Structure of an RNA/DNA dodecamer corresponding to the HIV-1 polypurine tract at 1.6 Å resolution

Pawel Drozdzal,[a] Karolina Michalska,[a] Ryszard Kierzek,[b] Lechoslaw Lomozik[a,c]* and Mariusz Jaskolski[a,b]*

[a]Faculty of Chemistry, A. Mickiewicz University, 60-780 Poznan, Poland, [b]Institute of Bioorganic Chemistry, Polish Academy of Sciences, 61-704 Poznan, Poland, and [c]Faculty of Chemical Technology and Engineering, University of Technology and Life Sciences, 85-225 Bydgoszcz, Poland

Correspondence e-mail: lomozik@amu.edu.pl, mariuszj@amu.edu.pl

The crystal structure of an RNA/DNA hybrid dodecamer, r(5′-uaaaagaaaagg):d(5′-CCTTTTCTTTTA), which contains three-quarters of the polypurine tract (PPT) sequence of the HIV RNA genome is reported. The hybrid structure was determined at 1.6 Å resolution and was found to have the A-form conformation. However, the presence of alternate conformations along the RNA template strand indicated increased flexibility of the PPT sequence. Two segments (at nucleotides 1–2 and 6–8) of the RNA chain have two conformations exhibiting differences in torsion and pseudorotation angles. For conformation $I_{(1–2, 6–8)}$, 25% of the RNA sugars have the C2′-exo pucker and the rest have the expected C3′-endo pucker. The $II_{1–2}$ and $II_{6–8}$ conformations of the RNA strand have one sugar with the C2′-exo pucker. None of the ribose rings exist in the C2′-endo form, in contrast to a previous report which postulated a C2′-endo ribose as a key structural element of the PPT. The widths of the minor groove for conformations $I_{(1–2, 6–8)}$ and $II_{(1–2, 6–8)}$ of the RNA strand are 9.2–10.5 and 9.4–10.7 Å, respectively. Both ranges are very close to the intervals accepted for A-form RNA duplexes. On the opposing DNA primer strand most of the sugars are C3′-endo, except for the 3′-terminal sugars, which are C2′-endo (T22) or O4′-endo (T23 and A24). The duplex includes a noncanonical u1(anti)·A24(syn) base interaction with only one hydrogen bond between the bases. This noncanonical base interaction at the 5′-end of the template distorts the values of the helical parameters of the adjacent base pair.

1. Introduction

RNA/DNA hybrid duplexes are of great importance in medical and biological applications, *e.g.* in gene therapy and antisense technology. Hybrid duplexes also appear during replication, transcription and the synthesis of retroviral DNA by reverse transcriptase (Shaw & Arya, 2008). HIV reverse transcriptase (HIV-RT) has both DNA polymerase activity and RNase activity. The latter activity is responsible for the degradation of the RNA template and occurs only when it is presented in an RNA/DNA hybrid. While negative-strand DNA synthesis is primed using a host-derived tRNA annealed to the RNA template, positive-strand DNA synthesis is initiated from a purine-rich segment of the viral RNA known as the polypurine tract (PPT; 5′-aaaagaaaaggggggg-3′ in HIV). In a remarkable and still poorly understood process, the PPT of HIV is not digested by the RNase H domain of HIV-RT. Mutation studies indicate that changes in the PPT sequence

have no apparent effect on RNase H activity towards the hybrid duplex (Rattray & Champoux, 1989). *In vivo*, the intact PPT sequence is essential as it serves as the primer for the positive-strand DNA synthesis and helps to define the precise end of the viral DNA, which is critical for the next enzymatic step, which is controlled by the retroviral integrase.

