

# An Unusual Sugar Conformation in the Structure of an RNA/DNA Decamer of the Polypurine Tract May Affect Recognition by RNase H

Mary L. Kopka<sup>1\*</sup>, Laurence Lavelle<sup>2</sup>, Gye Won Han<sup>1</sup>, Ho-Leung Ng<sup>1</sup>  
and Richard E. Dickerson<sup>1,2</sup>

<sup>1</sup>Molecular Biology Institute  
University of California, Los  
Angeles, CA 90095-1570, USA

<sup>2</sup>Department of Chemistry and  
Biochemistry, University of  
California, Los Angeles, CA  
90095-1570, USA

Retroviral conversion of single-stranded RNA into double-stranded DNA requires priming for each strand. While host cellular t-RNA serves as primer for the first strand, the viral polypurine tract (PPT) is primer for the second. Therefore, polypurine tracts of retroviruses are essential for viral replication by reverse transcriptase (RT). These purine tracts are resistant to cleavage during first strand synthesis. In obtaining the primer for second strand synthesis, the RNase H function of RT must cleave the PPT exactly for *in vivo* transcription to proceed efficiently and proper integration to occur. At the RNase H active site the protein makes contacts primarily along the backbone, with hydrogen bonds to the sugar–phosphate oxygen atoms. A high-resolution structure (1.10 Å) of the first ten base-pairs of the RNA/DNA hybrid PPT, r-(c-a-a-g-a-a-a-g)/d-(C-T-T-T-C-T-T-T-G), contains the highly deformable r-(a-g-a) steps found in retroviral polypurine tracts. This r-(a-g-a) motif is utilized in the “unzipping” or unpairing of bases that occurs when RT binds a malleable PPT. Another unusual feature found in our high-resolution PPT structure is the sugar switch at RNA adenine 2. All the RNA sugars are the expected C3'-endo, except sugar 2, which is C2'-endo, characteristic of B-form sugars. This local A-to-B conversion adversely affects the pattern of hydrogen bonds from protein to sugar–phosphate backbone, disrupting the catalytic site. Disruption could cause the enzyme to pause at the 5'-end of the PPT, leaving it intact. Pyrimidine–purine (YR) steps are most deformable and the T-A step especially can undergo A-to-B transitions readily.

© 2003 Elsevier Ltd. All rights reserved.

**Keywords:** polypurine tract; RNA/DNA hybrid; RNase H; recognition; U-box

\*Corresponding author

## Introduction

Before the human immunodeficiency virus (HIV) can successfully infect its host, the viral RNA must be copied into DNA. To accomplish this task the virus utilizes a multi-functional enzyme, reverse

transcriptase (RT), that converts single-stranded viral RNA into double-stranded DNA (polymerase function) and then degrades the RNA template (RNase H function). The three X-ray structures of HIV-RT complexed with a drug, a bound nucleic acid and the unliganded enzyme<sup>1–3</sup> showed the molecule to have a “hand” motif similar to that found in other polymerases, with fingers, palm and thumb to guide the nucleic acid to the two active sites. Two “primer grip” regions of the molecule have been defined, the polymerase grip and the RNase H primer grip.<sup>4–6</sup> These grip regions are situated adjacent to both active sites.

The RNase H domain has both endonuclease and 3',5'-exonuclease activity.<sup>7</sup> During (–)-strand

L.L. & G.W.H. contributed equally to this work.

Present address: H.-L. Ng, Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720-3206, USA.

Abbreviations used: PPT, polypurine tract; RT, reverse transcriptase.

E-mail address of the corresponding author:  
kopka@mbi.ucla.edu

synthesis, RNase H degrades the RNA template (exonuclease activity) as the DNA strand is being copied, but leaves intact (endonuclease activity) a purine-rich run of 15 bases (Figure 1(A)) known as the polypurine tract.<sup>8</sup> This PPT serves as the primer for second or (+)-strand DNA synthesis. The RNase H subunit, consisting of approximately 120 + amino acid residues, resides in the p66 C-terminal domain of the p66/p51 RT heterodimer, and its crystal structure has been reported.<sup>9</sup> The RNase H primer grip region makes extensive contacts with the DNA primer,<sup>5</sup> and mutations of amino acids in both the active-site RNase H grip regions have been examined.<sup>10–12</sup> The structure of RT complexed with an RNA/DNA primer-template containing the PPT<sup>5</sup> has shown that contacts between the RNA/DNA hybrid bases, and amino acid residues in the RNase H primer grip and its catalytic center, occur primarily with the sugar-phosphate backbone. There are relatively few hydrogen bonds with the hybrid bases. Only three direct protein contacts (H-bonds) to the bases occur, two of which are found in the RNase H active site where the scissile bond is cleaved.

Knowledge of the structure of nucleic acid duplexes utilized by RT is important to understanding the multi-step process of reverse transcription. Structural biologists have investigated by NMR and crystallography various duplex structures that the enzyme contacts. To initiate copying of the minus strand at the start of reverse transcription, the viral RNA recruits a host tRNA forming an RNA/RNA duplex. However once (–)-strand DNA synthesis begins, a chimeric hybrid, an Okazaki fragment, is formed at the junction of the growing DNA chain with the tRNA/RNA. This junction has been studied by both crystallography and NMR,<sup>13,14</sup> as has the Okazaki fragment generated at the completion of synthesis when the (–)-strand primer is removed from the RNA/

DNA junction.<sup>15</sup> Another decamer RNA/DNA hybrid beginning one RNA nucleotide 5' of the cleavage site has been looked at crystallographically.<sup>16</sup> All of these structures are related to the template-primer for (–)-strand DNA synthesis.

Unusual backbone conformations have been seen by both crystal and solution studies in base-pairs at these Okazaki junctions recognized by RT.<sup>13–15</sup> The NMR structures have a bend in the junction region.<sup>14,15</sup> This conformational backbone flexibility, primarily at one base removed in the 5' direction from the junction, has been proposed as a recognition signal for RNase H endolytic cleavage.<sup>13</sup> Szyperski *et al.*<sup>14</sup> agree that the backbone conformations in this region are important. They make particular note of a sugar pucker variation at the hybrid junction and a surrounding base-step, where two of the DNA sugars interconvert between C3'-endo and C2'-endo.

The template-primers for both plus and minus strand are resistant to hydrolytic cleavage until DNA synthesis of the two strands is near completion, prior to integration. To obtain the proper primer for (+)-strand synthesis, the RNase H must cleave the PPT accurately for complete transcription to proceed efficiently.<sup>8,17–19</sup> The binding and endolytic cleavage of the PPT by RNase H have been the subject of many reports.<sup>5,11,12,17,20–22</sup> Only one PPT sequence has been investigated structurally (by NMR), the 3'-end of the PPT along with four bases in the immediate flanking sequence. However, these investigators<sup>23</sup> have substituted an adenine base, e.g. *gagg* for *gggg*, in a critical region in the guanine tract that is essential for proper extension, cleavage and primer removal.<sup>20</sup>

In NMR studies of hybrids the DNA strands have B-like sugar conformations, e.g. O4'-endo, C1'-exo, C2'-endo, while RNA sugars are C3'-endo.<sup>14,15,21,24,25</sup> This has not been the case for the crystal structures. Generally, these structures have been A-like helices with both RNA and DNA strands having C3'-endo sugar conformations. However, exceptions in sugar conformation at individual base-steps have been seen crystallographically, as well as other conformational variations along the backbones, such as  $\alpha\beta\gamma$  crankshaft transitions.<sup>13,16,26</sup> The NMR solution studies note a difference in groove widths from standard A and B-form helices, with most hybrids having a minor groove width intermediate between A and B. The crystal structures, on the other hand, tend to have A-like minor groove widths of 9–10 Å.<sup>13,16,26</sup>

This report focuses on the questions: what specific features of the hybrid RNA/DNA PPT make it resistant to RNase H digestion, and how does RNase H recognize the proper 5'-cleavage site to define adequately the 5'-end of the PPT? In the RT-hybrid complex, the "unzipping" of the 5' A-tract<sup>5</sup> and a narrower minor groove in the A-tracts<sup>5,21,27</sup> have been reported as features of the



**Figure 1.** (A) The HIV-1 hybrid polypurine tract sequence (bold). RNA is shown in lower case, DNA in upper case. Also shown are the 5' leading "U-tract", and two base-pairs 3' to the PPT. Bold arrows indicate the position where RNase H cuts to remove the PPT. Grey arrow indicates start of "U- or T-box". (B) The sequence of the RNA/DNA decamer analyzed in this report, where the first adenine is replaced by a cytosine to improve crystal packing. Numbering of bases is indicated.

PPT important in recognition. In the high-resolution (1.10 Å) structure reported here, we look at the first ten bases of the polypurine tract (Figure 1(B)) in which the first 5'-adenine base has been replaced by cytosine to improve crystal packing.

## Results

The high-resolution (1.10 Å) crystals of the hybrid r-(c-a-a-a-g-a-a-a-g)-d-(C-T-T-T-T-C-T-T-T-G) using two cations, Mg and Ca, at two pH values, and solved by two different methods, molecular replacement (this work) and direct methods<sup>28</sup> provide a look at the first ten bases of the PPT at atomic resolution. These crystals contain the two A-tracts, but not the G-tracts found at the 3'-end of the PPT. As is commonly done to facilitate good stacking and packing contacts in the crystal, there is a substitution of cytosine for the first adenine base at the 5'-end (for a view of packing contacts, see Figure 5 of Han<sup>28</sup>).

