement the CC reached 0.8787 and five additional atoms (C4′A2, O1PA4, C5′G5, C3′T15 and C5′G20) were located. A total of 502 + 7 + 5 = 514 non-H atoms were then submitted to least-squares structural refinement.

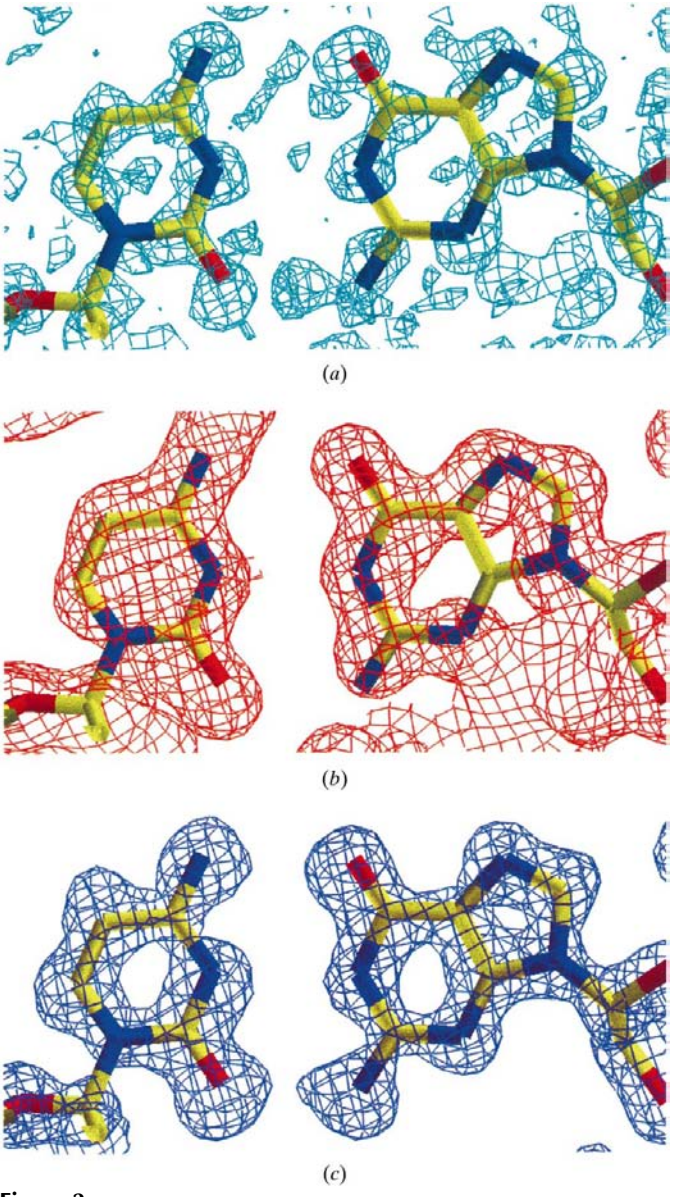

Figure 2
(*a*) *E* map and (*b*) F_o map generated by using 600 peaks from *SnB*. (*c*) The same F_o map region after *CROQUE* phase refinement.

Table 2
X-ray data counting statistics.

	$I/\sigma(I) > 20$		Total	
Resolution (Å)	No. of reflections	% of reflections	No. of reflections	Completeness (%)
100.00–2.71	1219	80.7	1398	92.5
2.71–2.15	1234	87.7	1396	99.2
2.15–1.88	1184	85.6	1364	98.6
1.88–1.71	1104	80.7	1347	98.5
1.71–1.59	1006	72.6	1351	97.5
1.59–1.49	848	62.2	1324	97.1
1.49–1.42	732	53.9	1311	96.6
1.42–1.36	604	45.1	1291	96.5
1.36–1.30	505	37.1	1298	95.4
1.30–1.26	429	31.9	1281	95.4
1.26–1.22	388	29.1	1262	94.7
1.22–1.18	301	22.1	1288	94.6
1.18–1.15	199	15.0	1245	93.9
1.15–1.13	140	10.5†	1143	85.4†
1.13–1.10	26	1.9†	882	65.7†

† Weaker counting statistics.