Elucidation of the multi-step process of reverse transcription requires knowledge of the structure of the nucleic acid duplexes utilized by HIV-RT. The crystal structure of full-length HIV-RT in complex with RNA/DNA substrate containing the PPT has shown that contacts between the RNA/DNA hybrid molecule and amino-acid residues in the RNase H primer grip and in its catalytic centre primarily occur with the sugar–phosphate backbone (Sarafianos *et al.*, 2001). Various fragments of the duplex sequence that are important for proper recognition and excision by the enzyme have been studied structurally by NMR and crystallography. Generally, hybrid helices adopt A-like geometry in the crystal structures, with both RNA and DNA strands having C3′-endo sugar conformations. Nevertheless, exceptions in sugar conformation at individual base steps have been detected using crystallography, for instance by Kopka *et al.* (2003). According to that report, which describes the structure of a PPT fragment at 1.1 Å resolution, a sugar switch from C3′-endo to C2′-endo at an RNA adenine (a3 in the present numbering) could cause the enzyme to pause at the 5′-end of the PPT as the result of an altered pattern of hydrogen bonds between the enzyme and the sugar-phosphate backbone. In contrast, NMR studies of RNA/DNA hybrids reveal that the DNA strands show B-like sugar conformations (*e.g.* C2′-endo, C1′-exo and O4′-endo), while the RNA sugars are C3′-endo (Fedoroff *et al.*, 1993). Yi-Brunozzi *et al.* (2008) employed high-resolution NMR spectroscopy to examine the full-length PPT of HIV and found normal Watson–Crick base pairing and A-type conformation throughout the entire PPT sequence, except at the a3 and a7 positions, which had a mixed C3′/C2′-endo conformation. These observations were also confirmed when the PPT hybrids contained fluorinated or changed bases.

In RNA/DNA duplexes the DNA sugars tend to shift towards a C3′-endo-like pucker, with average phase angles of pseudorotation in the range of O4′-endo puckers (Cross *et al.*, 1997). The structures of different RNA/DNA hybrids can

```
        1              12
5′r uaaaagaaaagg
3′d ATTTTCTTTTCC
    24              13
            (a)

        2              11
5′r caaagaaaag
3′d GTTTCTTTTC
    23              14
            (b)
```
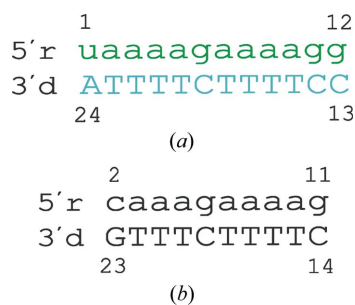
Figure 1
(a) The nucleotide sequence of the RNA/DNA dodecamer analyzed in this report. The RNA sequence is shown in green in lower case and the DNA sequence is shown in cyan in upper case. (b) Nucleotide sequence of the duplex analyzed by Kopka *et al.* (2003) with base numbering adopted from (a).

Table 1
Data-collection and refinement statistics.

Values in parentheses are for the last resolution shell.

Data collection	
Radiation source	X13, DESY, Hamburg
Wavelength (Å)	0.8010
Temperature (K)	100
Space group	$P3_221$
Unit-cell parameters (Å)	$a = 41.9$, $c = 57.2$
Resolution range (Å)	50.0–1.60 (1.66–1.60)
No. of reflections	8012
Completeness (%)	99.5 (99.6)
Multiplicity	6.7 (5.5)
$\langle I/\sigma(I)\rangle$	41.7 (2.85)
R_{int}† (%)	4.0 (55.0)
Refinement	
Refinement program	*REFMAC*5
Resolution (Å)	15.32–1.60
No. of reflections in working set	7204
No. of reflections in test set	790
R/R_{free}‡ (%)	18.1/22.9
No. of atoms	
Nucleic acids	497
Solvent	76
Mg^{2+}	3
$\langle B\rangle$ (Å^2)	
Nucleic acid chain *A*	28.7
Nucleic acid chain *B*	27.0
Solvent	37.1
Mg^{2+}	23.3
R.m.s. deviations from ideal	
Bond lengths (Å)	0.020
Bond angles (°)	2.26

† $R_{int} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of observation i of reflection hkl. ‡ $R = \sum_{hkl} ||F_{obs}| - |F_{calc}|| / \sum_{hkl} |F_{obs}|$, where F_{obs} and F_{calc} are observed and calculated structure factors. R_{free} is calculated using reflections excluded from refinement.

differ depending on whether the DNA strand consists of purine or pyrimidine nucleotides (Xiong & Sundaralingam, 1998).