The Ca-crystal solved by molecular replacement will be the structure to which the other two structures are compared. All three structures are essentially the same. The rms fit between structures is quite small, 0.231 Å for Mg-crystal and 0.137 Å for the direct method solution of the Ca-crystal (Table 1). Both molecular replacement solutions exhibited alternate conformations along the backbone at sugar-phosphate 2–3 and 19–20. The Ca-crystal also had an alternate conformation at O5' on base 11. The unusual C2'-endo sugar pucker at sugar 2 on the RNA purine strand and the deformable a-g-a steps are the defining structural features of this hybrid helix (Figure 2). One can see that the hybrid is A-like with a high degree of bending on the bottom half of the helix. Two sugars, adenine 2 and guanine 20, switch conformation from C3'-endo, as found elsewhere throughout the structure, to a C2'-endo conformation. The

sugar switch at sugar 2 is very apparent in the global view of the hybrid, where the gold-colored O2' atom (Figure 2) now sits in the major groove, not readily accessible for recognition by RNase H.

## Helix parameters

Nucleic acid helix parameters are indicative of helical type, determining whether a helix falls into the A, B or Z family. With the PPT hybrid, most helical values are typical of an A-form helix, which has a large roll, inclination and minor groove width, with negative slide and negative x-displacement. The following are average values for this hybrid, with A-helix values given in parentheses for comparison: roll = +6.65° (+9° to +10°), inclination = +17.2° (+20°), slide = -1.35 Å (-1.5 to -1.7 Å), x-displacement = -4.2 Å (-4.5 Å), rise = 2.7 Å (2.7–2.9 Å). The minor groove width across the phosphate groups, subtracting 5.8 Å for two phosphate group radii, is 9.95 Å (11.8 Å). The distance measured across the groove from O4' to O4' on the sugar rings is much less, 6.1 Å (7.5 Å). This alternation in groove dimensions, wide across the phosphate groups and narrow between the sugar O4' atoms, is typical of A-form helices. Thus far the parameters of these first ten bases of the PPT point to an A-type helix. But other parameters, the base-pair normal vectors and sugar-phosphate conformations, indicate local differences.

By plotting the orientation of normal vectors (vectors perpendicular to the base-pairs) it is possible to differentiate between different helical types, as well as determine whether the helix is bent or straight.<sup>29</sup> In an ideal A-helix of sequence c-a-a-a-g-a-a-a-g, the base-pairs are inclined 15–20° to the helical axis, so the tips of their vectors sweep around the axis in a writhe (Figure 3A, ■). Because base-pairs in B-DNA are approximately perpendicular to the helix axis, their normal vectors cluster near the origin at 0.0,

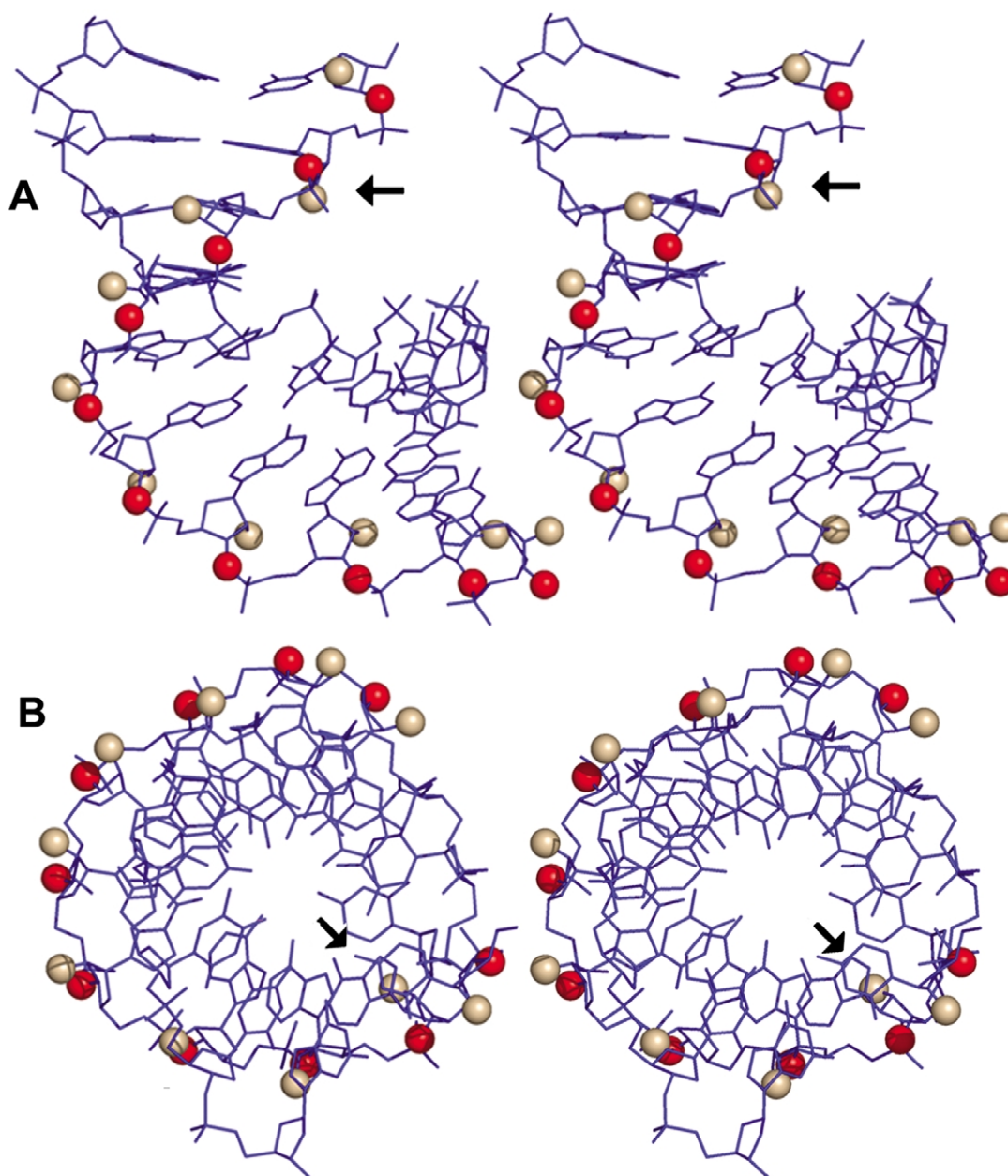
**Table 1.** Data collection and refinement

Cation	Mg	Ca	Ca <sup>28</sup>
Refinement method	Molecular replacement	Molecular replacement	Direct methods
Space group	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Unit cell parameters			
<i>a</i> , <i>b</i> , <i>c</i> (Å)	25.60, 40.76, 45.74	25.70, 41.07, 46.12	25.70, 41.07, 46.11
α, β, γ (deg.)	90, 90, 90	90, 90, 90	90, 90, 90
Resolution (Å)	20–1.10	50–1.15	9.6–1.15
Completeness (%) <sup>a</sup>	97.0 (74.6)	96.1 (93.9)	96.1 (93.9)
<i>R</i> <sub>merge</sub> <sup>a</sup>	5.8 (49.2)	8.1 (26.1)	8.1 (26.1)
No. reflections	18,327	17,368	17,309
<i>I</i> /σ <sup>a</sup>	33.77 (4.06)	29.61 (11.09)	29.61 (11.09)
<i>R</i> <sub>cryst</sub> ( <i>R</i> <sub>free</sub> ) <sup>a,b</sup>	11.7 (15.6)	12.7(15.8)	14.3 (18.6)
Average <i>B</i> -factor (Å <sup>2</sup> )	13.76	10.71	11.07
rms bond length (Å) <sup>c</sup>	0.019 (0.030)	0.009 (0.010)	0.017 (0.020)
rms bond angle (deg.) <sup>c</sup>	0.043 (0.060)	0.027 (0.028)	0.019 (0.040)
rms between structures (Å)	0.231	–	0.137

<sup>a</sup> Parentheses denote statistics in the highest-resolution shell of 1.18–1.15 Å or 1.13–1.10 Å.

<sup>b</sup>  $R_{\text{cryst}} = \sum |F_{\text{obs}} - F_{\text{calc}}| / \sum |F_{\text{obs}}|$ ; where  $F_{\text{obs}}$  and  $F_{\text{calc}}$  are the observed and calculated structure factor amplitudes.  $R_{\text{free}}$  is calculated as  $R_{\text{cryst}}$  but using only 5% of the data, which was not included in the refinement.

<sup>c</sup> Target values (σ) are given in parentheses.

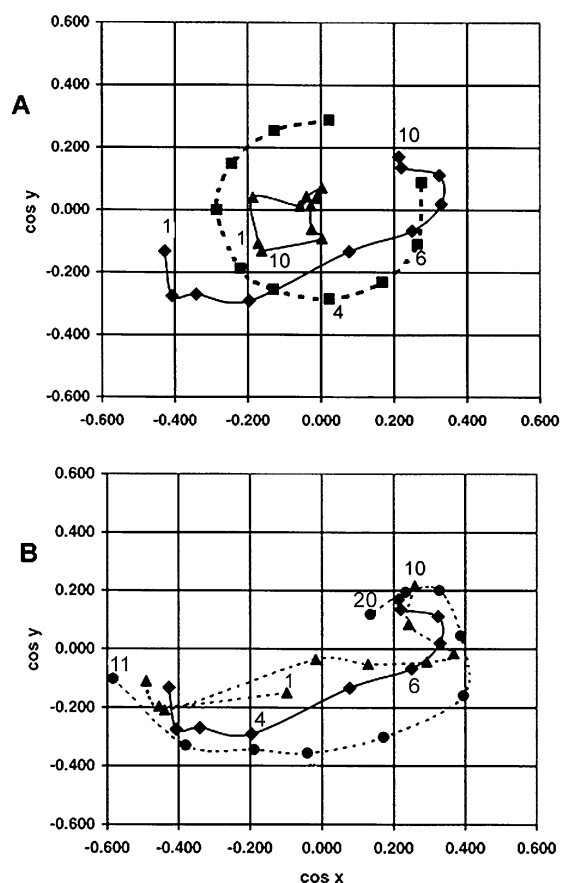


**Figure 2.** A, Stereo view of minor groove of RNA/DNA hybrid. B, Axial view of hybrid. Red spheres along the RNA strand are O3' phosphate oxygens, gold are O2' hydroxyls. Arrows indicate C2'-endo O2' hydroxyl on sugar 2. Figure generated using PYMOL (<http://pymol.sourceforge.net/>).