3.3. Least-squares refinement of the model

A total of 514 atoms, including three Ca^{2+} ions, were refined by least-squares using *SHELXL* (Sheldrick & Schneider, 1997). The R factor from the initial first 50 cycles of positional and isotropic B-factor refinement using the conjugate-gradient method was 22.9% (R_{free} = 0.248). Six missing atoms (C5′A2, C3′A2, O5′A3, C5′C11, C5′T14 and C4′G20) appeared clearly in the difference Fourier map at this stage. Four rounds of refinement were performed and in each round more water molecules were included where difference electron density showed a 3σ peak with stereochemically reasonable hydrogen bonds. Using all data, the R factor of the present model is 0.143 (R_{free} = 0.186) after positional refinement and anisotropic B-factor refinement for the RNA/DNA molecule and isotropic B-factor refinement for the solvent molecules. This model contains 562 atoms including 414 from the RNA/DNA helix dimer, three calcium ions, one MPD and 137 waters. Fig. 4 shows the crystal packing of the RNA/DNA hybrid caaagaaaag/CTTTTCTTTG. Full analysis of the structure will be presented elsewhere (Kopka *et al.*, in preparation).

4. Discussion

So far, it has been impossible to solve large structures using the *SnB* procedure if the resolution of the data is worse than 1.2 Å. Our data at 1.15 Å are about at the limit of what might work with direct methods, according to the database of macromolecular structures solved by these methods thus far. However, we were able to solve the structure successfully by

ab initio direct methods using *SnB* version 2.0 and the *CROQUE* phase-refinement procedure.

The results presented above show that direct methods, as implemented in the *SnB* procedure, can solve unknown crystal structures containing as many as 562 non-H atoms in the asymmetric unit using data as low as 1.15 Å resolution. This example is one of the largest previously unsolved structures determined by *ab initio* direct methods at this resolution as taken from the *SnB* web page (http://www.buffalo.edu/SnB/SnBsuccesses.htm). It is important to note that although the correct *SnB* solution was identified on the basis of a minimum function $R(\Phi)$, additional phase refinement by *CROQUE* was required to improve the interpretability of those solutions.