It has been suggested that the specificity of RNase H for the RNA/DNA substrate depends on the width of the minor groove, the flexibility of the RNA/DNA encompassing the PPT and the sugar pucker of the substrate nucleotides, but it is not clear to what extent the enzyme can tolerate conformational variability of the RNA/DNA substrate (Fedoroff *et al.*, 1993; Tonelli *et al.*, 2003). It has been observed that hybrid helices with different (*e.g.* non-A) minor-groove widths are cleaved by HIV RNase H much less efficiently (Sarafianos *et al.*, 2009). In solution, most hybrids have a minor-groove width intermediate between standard A-form and B-form double helices. On the other hand, the crystal structures tend to have A-like minor-groove parameters. The unzipping of the protein-bound hybrid double helix found in the crystal structure of HIV-RT complexed with the RNA/DNA substrate occurs at the agaaa base-pair steps (Sarafianos *et al.*, 2001). This sequence motif (or a shorter motif; *e.g.* agaa in FeLV) is found in the RNA strands of many retroviruses (*e.g.* CAEV, FIV and SIV; Kopka *et al.*, 2003).

In the present study, we describe the crystal structure of an RNA/DNA dodecamer duplex containing the 11 5′ bases of the PPT motif of HIV-1 preceded by a u residue from the 'U-box' (Ilyinskii & Desrosiers, 1998; Fig. 1) and discuss its

Table 2
Average base-pair and local base-pair step helical parameters.

The structural features of DNA and RNA helices are taken from Lu & Olson (2003) and Arnott (1999).

	Helical twist Ω_h (°)	Helical rise h (Å)	Inclination η (°)	Shift (Å)	Slide (Å)	Rise (Å)	Tilt (°)	Roll (°)	Twist (°)
A-DNA	32.70	2.55	19.80†	0.00	−1.40	3.30	0.0	12.4	30.3
A-RNA	32.70	2.81	15.50	−0.08	−1.48	3.30	−0.4	8.6	31.6
A′-RNA	30.00	3.00	10.60	0.05	−1.88	3.39	−0.1	5.4	29.5
This work (except u1·A24 base pair)	33.73	2.82	15.70	0.11	−1.14	3.27	0.19	8.83	32.22
PDB entry 1pjo‡	32.64	2.89	13.97	0.00	−1.26	3.23	−1.19	7.91	31.23

† In duplexes of the A-type, η is usually between 10 and 20° (Kennard & Hunter, 1991) but can be as low as 7° (Heinemann *et al.*, 1987). ‡ Kopka *et al.* (2003).

geometry in comparison with previously reported models of PPT.

2. Materials and methods

2.1. Oligonucleotide synthesis, purification and crystallization

DNA and RNA dodecamers were synthesized on an Applied Biosystems DNA/RNA synthesizer using phosphoramidite chemistry. Oligonucleotides were cleaved from the solid support using 3:1(*v*:*v*) ammonia:ethanol and were incubated overnight in the same solution at 328 K for deprotection. Purification was performed by thin-layer chromatography on silica-gel plates (0.5 mm, Merck) in 55:35:10(*v*:*v*:*v*) 1-propanol:ammonia:water (Xia *et al.*, 1998). A 1 m*M* solution of the hybrid duplex in water was annealed at 338 K for 10 min. Single crystals of the RNA/DNA hybrid were grown at 292 K by the hanging-drop vapour-diffusion method by mixing 2 µl nucleic acid solution and 2 µl precipitating solution consisting of 7%(*v*/*v*) 2-methyl-2,4-pentanediol (MPD), 40 m*M* sodium cacodylate pH 5.5, 40 m*M* LiCl, 40 m*M* Co[(NH$_3$)$_6$]Cl$_3$ and 20 m*M* MgCl$_2$. The drops were equilibrated against 0.5 ml 40%(*v*/*v*) MPD. Crystals appeared within two weeks and grew to dimensions of 0.3 × 0.1 × 0.1 mm.

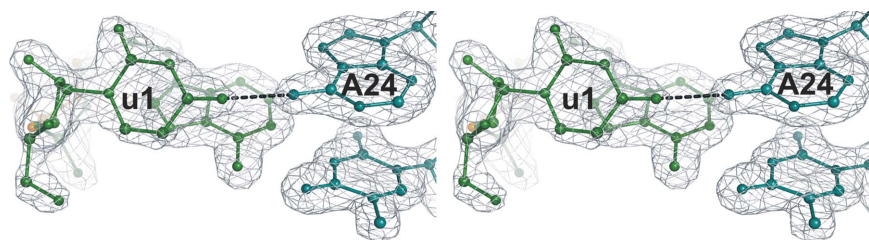### 2.2. Data collection, structure solution and refinement