0.0, as shown in the Han *et al.*<sup>27</sup> structure of the same sequence (Figure 3A, ▲). In this structure, the A-tracts exhibit little or no bend (points 2–4 and 6–9 on the graph). Our PPT hybrid plot (Figure 3A, ◆) reveals: (1) that its two A-tracts (a2-a3-a4 and a6-a7-a8-a9) are almost as straight as in the B-DNA helix of the same sequence with its points 2–4 and 6–9 clustered at opposite ends of the plot; and (2) that a large bend of 14.5° occurs over the a-g-a steps, points 4–6. Roll values (Vrol) are very high at these steps, with the a4-g5 and g5-a6 steps having positive roll angles of 15.97° and 10.27°, respectively. Buckle at g5-a6 and a6-a7 is also high, 18.46° and 16.48°, respectively (see Table 2). The a4-g5 purine step is unstacked,

which was not the case in the DNA/DNA B-helix.<sup>27</sup> This destacking, high propeller, roll and buckle occur in the region of the hybrid where the helix unzips in the Arnold and co-workers protein–PPT complex.<sup>5</sup> The overall writhe in the helix is 44°. Both the bend at the a-g-a steps and the overall writhe are very apparent in Figure 2.

Figure 3B reveals how each strand contributes to the overall normal vector curve. The two individual strands, rR and dY, are shown as dotted lines. Surprisingly, the DNA strand (◆) writhes in a way that smoothes the bend occurring at the a-g-a steps on the RNA strand, and compensates for the straight A-tracts at the two ends of the



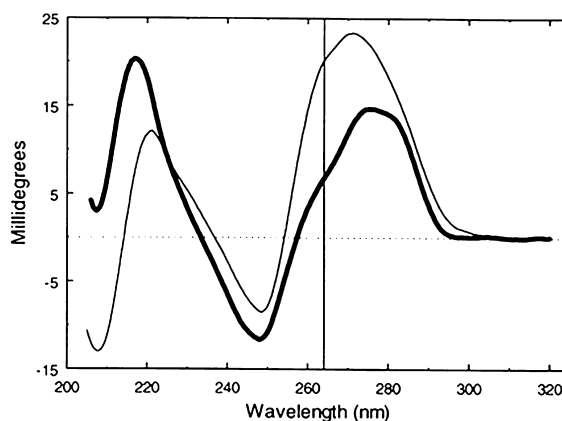
**Figure 3.** A, Normal vector plots of (■) RNA/RNA ideal A-helix, (▲) DNA/DNA B-helix<sup>27</sup> and, (◆) RNA/DNA hybrid helix. B, Repetition of the overall RNA/DNA hybrid helix curve (◆) along with curves for each individual strand (dotted lines): (●) DNA strand (dY), (▲) RNA strand (rR).

RNA (▲). Adenine-over-adenine stacking is the predominant interaction that makes A-tracts straight. The large disjuncture at the a-g-a steps contributes to the overall writhe.

**Table 2.** Helix parameters (°)

	DNA			Hybrid		
	Propeller	Vrol	Buckle	Propeller	Vrol	Buckle
C	-15.57	-8.51	2.23	-12.44	7.95	27.53
A	-11.99	6.49	2.75	-7.66	0.82	-4.07
<b>A</b>	-15.47	0.68	-1.52	<b>-21.19</b>	7.42	-6.84
<b>A</b>	-2.62	<b>2.83</b>	<b>-8.01</b>	<b>-24.76</b>	<b>15.97</b>	-0.92
<b>G</b>	-5.52	-2.04	<b>-10.02</b>	-4.78	<b>10.27</b>	<b>18.46</b>
<b>A</b>	-11.99	-1.36	-1.65	-4.93	6.42	<b>16.48</b>
<b>A</b>	-17.82	-3.32	-2.42	-6.09	5.30	4.76
<b>A</b>	-17.01	-1.92	0.02	-12.53	4.68	1.39
<b>A</b>	-12.48	9.74	-4.54	-3.98	1.03	5.00
<b>G</b>	-14.26	-	-9.71	-6.59	-	-6.64

Nucleic acid parameters calculated with FREEHELIX.<sup>29,42</sup> The RT-PPT parameters are not included in this Table because the available helix analysis programs do not allow for unpaired bases, instead matching each base with its supposed partner whether that partner is paired with it or not. Most global parameters in analysis of the RT-PPT helix would be correct. Base-pair parameters in the region of mispairing are in question.



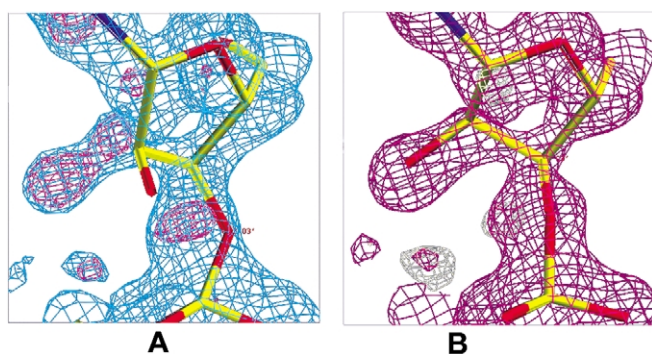
**Figure 4.** CD spectra of decamer duplexes of the PPT sequence. Thin line = RNA/DNA hybrid (this work). Thick line = DNA/DNA analog.<sup>27</sup>

### Circular dichroism

The CD spectra for the DNA duplex and the hybrid are compared in Figure 4. The DNA duplex has a typical B-DNA spectrum with positive peaks at 217 nm and 275 nm, and a negative band at 248 nm.<sup>30</sup> The hybrid duplex has spectral components with both A and B-type conformations. A-form characteristics are a large peak at 271 nm, and the trough at 208 nm characteristic of an r-purine rich strand in a hybrid duplex.<sup>20,30</sup> B-form character is shown by a peak at 221 nm and a negative band at 248 nm. The PPT sequence with two short A-tracts has 70% AT base-pairs. Both the DNA duplex and the hybrid have a shoulder at 264 nm that we interpret as a spectral feature of A-tracts. Sen & Grasslund<sup>31</sup> in their CD spectra of A-tract duplexes have a shoulder at 264 nm, but do not discuss this feature. This shoulder is unlikely to be a structural feature of AT base-pairs in mixed sequences, because d(AAG)<sub>8</sub>:d(CTT)<sub>8</sub> with 67% AT and no A-tracts has a classical B-DNA CD spectrum with no shoulder in this region.<sup>30</sup> Nor can this shoulder be attributed to nearest-neighbor interactions, because d(C-A-A-A-G-A-A-A-A-G)-d(C-T-T-T-T-C-T-T-T-G) and d(AAG)<sub>8</sub>:d(CTT)<sub>8</sub> have similar nearest-neighbor interactions.

### Backbone and sugar conformation

The single most striking feature of the hybrid helix is the adenine 2 ribose sugar switch from O3'-endo to O2'-endo. This B-DNA sugar conformation was found independently in all three structure analyses, the Ca and Mg crystals solved by molecular replacement and the Ca-crystal solved *a priori* by direct methods.<sup>28</sup> In the Conn *et al.*<sup>26</sup> hybrid model used for molecular replacement, all sugars were in the C3'-endo conformation. Difference maps ( $F_o - F_c$ ) indicated clearly that the ribose O2' atom on adenine 2 had to be moved (Figure 5A). In repositioning the O2' oxygen atom



**Figure 5.** A, Electron density of adenine sugar 2. Blue,  $2F_o - F_c$  density ( $1\sigma$ ) and pink,  $F_o - F_c$  density ( $3\sigma$ ), both indicating the incorrectness of the C3'-endo sugar conformation. B, Density after changing to a C2'-endo sugar and refinement.

into the density, the sugar pucker converted from C3' to C2'-endo (Figure 5B). This C2'-endo pucker is not caused by crystal packing contacts. On the opposing DNA strand, all sugars are in the C3'-endo conformation except guanine sugar 20, which has a C2'-endo pucker.