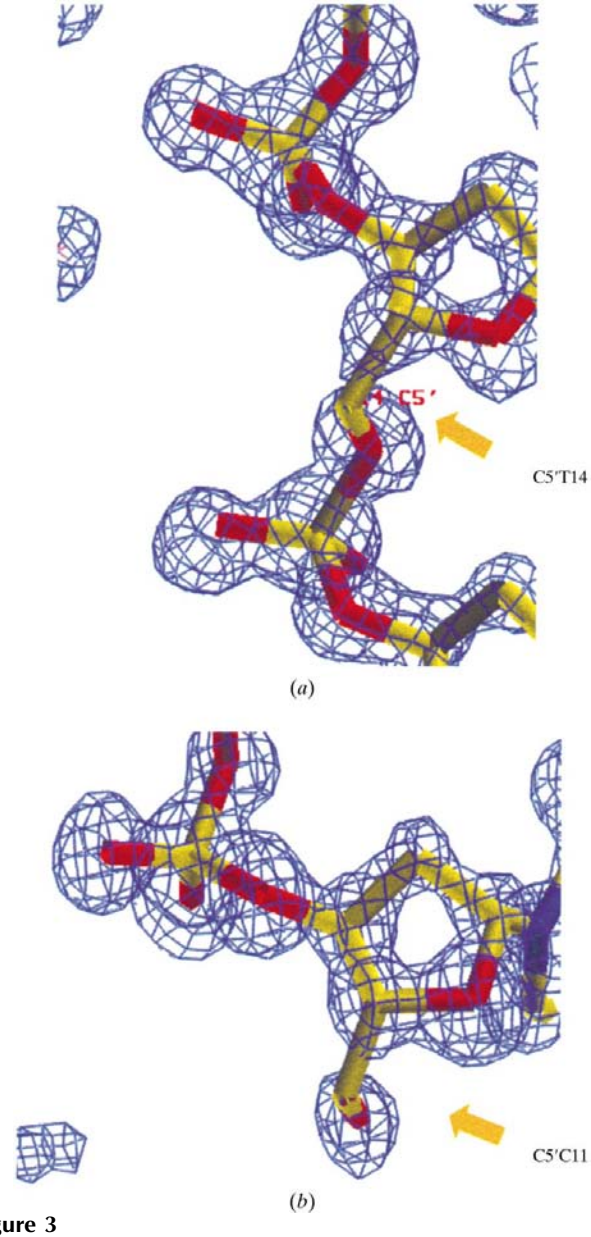

Figure 3
Illustration of the quality of the *CROQUE* F_o map at sugar and phosphate backbone regions prior to least-squares refinement. The electron density was missing from atoms (*a*) C5′T14, (*b*) C5′C11 after *CROQUE* phase refinement. Arrows indicate missing electron density around the C5′ atoms. Models are from the final refinement.

Until recently, no DNA structure was included among the low-to-medium molecular weight macromolecular structures solved by direct methods. The first DNA crystal structure solved by these methods was GCGTA^{M}TACGC (^{M}T is 2-methoxy-3′-methylenephosphonate-thymidine) at 0.83 Å (Egli *et al.*, 2000). In their original *SnB* map, 260 DNA atoms (out of a total of 408), 20 water molecules and one Mg^{2+} ion were located. In comparison, 95.7% of the RNA/DNA atoms which were located in our 1.15 Å *SnB* map improved by *CROQUE* phase refinement, implying that *CROQUE* was vital to this work.

As mentioned earlier, the success rate (the percentage of trial structures that are solved by direct methods) is known to be dependent on the space group. In all the test structures with *SnB*, it has been shown that there is usually a lower success rate for high-symmetry structures than those for *P*1. Most large $P2_12_12_1$ structures have less than 1% success rate. This is because an *N*-atom fragment in $P2_12_12_1$ has a smaller chance

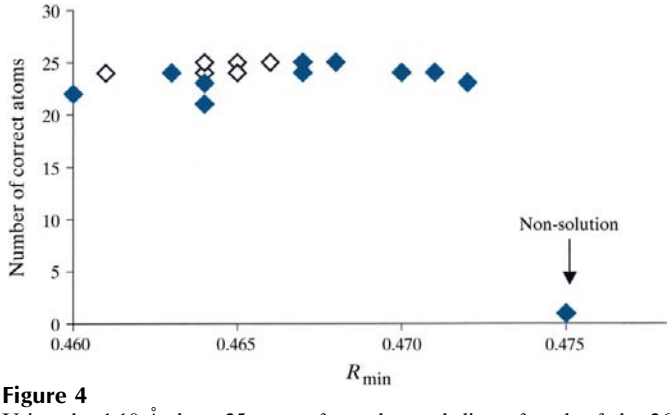

Figure 4
Using the 1.10 Å data, 25 atoms from the peak lists of each of the 26 solutions shown in Fig. 1(*b*) were compared with the refined structure. The number of correct atoms in each of the trials (R_{min} = 0.460–0.475) are plotted as a function of R_{min}. Atoms within 0.25 Å distance from the refined structure are considered correct (Weeks *et al.*, 1995). A total of 27 trials (26 solutions plus one non-solution) were plotted. Closed diamonds represent a single solution and open diamonds represent more than two solutions.