X-ray diffraction data were collected to 1.6 Å resolution on the EMBL X13 beamline at the DESY synchrotron in Hamburg. The mother liquor served as the cryoprotectant solution. The data were indexed, integrated and scaled using the *HKL*-2000 program suite (Otwinowski & Minor, 1997). The X-ray data statistics are summarized in Table 1. The

structure was solved by molecular replacement using *Phaser* (McCoy *et al.*, 2007) with PDB entry 1pjo (Kopka *et al.*, 2003) as a molecular probe. The missing base pairs in the model were added manually in the program *Coot* (Emsley *et al.*, 2010). The model was refined using *REFMAC*5 (Murshudov *et al.*, 2011) from the *CCP*4 program suite (Winn *et al.*, 2011). 10% of all reflections were selected at random and set aside for R_{free} calculations. The program *Coot* was used for visualization of electron-density maps and for manual rebuilding of the atomic model. The model was refined using isotropic *B* factors and TLS parameters (Painter & Merritt, 2006; six TLS groups, with three groups per strand of nucleic acid) and validated using the free *R* test (Brünger, 1992) and with the program *NuCheck* (Feng *et al.*, 1998). In the final stages of model building, two segments (u1–a2 and g6–a8) of the RNA molecule were modelled in two alternate conformations, designated I (major, 60%) and II (minor, 40%). The helical parameters were calculated using 3*DNA* (Lu & Olson, 2003) and *Curves+* (Lavery *et al.*, 2009) and the figures were generated in *PyMOL* (DeLano, 2002).

3. Results

3.1. Helix parameters: overall structure

The sequence of the duplex analyzed in this work consists of a uridine residue from the 'U-box' and the subsequent 11 5′ bases of the PPT of HIV. In the present hybrid duplex most of the conformational helix parameters are typical of the A-form double helix (Table 2). The u1·A24 bases interact *via* only one hydrogen bond (2.57 Å) between the N6 atom of the adenine and the carbonyl O4 atom of the uridine (Fig. 2) and this noncanonical base interaction distorts the average values of several helical parameters. The most pronounced changes are observed for helical twist, twist and inclination. With the exception of the u1·A24 bases, there are 10.67 residues per helical turn, with an average helical twist of 33.73°, which is consistent with A-DNA or A-RNA. The average inclination angle (15.70°) falls in the middle of the range observed for A-form duplexes. Considering the entire hybrid duplex, one obtains 12.65 base pairs per turn, which resembles the pattern of
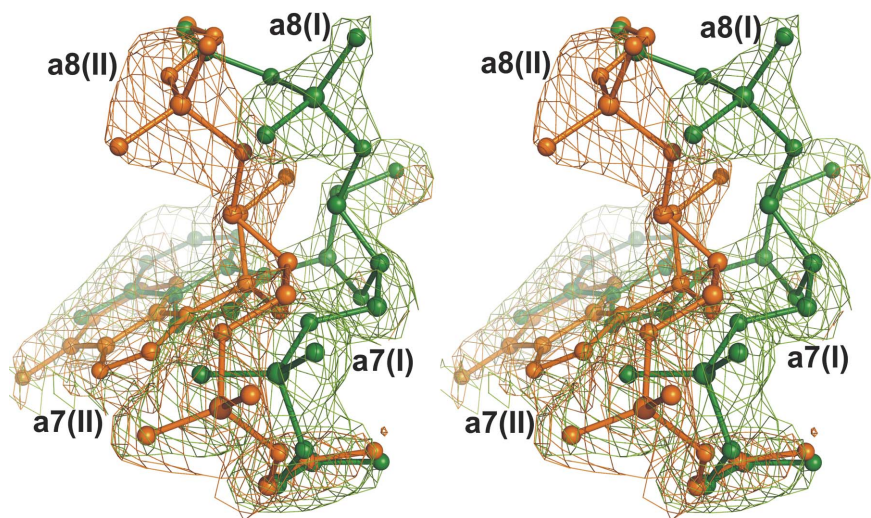

Figure 2
Stereoview of the u1·A24 bases with one hydrogen bond (2.57 Å, dashed line). The A24 nucleoside has the glycosidic angle in the *syn* conformation (χ = 64.7°). The $2F_o − F_c$ map is contoured at the 1σ level.