At high resolution alternate conformations are often indicated in the electron density. In this structure, alternate conformations are found along the backbone at RNA bases 2, 3 and 8, and DNA base 20. The major conformers at these positions have the following characteristics: The phosphate backbone angles,  $\epsilon$  and  $\zeta$ , at RNA sugar 2 are in the BII conformation. Adoption of the BII conformation requires that the sugar ring have a C2'-endo conformation.<sup>32,33</sup> Sugar 20 on the DNA strand also has a C2'-endo conformation. Along the backbone at bases 3 (RNA) and 20 (DNA) the alternate conformers try to achieve an extended conformation ( $t, t, t$ ) for the  $\alpha\beta\gamma$  angles, with two angles *trans* and the third angle tending towards *trans* but still in the  $g^+$  region. RNA sugar 8 has a completely extended backbone in which the  $\alpha\beta\gamma$  angles, the crankshaft linkage<sup>32,33</sup> are  $t, t, t$  rather than the standard  $g^-, t, g^+$ . Horton & Finzel<sup>16</sup> observed crystallographically, and Cheatham & Kollman<sup>34</sup> by molecular dynamics, that an unusual backbone conformation on one strand is matched frequently by an extended conformation one base-pair away on the opposite strand. If the strain is not released on the opposite strand, bending to relieve the strain is a possibility. In our PPT hybrid, the conformer switch from C3' to C2'-endo at adenine sugar 2 is matched by a similar reversal at guanine sugar 20. The position of the phosphate group within a two base-pair frame,  $Z_p$ , also indicates a B-like conformation at base-pairs 1–20 (0.23 Å) and 2–19 (0.93 Å), with A-form values  $> 1.5$  Å.<sup>35</sup>

### Deformability of the a-g-a steps

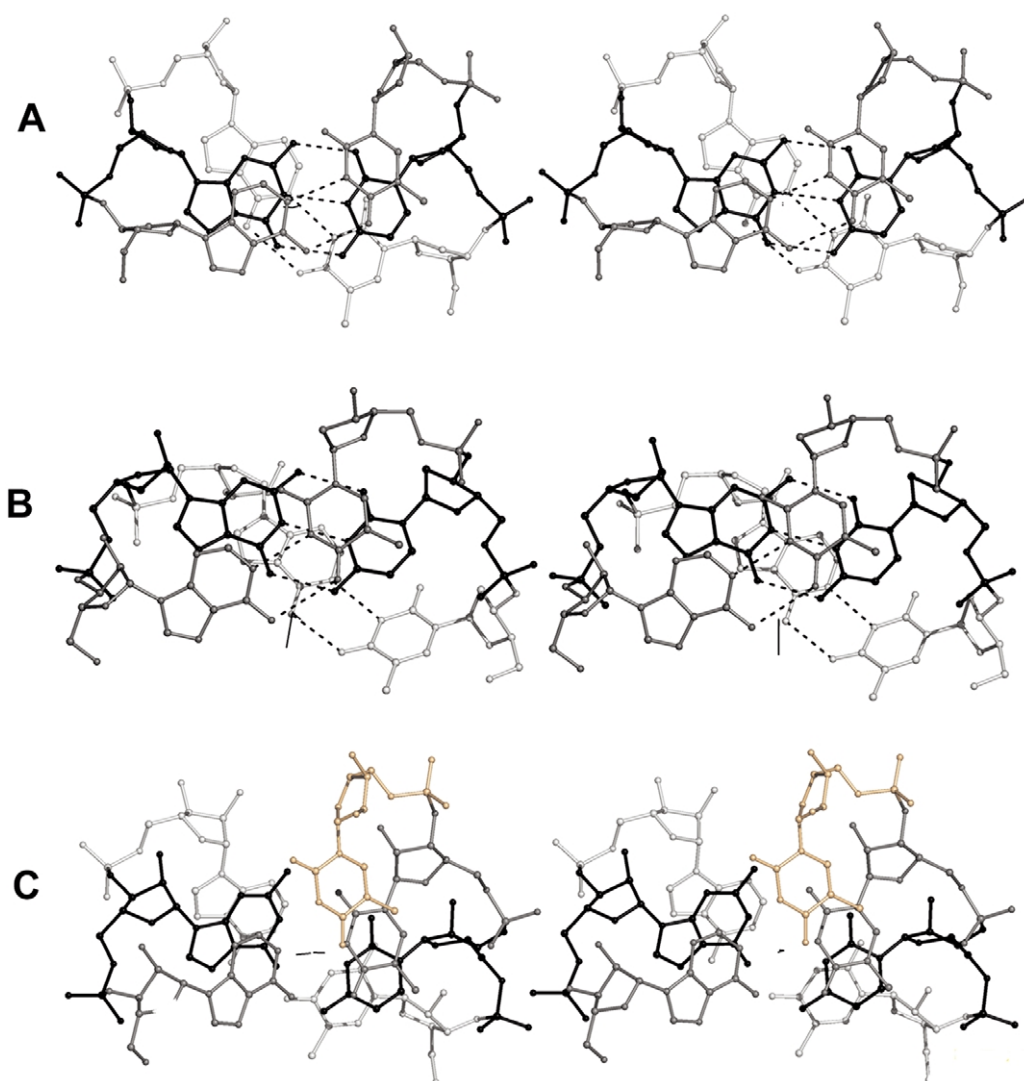
In DNA helices, the six-membered rings of purine bases are known to stack ring-over-ring. This is particularly true of runs of adenine bases. Introduction of a single guanine base (Figure 6A) into the stack does not alter this pattern, as shown in the crystal structure of the DNA/DNA analog of the PPT.<sup>27</sup> However, a discontinuity in the stacking

occurs in the RNA/DNA hybrid PPT: where a4 is not stacked on g5, but g5 does stack with its six-membered ring on the five-membered ring of a6 (Figure 6B). Although the stacking in the hybrid PPT bound to RT<sup>5</sup> is similar to that of the uncomplexed hybrid reported here (Figure 6C), the complex also has unpaired bases. Helix parameters, propeller, roll and buckle, are utilized to bring about this destacking. In Table 2, which compares these parameters at each base-step for the DNA and the hybrid helices, one can see that for the DNA analog all changes occur at the A-G step, which is at the center of the decamer. In contrast, the hybrid spreads out radical changes in parameters over three base-steps, a-a-g-a. All these parameter changes contribute to the deformability of this sequence.

In the RT-PPT structure Arnold and co-workers<sup>5</sup> reported an “unzipping” of a seven base-pair region, a-a-a-g-a-a, of the PPT. This unzipping or unpairing of base-pairs involves one RNA purine and one DNA pyrimidine two base-pair steps apart on opposite strands. Their RNA base, a3 (numbered as in our RNA/DNA helix), is unpaired and an unpaired DNA base, C16, is found on the pyrimidine strand (Figure 7A). This unzipping is not found in our hybrid (Figure 7B). The deformability of the purine a-g-a steps apparently contributes to the destacking and unpairing that occurs during RT-hybrid recognition.

### Hydration and cations

Along the bottom half of the molecule between bases 6 to 8, O2' hydroxyl groups on the sugars are linked together by a string of water molecules. The string is broken in the top half of the molecule because of the sugar switch on base 2 from C3' to C2'-endo. This places the O2' hydroxyl group closer to the major groove rather than the minor groove where it usually resides. Because the ends of one helix pack into the minor groove of another, the minor groove hydration pattern is broken at base-pair 1–20 as well as bases 4, 5 and 6 where packing contacts occur. The phosphate groups are highly hydrated, but again hydration patterns are broken by packing contacts. Water molecules on the phosphate groups have somewhat longer H-bonds (between 3.1 Å and 4.0 Å), whereas water molecules



**Figure 6.** Stacking of the A-G-A and a-g-a steps in (A) the DNA/DNA analog<sup>27</sup> and (B) the RNA/DNA hybrid. Grey = A4-T17 (nearest viewer), black = G5-C16, white = A6-T15 (farthest pair). In A the six-membered rings of purines stack atop one another, and an exocyclic oxygen of thymine or cytosine stacks over the following pyrimidine ring. In B the purine a4 is completely destacked, but its base-pair thymine stacks its exocyclic oxygen over the C16 ring below it. At the next base-pair step the six-membered ring of g5 stacks over the five-membered ring of a6 and C16 does not stack on T17 at all. C, The “unzipped” RT-bound hybrid helix,<sup>5</sup> where mispairing occurs. (Base numbering is that of our hybrid helix, Figure 1B.) a4 (grey) is no longer paired with its mate T17, but with the neighboring T18 (gold). g5 (black) is mispaired with T17 (grey). C16 (black) is unpaired and the A6–T15 base-pair (white) is back in proper register. In this structure the a4–T18 base-pair is destacked, g5 stacks on a6, and the remaining pyrimidines (T17, C16, T15) stack atop one another. Figure generated using PYMOL (<http://pymol.sourceforge.net/>).

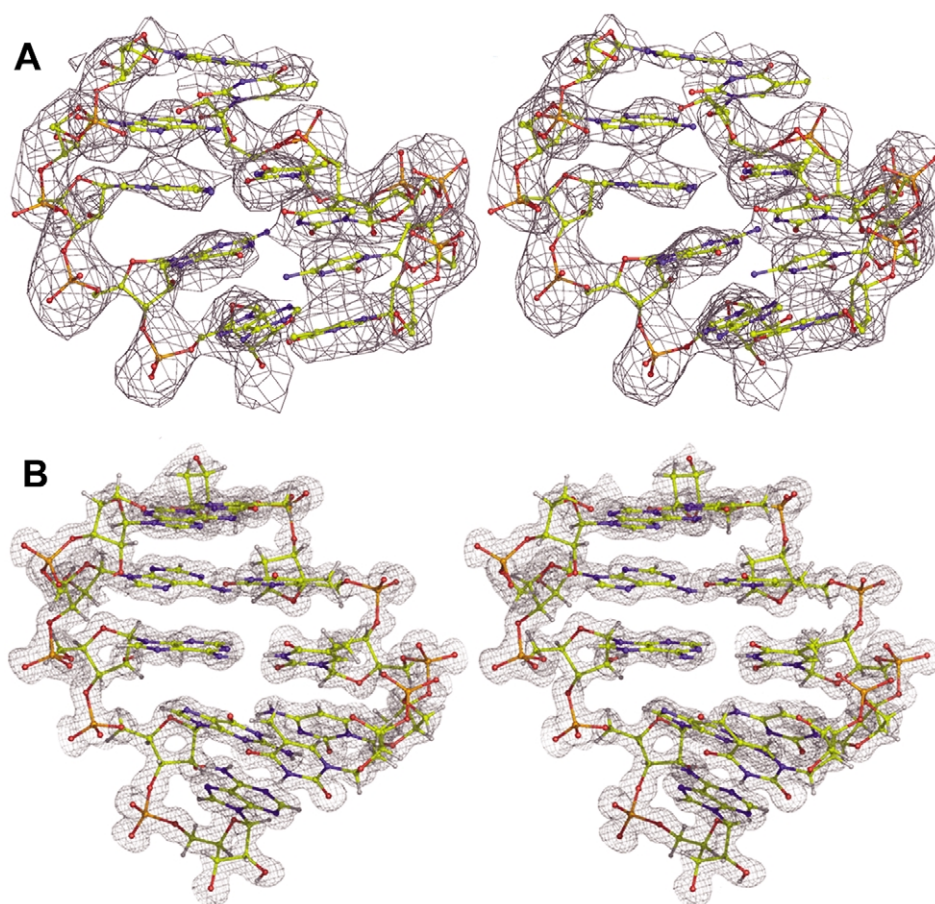
linking the O2' hydroxyl atoms are shorter (2.6–3.5 Å). The wide major groove, of course, is filled with water.