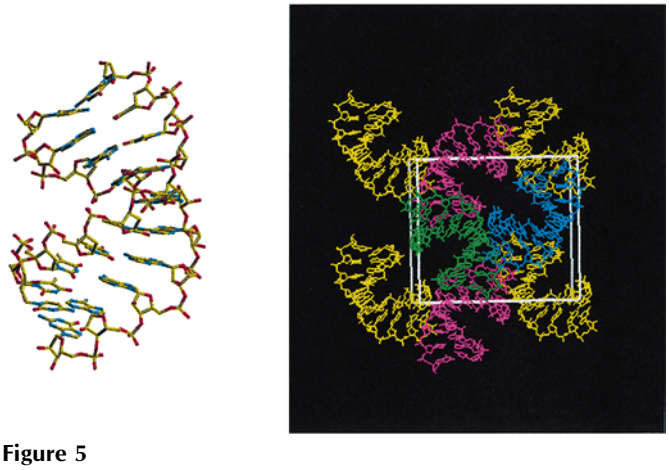

Figure 5
Left: final refined structure of the RNA/DNA hybrid of PPT indicating an A-helix conformation. Right: crystal packing of the hybrid in the unit cell, viewed along the *a* axis, onto the *bc* plane. Each colored molecule represents one of the four asymmetric units in this space group.

of fitting any one of 16 choices of origins compared with an infinite number of 'correct' structures for *P*1. However, our success rate of ∼1.5–3% (Fig. 1*a*) for the complete 1.15 Å data set is higher than those of most large $P2_12_12_1$ structures.

After the structure was solved, we used the data in the 1.15–1.10 Å resolution range, which has weaker counting statistics, to see if this improved the *SnB* distribution of solution from non-solutions (Table 2). The success rate was ∼2.6% (Fig. 1*b*). As the histograms in Figs. 1(*a*) and 1(*b*) show, clear bimodal distribution is seen in the higher resolution data even though it has weaker counting statistics (Fig. 1*b*). All of the 26 solutions in Fig. 1(*b*) were confirmed as true solutions. To do this, 25 atoms from the top peaks list were compared with the refined structure. In all true solutions, atoms were identified as correct when compared with the refined structure if the atoms were within 0.25 Å of the refined model. 21–25 atoms (84–100%) were correctly positioned within this range. Although the difference in $R_{\rm min}$ between solutions and non-solutions is quite small, there is a clear distinction between solutions and non-solutions as shown in Fig. 4. Only one correct atom was found from the 25 top peaks for a non-solution at $R_{\rm min}$ = 0.475. A 'postmortem' analysis showed that the mean phase error[2] of the original *SnB* model was 48° and that it was reduced to 29° by *CROQUE* prior to fitting the map.

A comparison of the same decamer sequences in DNA/DNA and RNA/DNA shows that the decamers are structurally different and have different crystal packing. The hybrid crystal solved here is an A-helix while the DNA/DNA helix was a B-helix, although one with the deeper and narrower minor groove as is expected in A-tracts or runs of consecutive A·T base pairs (Han *et al.*, 1997). This DNA/DNA PPT analogue packs end to end forming an infinite helix in the crystal. In contrast, the crystal packing of the RNA/DNA hybrid decamer is very typical of an A-helix, with terminal base pairs of symmetry-related molecules packed against one wall of the neighboring minor groove. The structure and crystal packing are shown in Fig. 5.

It has been demonstrated kinetically that long-chain drugs that bind within the minor groove of B-DNA are capable of inhibiting HIV reverse transcriptase (Filipowsky *et al.*, 1996). It also has been experimentally shown that netropsin and distamycin do not bind to A-form helices, but induce an A to B transition with DNA, RNA and the hybrid rA·dT (Ivanov *et al.*, 1974; Zimmer *et al.*, 1982). We expect that the addition of the drugs to the RNA/DNA hybrid decamer may induce a more B-like conformation, making it a prime site for enzyme inhibition by drugs, such as the tomaymycin–netropsin and tomaymycin–distamycin bis-linked drugs.

The author is indebted to Professor R. E. Dickerson for supporting work on this project and Mary L. Kopka for dedicated help throughout this polypurine project. The author also greatly appreciates Dr David Langs of the Hauptman–Woodward Medical Research Institute, Inc. for help in using the *CROQUE* program developed under NIH grant GM-46733. His expert advice and encouragement were crucial to solve the structure. Special thanks to Michael Sawaya for reading this manuscript and providing helpful comments and suggestions. The author would also like to thank Charles Weeks of the Hauptman–Woodward Medical Research Institute, Inc. for help in retrieving the solution trials using *SnB* for a 'postmortem' analysis. Thanks to Sarvenaz Saadat for help in the purification of the DNA strand. This work was supported by NIH Grant GM-31299 and Amgen Fellowship from UCLA Center for AIDS Research.