Figure 3
Stereoview of a portion of the hybrid duplex between RNA bases g6–a8 with two alternative backbone conformations (I_{6-8}, green; II_{6-8}, orange) shown in $F_o - F_c$ OMIT maps contoured at 2.5σ. The green map was calculated with the omission of conformation I_{6-8} and the orange map was calculated with the omission of conformation II_{6-8}.
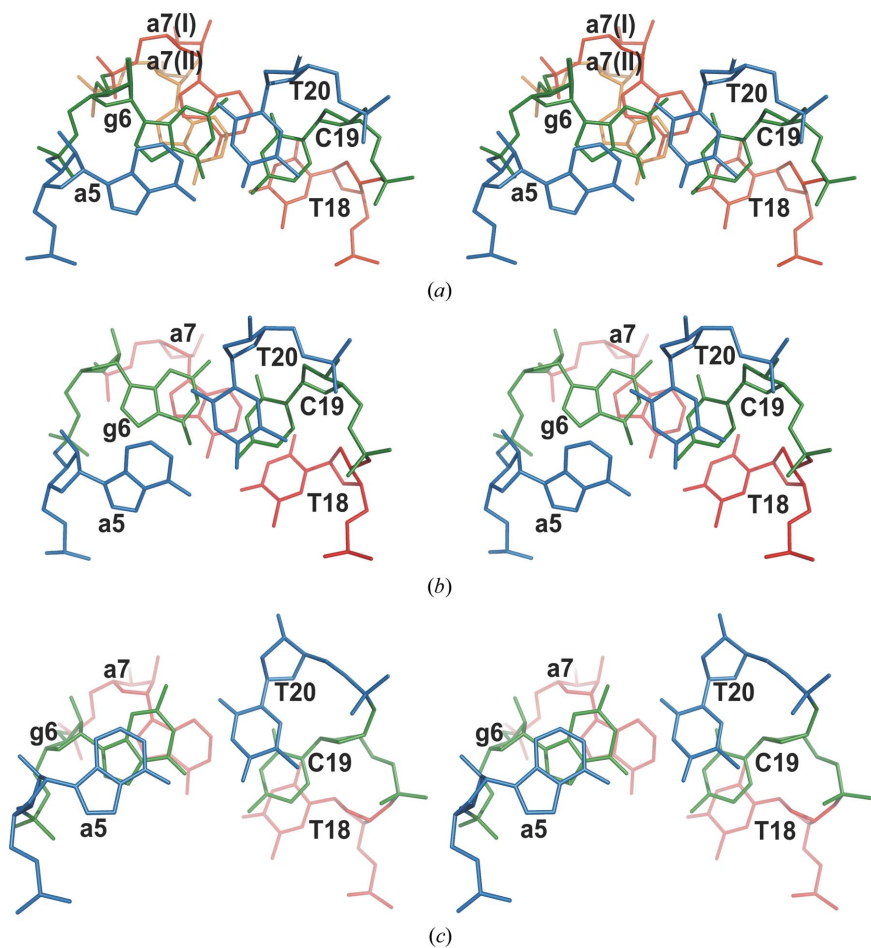

Figure 4
Stereoview of the stacking of the aga/TCT steps in (*a*) the present RNA/DNA duplex, (*b*) the RNA/DNA hybrid described by Kopka *et al.* (2003) and (*c*) the 'unzipped' RT-bound hybrid duplex (Sarafianos *et al.*, 2001). Blue, a5·T20; green, g6·C19; red/orange, a7·T18 in conformation I/II.

A′-RNA, while the average inclination (7.48°) would be outside the range attributed to A-type helices. The helical and backbone parameters of most of the nucleotides are classified as close to the canonical AI form (Svozil *et al.*, 2008; Schneider *et al.*, 2004). The most notable exceptions are nucleotides a2 and T23, which have an intermediate A/B form.

The DNA strand exists in a single well defined conformation with most of the sugars in the C3′-*endo* form, except for the 3′-terminal sugars, which are C2′-*endo* for T22 ($P = 154.4°$) or O4′-*endo* for T23 ($P = 84.6°$) and A24 ($P = 72.8°$).