All three structure solutions, Ca-direct methods (Ca-DM), Ca-molecular replacement (Ca-MR) and Mg-molecular replacement (Mg-MR), contain four cations at the same position with the following exceptions: 2-methyl-2,4-pentanediol (MPD) was modeled for one of the calcium ions in the Ca-DM solution, while in the Ca-MR refinement a calcium ion having two alternate positions was modeled for this density. The Mg-MR located an additional two magnesium ions that are not at 100% occupancy. All cations are hydrated, and magnesium ions maintain octahedral geometry.

Hydrated calcium ions are known to have irregular coordination geometry with six, seven or even eight water molecules in the coordination sphere. In this structure, calcium coordination numbers of six or seven are found. In some cases, a phosphate oxygen atom replaces one of the coordination water molecules, anchoring the cation complex to the hybrid.

## Discussion

Both polymerases and nucleases perform their enzyme action on the sugar-phosphate backbone of the nucleic acid duplexes to which they bind.



**Figure 7.** A,  $2F_o - F_c$  map of the PPT sequence, r-(a-a-a-g-a)/d-(T-C-T-T-T) as found in the Sarafianos *et al.*<sup>5</sup> PPT/RT complex, contoured at  $0.9\sigma$ . B, The same hybrid sequence from our RNA/DNA structure, contoured at  $1.2\sigma$ . In A the first base-pair a2-T19 is paired, a3 is unpaired, a4 is mispaired with T18, g5 is mispaired with T17, C16 is unpaired and the helix comes back into register with the last base-pair a6-T15. In B all bases are properly paired. The difference in map-cage density occurs because A is a 3.0 Å structure and B is a 1.1 Å structure. Base numbering as in Figure 1. Figure generated using PYMOL (<http://pymol.sourceforge.net/>).

The importance of the 2'-OH as the single distinguishing feature of hybrid helices that confers selectivity was recognized by Horton & Finzel in the first crystal structure of an RNA/DNA hybrid.<sup>16</sup> Earlier X-ray work on chimeric duplexes, where a few RNA bases are interspersed in a DNA strand, observed that enzymes that recognize chimeric sequences must perceive the presence of the 2'-hydroxyl group.<sup>36</sup> Other features such as groove width and bending play a role, but these are sequence-dependent. The sugar switch from 3'-endo to a 2'-endo pucker on adenine 2 of the hybrid duplex reported here changes the pattern of sugar oxygen atoms on the floor of the minor groove. Other local backbone changes occur at base-pairs 1–20 and 2–19 where the A-form helix becomes more B-like, as judged on sugar pucker and  $Z_p$ , crankshaft transitions and BI *versus* BII phosphate groups. B-to-A transitions have been found recently in several nucleic acid structures, including an A-B intermediate,<sup>37</sup> methylated and brominated hexamers<sup>38,39</sup> and an RNA tetraplex.<sup>40</sup> Protein–nucleic acid complexes analyzed by Olson and co-workers<sup>35</sup> demonstrated that B-to-A transitions occur with enzymes that perform cut-

ting or sealing operations at the O3' phosphodiester bond. B-to-A transitions selectively expose for enzymatic attack sugar phosphate atoms, such as the 3'-oxygen atom, that ordinarily are buried. The reverse transition, A-to-B, conceals the 3'-phosphate oxygen atom from attack.

In the structure reported here, an unusual sugar switch on the RNA strand from C3'-endo to C2'-endo at the 5'-end of the PPT may play a role in rejection of the PPT by RNase H. Before considering evidence implicating this switch in recognition, it is important to see what evidence exists for C2'-endo sugar puckers on RNA strands. Zimmerman & Pheiffer<sup>41</sup> in their X-ray diffraction study on poly(rA)·poly(dT) state that poly(rA)·poly(dT) is unique in its ability to adopt either A or B-form helices. They note that by increasing the humidity of the fibers that initially gave an A-form diffraction pattern, "the helical parameters derived from wetted fibers of poly(rA)·poly(dT) are similar but not identical with those of wetted DNA fibers". This makes poly(rA)·poly(dT) unique among hybrids, in that it can adopt either A or B-form.

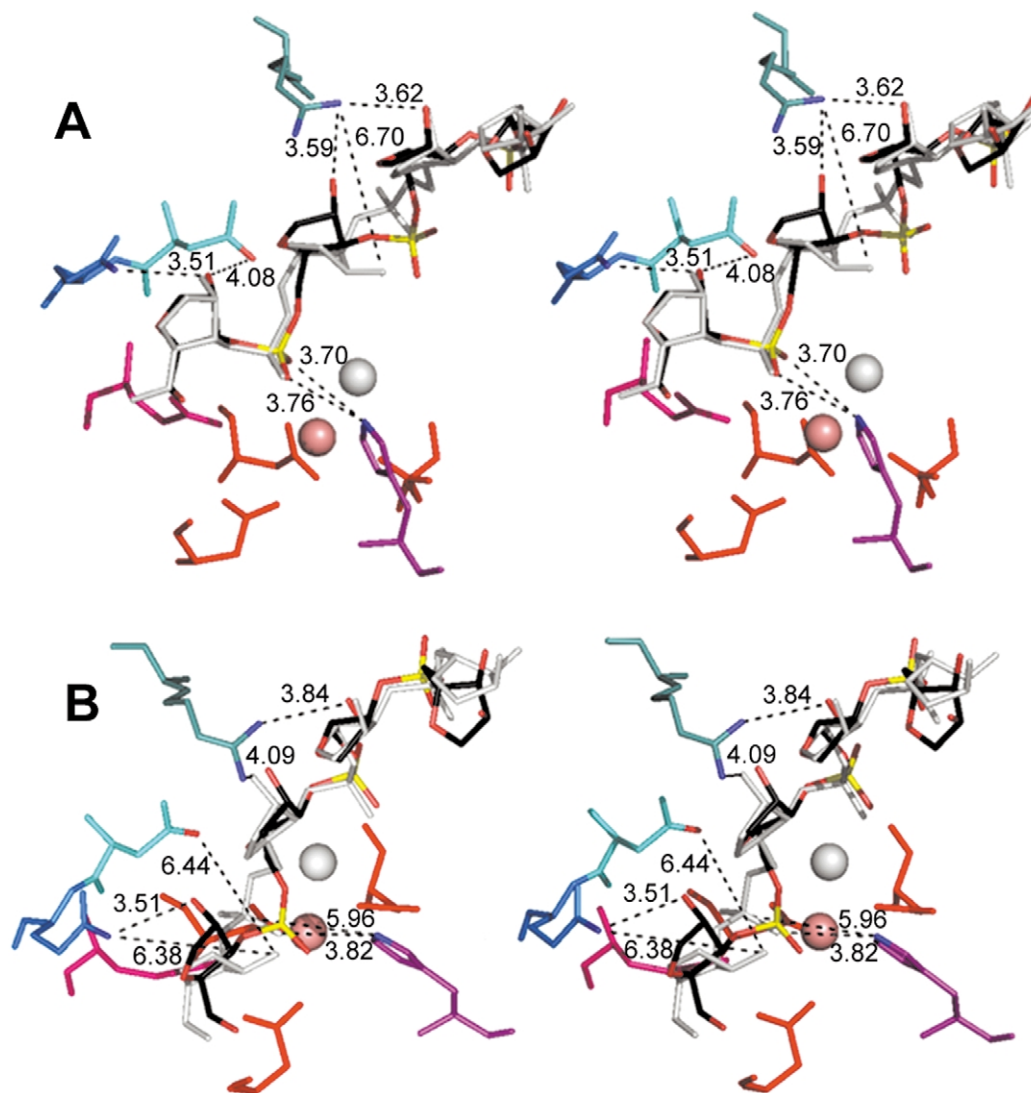
Molecular dynamic simulations<sup>34</sup> have shown

that “B-RNA” is energetically possible and can be converted to A-RNA by a change in sugar pucker from C2'-endo to C3'-endo. Sugar repuckering from C2' to C3'-endo and back occurs at a much lower rate than in DNA simulations. For this reason, the barrier to repuckering is adjudged to be higher in RNA than DNA, but is longer-lived when it occurs. As for the phosphate groups in the simulations, B I to B II transitions are observed in both DNA and RNA strands in “B-RNA” and hybrid helices. Although not conventional dogma, the authors conclude, “It was previously thought that the unacceptable stereochemistry of the O2' hydroxyl “bumping” into the following phosphate group, sugar ring and base would destabilize the B-form geometry and make B-RNA unfavorable... the interaction is not unfavorable.”