[2] Mean phase error = $\sum(F_{\rm obs} * \Delta{\rm phase})/\sum F_{\rm obs}$.

References

Anderson, D. H., Weiss, M. S. & Eisenberg, D. (1996). *Acta Cryst.* D**52**, 469–480.
Blessing, R. H. & Smith, G. D. (1999). *J. Appl. Cryst.* **32**(4), 664–670.
Dauter, Z., Sieker, L. C. & Wilson, K. S. (1992). *Acta Cryst.* B**48**, 42–59.
Deacon, A. M., Weeks, C. M., Miller, R. & Ealick, S. E. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 9284–9289.
Egli, M., Tereshko, V., Teplova, M., Minasov, G., Joachimiak, A., Sanishvili, R., Weeks, C. M., Miller, R., Maier, M. A., An, H., Cook, P. D. & Manaharan, M. (2000). *Biopolymers (Nucleic Acids Sci.)*, **48**(4), 234–252.
Filipowsky, M. E., Kopka, M. L., Zison, M., Clark, P., Hughes, S. H., Lown, J. W. & Dickerson, R. E. (1996). *Biochemistry*, **35**, 15397–15410.
Han, G. W., Kopka, M. L., Cascio, D., Grzeskowiak, K. & Dickerson, R. E. (1997). *J. Mol. Biol.* **269**, 811–826.
Hu, S.-H., Loughnan, M., Miller, R., Weeks, C. M., Blessing, R. H., Alewood, P. F., Lewis, R. J. & Martin, J. L. (1998). *Biochemistry*, **37**, 11425–11433.
Ivanov, V. I., Minchenkova, L. A. & Minyat, E. E. (1974). *J. Mol. Biol.* **87**, 817–833.
Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Koo, H.-S., Wu, H.-M. & Crothers, D. M. (1986). *Nature (London)*, **320**, 501–506.
Kopka, M. L., Han, G. W., Lavelle, L. & Dickerson, R. E. (2001). In preparation.
Langs, D. A. (1988). *Science*, **241**, 188–191.
Langs, D. A., Blessing, R. H. & Smith, G. D. (2000). *J. Appl. Cryst.* **33**, 174–175.
Loll, P. J., Bevivino, A. E., Korty, B. D. & Axelsen, P. H. (1997). *J. Am. Chem. Soc.* **119**, 1516–1522.
Loll, P. J., Miller, R., Weeks, C. M. & Axelsen, P. H. (1998). *Chem. Biol.* **5**, 293–298.
Miller, R., Gallo, S. M., Khalak, H. G. & Weeks, C. M. (1994). *J. Appl. Cryst.* **27**, 613–621.
Privé, G. G., Anderson, D. H., Wesson, L., Cascio, D. & Eisenberg, D. (1999). *Protein Sci.* **8**(7), 1400–1409.
Schäfer, M., Schneider, T. R. & Sheldrick, G. M. (1996). *Structure*, **4**, 1509–1515.
Sheldrick, G. M. & Gould, R. O. (1995). *Acta Cryst.* B**51**, 423–431.
Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **277**, 319–343.
Smith, G. D., Blessing, R. H., Ealick, S. E., Fontecilla-Camps, J. C. Hauptman, H. A., Housset, D., Langs, D. A. & Miller, R. (1997). *Acta Cryst.* D**53**, 551–557.
Weeks, C. M., DeTitta, G. T., Hauptman, H. A., Thuman, P. & Miller, R. (1994). *Acta Cryst.* A**50**, 210–220.
Weeks, C. M., Hauptman, H. A., Smith, G. D., Blessing, R. H., Teether, M. M. & Miller, R. (1995). *Acta Cryst.* D**51**, 33–38.
Weeks, C. M. & Miller, R. (1999). *J. Appl. Cryst.* **32**, 120–124.
Zimmer, C., Kakiuchi, N. & Guschlbauer, W. (1982). *Nucleic Acids Res.* **10**(5), 1721–1732.