It appears from the preceding arguments that a

B-form sugar pucker (C2'-endo) in an RNA strand of a hybrid is acceptable geometry. This sugar occurs on the RNA strand at a c-a step, which is a pyrimidine-purine step (YR). YR steps, including T-A, C-A and U-A, are the most deformable,<sup>13,42</sup> and it has been shown in protein/DNA complexes that the T-A step can adopt either A or B conformation.<sup>35</sup>

The RNase H catalytic pocket is composed of positively charged amino acid residues, a conserved histidine residue and divalent cations, either manganese or magnesium, that are required for catalysis.<sup>43,44</sup> Catalysis occurs by deprotonation of water to form a nucleophilic hydroxide group that attacks the scissile phosphate group on RNA. To illustrate the role the adenine 2 sugar conformation plays in RNase H recognition at the 5'-end of the PPT, Figure 8 shows a least-squares fit of



**Figure 8.** A, Least square superposition of bases 1 to 4 of the RNA strand of the RNA/DNA decamer, c-a-a-a, (white) onto the Sarafianos *et al.*<sup>5</sup> backbone (black, red, yellow) at the RNase H active site (rms = 1.19 Å). Two magnesium ions, pink (crystallographic  $B = 23.5$ ) and white (crystallographic  $B = 79.9$ ) in the Huang *et al.*<sup>6</sup> RT structure are least square fitted onto the Sarafianos structure<sup>5</sup> (rms = 1.10 Å). B, Least square superposition of bases 2 to 5 (white), a-a-a-g, onto the backbone (rms = 1.25 Å). Positively-charged amino acid side-chains, R448, N474, and Q475, (from top to bottom left) are blue-green. Negatively charged E478, D443 and D549 (from left to right), with D498 below D443, are in red. H539 at lower right is magenta. Figure generated using PYMOL (<http://pymol.sourceforge.net/>).

the backbone atoms of the RNA/DNA hybrid (in white) to the Arnold and co-workers<sup>5</sup> RNA/DNA PPT structure at the RNase H active site. Three amino acid residues capable of recognizing the polyanionic backbone of the nucleic acid, R448, N474 and Q475, “read” the 2'-hydroxyl groups of the first four bases of the hybrid decamer RNA strand (white). N474 and Q475 have been suggested as critical in determining RNase H specificity.<sup>11</sup> These residues make hydrogen bonds from N and O atoms on their amino acid side-chains to 2'-oxygen atoms on the RNA sugars (Figure 8A). The second sugar, with the 2'-endo pucker, is not synchronous with the reading frame, having a distance of 6.70 Å from its 2'-hydroxyl group to an arginine NH<sub>2</sub>. At the bottom of Figure 8A are the acidic amino acid residues, D443, E478, D498 and D549, of the RNase H catalytic site along with H539, which is implicated in the cutting mechanism. RNase H cuts at a 3'-OH. Two magnesium ions, pink ( $B = 23 \text{ \AA}^3$ ) and white ( $B = 79 \text{ \AA}^3$ ), found in the structure by Harrison and co-workers<sup>6</sup> have been superimposed. In Figure 8B the 2'-endo sugar has moved into the RNase H catalytic pocket. Here, the reading frame has been disrupted and H-bonds to the 2'-hydroxyl group of the C2'-endo sugar are larger than 6 Å. The contacts to the O1 and O2 phosphate atoms are also too long (3.82 and 5.96 Å), and the scissile 3'OH group on the RNA strand (white) has been moved out of reach of the catalysis mechanism.

If sugar pucker and phosphate backbone conformation are important for recognition, the question remains, what keeps the sugar in this C2'-endo conformation? An interesting report, investigating the role of the “U-box” in replication, observes that neither the mechanism for PPT resistance to RNase H, nor the requirements for its 5'-end cut have been established.<sup>45</sup> A string of five uridine bases in the RNA strand (thymine in the final transcript) immediately upstream, i.e. 5', to the PPT is highly conserved in retroviruses. Mutations in this region can block viral replication in cells and the authors<sup>45</sup> pinpoint the blocking event as occurring between the first strand jump and the start of second strand synthesis. It is possible this

string of uridine bases plays a structural role. Stacking of five uridine bases upon four adenine bases would have quite different stacking geometry than that of random sequence. In relieving the strain at the juncture between the pyrimidine (Y) stack and the purine (R) stack, the backbone may adjust at this YR step with a C3'-endo to C2'-endo sugar switch.

The unzipping of the protein-bound hybrid helix found by Sarafianos *et al.*<sup>5</sup> occurs in the five base-pair region, a-a-a-g-a, of the PPT that includes the a-g-a steps. This five base motif is found on RNA strands in a number of retroviruses including CAEV, FIV, and SIV (see Table 2 of Iliyinski & Desrosiers<sup>45</sup>). Some retroviruses do not have the full five base motif, but have a shortened motif, e.g. FeLV has a-a-g-a. This sequence, a-a-a-g-a, appears to be important in retroviral recognition and may be used by other polymerases that recognize hybrids.

In the Nucleic Acid Database<sup>46</sup> three hybrid structures (AH0001, AH0005 and DR0003) are found that have an RNA strand with only purines and a DNA strand entirely of pyrimidines.<sup>26,47,48</sup> All of these sequences, r-(<sup>5'</sup>g-a-a-g-a-g-a-g-c<sup>3'</sup>), r-(<sup>5'</sup>g-a-a-g-a-a-g-a-g<sup>3'</sup>) and r-(<sup>5'</sup>g-a-a-g-a-a-g-a-a<sup>3'</sup>), contain deformable a-g-a steps. The Conn *et al.*<sup>26</sup> sequence crystallizes in the same space group,  $P2_12_12_1$ , as the hybrid reported here; the other two crystallize in space group  $P6_1$ . Each structure has two a-g-a steps (Table 3). At the 5'-end of the Conn *et al.*  $P2_12_12_1$  structure, adenine stacks over guanine, but not as in a DNA helix, where six-membered rings lie over each other. The six-membered ring of the first adenine stacks over the five-membered ring of guanine. The guanine six-membered ring then stacks only partially over the following five-membered adenine ring. At the a-g-a steps at the 3'-end, six-membered over five-membered stacking predominates. At the 5'-end of both Xiong & Sundaralingam crystal structures<sup>47,48</sup> the a-g step is de-stacked as in the hybrid reported here (Figure 6B), while at both 3'-ends, the a-g steps are stacked with the six-membered ring over the five. This leads to the conclusion that a-g-a steps in a hybrid helix are deformable, and this deformability appears to be context-dependent.

**Table 3.** The a-g-a steps in hybrid crystal structures

Author	NDB or PDB No.	Sequence (RNA strand)	Space group (Å)	Resolution	Stacking of a-g step
This work	UH0005 UH0006	c-a-a-a-g-a-a-a-g	$P2_12_12_1$	1.15	Unstacked
Han <sup>28</sup> Conn <i>et al.</i> <sup>26</sup>	AH0012 AH0001	g-a-a-g-a-g-a-a-g-c	$P2_12_12_1$	2.5	<sup>5'</sup> -a-g stacked <sup>3'</sup> -a-g partially stacked
Xiong & Sundaralingam <sup>48</sup>	DR0003	g-a-a-g-a-a-g-a-g	$P6_1$	1.8	<sup>5'</sup> -a-g unstacked <sup>3'</sup> -a-g stacked
Xiong & Sundaralingam <sup>47</sup>	AH0005	g-a-a-g-a-a-g-a-a	$P6_1$	1.8	<sup>5'</sup> -a-g unstacked <sup>3'</sup> -a-g stacked
Sarafianos <i>et al.</i> <sup>5</sup>	1HYS	a-a-a-a-g-a-a-a-a-g (seq. located within a 31-mer)	$P3_212$	3.0	Unstacked with mispaired bases before and after the unstacking

Only one protein-hybrid structure in the PDB database<sup>49</sup> contains the a-a-a-g-a sequence, the Sarafianos *et al.*<sup>5</sup> RT-PPT structure. In fact, only five structures report utilization of these five bases in their recognition sequence. Three of these are proteins bound to DNA targets, one is the hybrid RT-PPT structure, and the last is a ribosomal protein bound to RNA.

## Conclusions

Two sugar switches have been found in nucleic acid sequences utilized for both plus and minus strand initiation. Both occur at a YR step located one base-step away from the scissile bond. In the NMR structure of the Okazaki junction formed at minus-strand initiation, a sugar switch from C3'-endo to C2'-endo is found one step away from the scissile bond. This cut removes the tRNA<sup>Lys</sup> primer that is annealed to the viral RNA.<sup>14</sup> A narrow minor groove is found in the hybrid portion of this junction as well. Sarafianos *et al.*<sup>5</sup> also found a narrow minor groove in the PPT bound to RT. Neither the crystal structure of (–)-strand Okazaki fragment<sup>13</sup> nor the PPT decamer sequence reported here have a narrow minor groove. Possibly an induced fit between protein and the malleable nucleic acid with concomitant groove narrowing is occurring. The narrow minor groove, the sugar switch to C2'-endo on the RNA strand one base removed from the scissile bond, along with the unzipping of the bases found in the PPT bound to protein, facilitated by the deformable a-g-a steps, appear to be important signals for RNase H recognition.

In conclusion, the RT–nucleic acid structures to date have all been at or near 3.0 Å resolution. At this resolution, sugar conformation is indeterminate. To be able to release sugar constraints during refinement, data at or near 2.3 Å or less are needed. High-resolution structures of hybrids containing uridine tracts elucidating their stacking properties would also be helpful, as well as a structure of the 3' end of the PPT which contains the six guanine bases.

## Materials and Methods

### Crystallization and structure solution

The deoxy-pyrimidine strand was synthesized by solid-phase phosphoramidite methods and purified by anion-exchange chromatography on a Whatman DE-52 column. The ribonucleotide purine strand was purchased from Yale University's Keck Oligonucleotide Synthesis facility and purified by PAGE on 50 cm × 33 cm × 0.3 cm gels. The RNA was eluted from the gel using a Schleicher & Schuell Elutrap Electroeluter. Gel-purification of the RNA strand was essential for obtaining hybrid crystals that diffract to high resolution (1.10 Å). Hybrid crystals that resulted when the RNA strand was purified by chromatography diffracted only

between 2.5 Å and 4.5 Å. The two strands were annealed in a waterbath at 65 °C and allowed to cool slowly overnight to room temperature. Crystals were grown at room temperature by vapor-diffusion in micro-bridges (Hampton Research) using two different cations, magnesium (Mg-crystal) and calcium (Ca-crystal), and two pH values, 5.8 and 6.8. The Ca-crystal was grown from 0.3 mM hybrid duplex, 12 mM calcium acetate, 0.6 mM spermidine hydrochloride (pH 6.8), 0.075% β-octylglucoside, 12 mM sodium cacodylate (pH 6.8) and 12% (v/v) MPD versus 40% (Ca-crystal) and 50% (Mg-crystal) MPD in the crystallization reservoirs. All crystallization components were the same for the Mg-crystal with the exception of substituting 12 mM magnesium acetate for the calcium acetate and 12 mM sodium cacodylate at pH 5.8. The two crystals were isomorphous. Data were collected at the National Synchrotron Light Source at Brookhaven National Laboratories at –180 °C. The MPD used in crystallization was a sufficient cryoprotectant for both crystals. Data were processed using DENZO and SCALEPACK.<sup>50</sup> The Ca-crystal was solved independently by direct methods.<sup>28</sup> Comparison data statistics for all three structure solutions are shown in Table 1.

The Mg and Ca structures were solved by molecular replacement with EPMR<sup>51</sup> using a hybrid duplex model, NDB AH0001,<sup>26</sup> consisting of an RNA purine strand r-(g-a-a-g-a-g-a-a-g-c) and a DNA pyrimidine strand d-(G-C-T-T-C-T-C-T-T-C). (Statistics for the Ca and Mg data sets are reported together, separated by a slash between values, e.g. Ca/Mg.) Using data from 8–3 Å in EPMR for the Ca-crystal and 15–4 Å for the Mg-crystal, the starting  $R_{\text{xtal}}$  was (Ca/Mg) 54.8/55.8%. Rigid-body refinement in CNS<sup>52</sup> dropped (Ca/Mg)  $R_{\text{xtal}}$  to 45.9/50.3% and  $R_{\text{free}}$  to 44.2/49.4%. Simulated annealing plus B-factor refinement at 2.0 Å/1.5 Å (Ca/Mg) gave  $R_{\text{xtal}} = 31.3/35.0\%$ ,  $R_{\text{free}} = 38.8/38.2\%$ . SHELXL-97<sup>53</sup> refinement was begun (Ca/Mg) 8–2 Å/8–1.6 Å followed by anisotropic refinement when the majority of the data had been added. Final anisotropic R-factors were (Ca/Mg)  $R_{\text{xtal}} = 12.7/11.7\%$  and  $R_{\text{free}} = 15.8/15.6\%$ . The Ca-structure has four hydrated calcium ions, with one showing 50% occupancy at two positions, and 127 water molecules. The Mg-structure has six hydrated magnesium ions with two at lower occupancy (70% and 30%) and 128 water molecules. Because negative difference density was observed on two of the magnesium cations, it became apparent that this high-quality data required a reduced scattering factor for Mg<sup>2+</sup>. For this reason, Ne scattering factors, isoelectronic with Mg<sup>2+</sup>, were used. Water molecules were positioned where peaks greater than 1–1.25σ in the  $2F_o - F_c$  and at least 3σ in the  $F_o - F_c$  maps appeared simultaneously.

### CD spectroscopy

Microcrystals were dissolved in double-distilled water and the solution desalted in a Centricon 3 filter (three washes). The final solutions contained each strand at a concentration of 4.8 μM in a buffer of 150 mM NaCl, 5 mM MgCl<sub>2</sub>, 10 mM sodium cacodylate (pH 7.0). Spectra were recorded at 5 °C on a computer-driven AVIV 62DS CD spectrophotometer using a 1.0 nm band-width. The cell compartment was purged continuously with N<sub>2</sub>. The data were least-squares fit with a polynomial of seventh order. Extinction coefficients used were: r/d(C-A-A-A-G-A-A-A-A-G) ε = 115.3 cm<sup>-1</sup> mM<sup>-1</sup>, d(C-T-T-T-T-C-T-T-T-G) ε = 83.94 cm<sup>-1</sup> mM<sup>-1</sup>

### Coordinates accession number

Atomic coordinates and structure factors have been deposited in the Protein Data Bank (PDB)<sup>†</sup> and the Nucleic Acid Database (NDB) (PDB ID code 1PJG and 1PJO; NDB ID code UH0005 and UH0006).

### Acknowledgements

We thank Dr Duilio Cascio for data collection assistance, Dr Ann Maris for technical assistance and discussion, and Dr Michael Sawaya for assistance with graphics. This study was supported by National Institute of Health grant GM31299.

### References

- Kohlstaedt, L. A., Wang, J., Friedman, J. M., Rice, P. A. & Steitz, T. A. (1992). Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science*, **256**, 1783–1790.
- Jacobo-Molina, A., Ding, J., Nanni, R. G., Clark, A. D., Jr., Lu, X., Tantillo, C. *et al.* (1993). Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA. *Proc. Natl Acad. Sci. USA*, **90**, 6320–6324.
- Rodgers, D. W., Gamblin, S. J., Harris, B. A., Ray, S., Culp, J. S., Hellmig, B. *et al.* (1995). The structure of unliganded reverse transcriptase from the human immunodeficiency virus type 1. *Proc. Natl Acad. Sci. USA*, **92**, 1222–1226.
- Ding, J., Das, K., Hsiou, Y., Sarafianos, S. G., Clark, A. D., Jr, Jacobo-Molina, A., Tantillo, C. *et al.* (1998). Structure and functional implications of the polymerase active site region in a complex of HIV-1 RT with a double-stranded DNA template-primer and an antibody Fab fragment at 2.8 Å resolution. *J. Mol. Biol.* **284**, 1095–1111.
- Sarafianos, S., Das, K., Tantillo, C., Clark, A. D., Jr, Ding, J., Whitcomb, J. M. *et al.* (2001). Crystal structure of HIV-1 reverse transcriptase in complex with a polypurine tract RNA:DNA. *EMBO J.* **20**, 1449–1461.
- Huang, H., Chopra, R., Verdine, G. L. & Harrison, S. C. (1998). Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science*, **282**, 1669–1675.
- Le Grice, S. J. (1993). Human immunodeficiency virus reverse transcriptase. In *Reverse Transcriptase* (Skalka, A. M. & Goff, S. P., eds), pp. 175–191, Cold Spring Harbor Laboratory Press, Plainview, NY.
- Coffin, J. M., Hughes, S. H. & Varmus, H. E. (1997). Chapter 4. In *Retroviruses*, Cold Spring Harbor Laboratory Press, Plainview, NY pp. 129–130.
- Davies, J. F., II, Hostomska, Z., Hostomsky, Z., Jordan, S. R. & Matthews, D. A. (1991). Crystal structure of the ribonuclease H domain of HIV-1 reverse transcriptase. *Science*, **252**, 88–95.
- Tisdale, M., Schulze, T., Larder, B. A. & Moelling, K. (1991). Mutations within the RNase H domain of human immunodeficiency virus type 1 reverse transcriptase abolish virus infectivity. *J. Gen. Virol.* **72**, 59–66.
- Julias, J. G., McWilliams, M. J., Sarafianos, S. G., Arnold, E. & Hughes, S. H. (2002). Mutations in the RNaseH domain of HIV-1 reverse transcriptase affect the initiation of DNA synthesis and the specificity of RNase H cleavage *in vivo*. *Proc. Natl Acad. Sci. USA*, **99**, 9515–9520.
- Rausch, J. W., Lener, D., Miller, J. T., Julias, J. G., Hughes, S. H. & Le Grice, S. F. J. (2002). Altering the RNase H primer grip of human immunodeficiency virus reverse transcriptase modifies cleavage specificity. *Biochemistry*, **41**, 4856–4865.
- Mueller, U., Maier, G., Onori, A. M., Cellai, L., Heumann, H. & Heinemann, U. (1998). Crystal structure of an eight base pair duplex containing the 3'-DNA-RNA-5' junction formed during initiation of minus strand synthesis of HIV replication. *Biochemistry*, **37**, 12005–12011.
- Szyperski, T., Gotte, M., Billeter, M., Perola, E., Cellai, L., Heumann, H. & Wuthrich, K. (1999). NMR structure of the chimeric hybrid duplex r(gcaguggc)-r(gcca)d(CTGC) comprising the tRNA-DNA junction formed during initiation of HIV-1 reverse transcription. *J. Biomol. NMR*, **13**, 343–355.
- Fedoroff, O. Y., Salazar, M. & Reid, B. R. (1996). Structural variation among retroviral primer-DNA junctions: solution structure of the HIV-1 (–)-strand Okazaki fragment r(gcca)d(CTGC)-d(GCAGTGGC). *Biochemistry*, **35**, 11070–11080.
- Horton, N. C. & Finzel, B. C. (1996). The Structure of an RNA/DNA hybrid: a substrate of the ribonuclease activity of HIV-1 reverse transcriptase. *J. Mol. Biol.* **264**, 521–533.
- Powell, M. D., Ghosh, M., Jacques, P. S., Howard, K. J., Le Grice, S. F. J. & Levin, J. G. (1997). Alanine-scanning mutations in the “primer grip” of p66 HIV-1 reverse transcriptase result in selective loss of RNA priming activity. *J. Biol. Chem.* **272**, 13262–13269.
- Randolph, C. A. & Champoux, J. J. (1994). The use of DNA and RNA oligonucleotides in hybrid structures with longer polynucleotide chains to probe the structural requirements for Moloney murine leukemia virus plus strand priming. *J. Biol. Chem.* **269**, 19207–19215.
- Fuentes, G. M., Rodriguez-Rodriguez, L., Fay, P. J. & Bambara, R. A. (1995). Use of an oligoribonucleotide containing the polypurine tract sequence as a primer by HIV reverse transcriptase. *J. Biol. Chem.* **47**, 28169–28176.
- Powell, M. D. & Levin, J. G. (1996). Sequence and structural determinants required for priming of plus-strand DNA synthesis by the human immunodeficiency virus type 1 polypurine tract. *J. Virol.* **70**, 5288–5296.
- Fedoroff, O. Y., Salazar, M. & Reid, B. R. (1993). Structure of a DNA:RNA hybrid duplex. Why RNase H does not cleave pure RNA. *J. Mol. Biol.* **233**, 509–523.
- Schultz, S. J., Zhang, M., Kelleher, C. D. & Champoux, J. J. (1999). Polypurine tract primer generation and utilization by Moloney murine leukemia virus reverse transcriptase. *J. Biol. Chem.* **274**, 34547–34555.
- Fedoroff, O. Y., Ge, Y. & Reid, B. R. (1997). Solution structure of r(gaggacug):d(CAGTCCTC) hybrid: implications for the initiation of HIV-1 (+)-strand synthesis. *J. Mol. Biol.* **269**, 225–239.
- Gao, X., Jeffs, P. W. & Jeffs, P. W. (1994). Sequence-dependent conformational heterogeneity of a hybrid

<sup>†</sup> [www.rcsb.org](http://www.rcsb.org)

- DNA-RNA dodecamer duplex. *J. Biomol. NMR*, **4**, 367–384.
25. Yi, J. I., Lane, A., Conn, G. L. & Brown, T. (1998). Solution structures of DNA-RNA hybrids with purine-rich and pyrimidine-rich strands: comparison with the homologous DNA and RNA duplexes. *Biochemistry*, **37**, 73–80.
  26. Conn, G. L., Brown, T. & Leonard, G. A. (1999). The crystal structure of the RNA/DNA hybrid r(GAAGAGAAGC)-d(GCTTCTCTTC) shows significant differences to that found in solution. *Nucl. Acids Res.* **27**, 555–561.
  27. Han, G. W., Kopka, M. L., Cascio, D., Grzeskowiak, K. & Dickerson, R. E. (1997). Structure of a DNA analog of the primer for HIV-1 RT second strand synthesis. *J. Mol. Biol.* **269**, 811–826.
  28. Han, G. W. (2001). Direct-methods determination of an RNA/DNA hybrid decamer at 1.15 Å resolution. *Acta Crystallog. sect. D*, **57**, 213–218.
  29. Dickerson, R. E. & Chiu, T. K. (1998). Helix bending as a factor in protein/DNA recognition. *Biopolymers (Nucl. Acid Sci.)*, **44**, 361–403.
  30. Hung, S. H., Yu, Q., Gray, D. M. & Ratliff, R. L. (1994). Evidence from CD spectra that d(purine)-r(pyrimidine) and r(purine)-d(pyrimidine) hybrids are in different structural classes. *Nucl. Acids Res.* **22**, 4326–4334.
  31. Sen, A. & Graslund, A. (2000). Structural constraints regulating triple helix formation by A-tracts. *Biophys. Chem.* **88**, 69–80.
  32. Fratini, A. V., Kopka, M. L., Drew, H. R. & Dickerson, R. E. (1982). Reversible bending and helix geometry in a B-DNA dodecamer: CGCGAATTBRGCGC. *J. Biol. Chem.* **257**, 14686–14707.
  33. Dickerson, R. E., Kopka, M. L. & Drew, H. R. (1983). Structural correlations in B-DNA. In *Structure and Dynamics: Nucleic Acids and Proteins* (Clementi, E. & Sarma, R. H., eds), pp. 149–179, Adenine Press, Guilderland, NY.
  34. Cheatham, T. E., III & Kollman, P. A. (1997). Molecular dynamics simulations highlight the structural differences among DNA:DNA, RNA:RNA, and DNA:RNA hybrid duplexes. *J. Am. Chem. Soc.* **119**, 4805–4825.
  35. Lu, X.-J., Shakked, Z. & Olson, W. K. (2000). A-form conformational motifs in ligand-bound DNA structures. *J. Mol. Biol.* **300**, 819–840.
  36. Egli, M., Usman, N. & Rich, A. (1993). Conformational influence of the ribose 2'-hydroxyl group: crystal structures of DNA-RNA chimeric duplexes. *Biochemistry*, **32**, 3221–3237.
  37. Ng, H.-L., Kopka, M. L. & Dickerson, R. E. (1999). The structure of a stable intermediate in the A ↔ B DNA helix transition. *Proc. Natl Acad. Sci. USA*, **97**, 2035–2039.
  38. Vargason, J. M., Eichman, B. F. & Ho, P. S. (2000). The extended and eccentric E-DNA structure induced by cytosine methylation or bromination. *Nature Struct. Biol.* **7**, 758–761.
  39. Vargason, J. M., Henderson, K. & Ho, P. S. (2001). A crystallographic map of the transition from B-DNA to A-DNA. *Proc. Natl Acad. Sci. USA*, **98**, 7265–7270.
  40. Deng, H., Xiong, Y. & Sundaralingam, M. (2001). X-ray analysis of an RNA tetraplex (UGGGGU)<sub>4</sub> with divalent Sr<sup>2+</sup> ions at subatomic resolution (0.61 Å). *Proc. Natl Acad. Sci. USA*, **98**, 13665–13670.
  41. Zimmerman, S. B. & Pfeiffer, B. H. (1981). A RNA-DNA hybrid that can adopt two conformations: an X-ray diffraction study of poly(rA)-poly(dT) in concentrated solution or in fibers. *Proc. Natl Acad. Sci. USA*, **78**, 78–82.
  42. Dickerson, R. E. (1998). DNA bending: the prevalence of kinkiness and the virtues of normality. *Nucl. Acids Res.* **26**, 1906–1926.
  43. Goedken, E. R. & Marqusee, S. (2001). Co-crystals of *Escherichia coli* RNase H with Mn<sup>2+</sup> ions reveals two divalent metals bound in the active site. *J. Biol. Chem.* **10**, 7266–67211.
  44. Keck, J. L., Goedken, E. R. & Marqusee, S. (1998). Activation/attenuation model for RNaseH. *J. Biol. Chem.* **273**, 34128–34133.
  45. Iliyinski, P. O. & Desrosiers, R. C. (1998). Identification of a sequence element immediately upstream of the polypurine tract that is essential for replication of simian immunodeficiency virus. *EMBO J.* **17**, 3766–3774.
  46. Berman, H. M., Olson, W. K., Beveridge, D. L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S.-H., Srinivasan, A. R. & Schneider, B. (1992). The Nucleic Acid Database: a comprehensive relational database of three-dimensional structures of nucleic acids. *Biophys. J.* **63**, 751–759.
  47. Xiong, Y. & Sundaralingam, M. (1998). Crystal structure and conformation of a DNA-RNA hybrid duplex with a polypurine RNA strand: d(TTCTTBr<sup>3</sup>CTTCC)-r(GAAGAAGAA). *Structure*, **6**, 1493–1501.
  48. Xiong, Y. & Sundaralingam, M. (2000). Crystal structure of a DNA-RNA hybrid duplex with a polypurine RNA r(gaagaagag) and a complementary polypyrimidine DNA d(CTCTTCTTC). *Nucl. Acids Res.* **28**, 2171–2176.
  49. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H. *et al.* (2000). The Protein Data Bank. *Nucl. Acids Res.* **28**, 235–242.
  50. Otwinowski, Z. & Minor, W. (1996). Macromolecular crystallography. In *Methods Enzymol.* (Carter, J. W. & Sweet, R. M., eds), vol. 276, Academic Press, New York, pp. 307–326.
  51. Kissinger, C. R., Gelhaar, D. K. & Fogel, D. B. (1999). Rapid automated molecular replacement by evolutionary search. *Acta Crystallog. sect. D*, **55**, 484–491.
  52. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kuntzle, W. *et al.* (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallog. sect. D*, **54**, 905–921.
  53. Sheldrick, G. M. & Schneider, T. R. (1997). SHELXL: high resolution refinement. In *Methods Enzymol.* (Carter, J. W. & Sweet, R. M., eds), vol. 277, Academic Press, New York, pp. 319–343.

Edited by K. Morikawa

(Received 17 July 2003; received in revised form 17 September 2003; accepted 25 September 2003)