3.2. Disorder in backbone and sugar conformation of the RNA strand

The sugar-phosphate backbone of the hybrid duplex exhibits local conformational flexibility of the RNA strand. Electron-density maps reveal that two segments of the RNA strand exist in two conformations [major (I) and minor (II)]: residues u1–a2 (conformations I_{1-2} and II_{1-2}) and residues g6–a8 (conformations I_{6-8} and II_{6-8}) (Fig. 3). The quality of the electron-density maps in the disordered areas is generally good, except for the sugar residue of the ribonucleotide a7 in conformation II, which has broken $2F_o - F_c$ electron density at the 1σ level. However, the backbone in this area can be traced with confidence in OMIT maps, especially as the adjacent phosphate groups, which serve as points of attachment, have very clear electron density. Also, the removal of conformation II_{6-8} from the model increased R and R_{free} by 1.3% and 1.4%, respectively. The disordered backbone fragments are anchored to common base moieties, with the exception of the a7 base, which has been modelled in two positions (I and II). These two conformations along the RNA strand exhibit differences in torsion and pseudorotation (P) angles.

Terminal base pairs (especially when they are weaker than cytosine·guanine) may be disordered owing to their lowered stability and owing to a lack of stacking interactions, and in consequence may be less well defined in the electron-density maps (Xiong & Sundaralingam, 1998). In the present structure, the u1·A24 bases at the 5′-end of the RNA template form only one, noncanonical hydrogen-bond interaction (see above) and the A24 nucleotide has an unusual *syn*

Table 3
Helical parameters (°) corresponding to stacking interactions in the present structure (upper values) and at the corresponding residues of the PDB model 1pjo (Kopka *et al.*, 2003) (bottom values).

Values in parentheses refer to conformation II of the RNA chain in the present structure.

A-form DNA†	B-form DNA†	Base	u1	a2	a3	a4	a5	g6	a7	a8	a9	a10	g11	g12	Average without u1·A24	Total bend
−0.1	0.5	**Buckle**	−6.15	−0.67	−2.95	0.72	−6.61	−6.23	−12.27 (−1.93)	−10.27	−6.41	−4.15	−2.16	0.78	−4.57	
			—	−27.19	1.40	6.45	2.83	−18.82	−16.76	−3.99	−0.34	−7.42	5.84	—	−5.80	
−11.8	−11.4	**Propeller**	−42.85	−18.86	−12.65	−24.63	−19.29	−11.65	−21.97 (−20.25)	−16.38	−16.60	−15.36	−9.13	−5.65	−15.65	
			—	−10.66	−6.55	−20.80	−22.89	−2.12	−5.07	−6.62	−14.35	−4.17	−7.60	—	−10.08	
8.0	0.6	**Roll**	23.37	1.93	12.07	9.42	8.19	8.23	18.52 (18.81)	9.68	6.92	7.90	5.46	—	8.83	
			—	8.89	2.27	6.33	20.01	10.32	7.51	6.17	6.22	3.43	—	—	7.91	
		Helical axis bend	—	4.5 (4.3)	3.3	2.4	2.2	2.4 (2.6)	1.4 (1.4)	2.0 (1.9)	1.6	1.6	1.3	1.0	—	10.6 (10.9)
			—	—	3.2	3.1	3.1	3.4	4.0	3.4	2.9	2.2	1.6	—	—	24.6

† Values for high-resolution A-DNA and B-DNA structures are‘ taken from the survey by Olson *et al.* (2001).

conformation. Analysis of the u1·A24 geometry suggests that it is facilitated by the low energy barrier (1.3 kJ mol^{-1}; Neidle, 2008) of the A24(*anti*)→A24(*syn*) transition, a more favourable stacking interaction with A23 and the possibility of creating an additional hydrogen bond with the a2 base (N6 of A24 is involved in a bifurcated hydrogen bond with N1 of a2 at a distance of 3.35 Å). Large differences in the backbone angles ε and ζ are observed for u1 in the two conformations, I$_{1-2}$ and II$_{1-2}$, of the RNA strand. The angles are $\varepsilon = 223.1$, $\zeta = 297.7°$ and $\varepsilon = 185.0$, $\zeta = 346.8°$ for conformations I$_{1-2}$ and II$_{1-2}$, respectively. There are also differences between the α, β and γ angles of the two conformations: the angles are 293.6, 189.1 and 43.3°, respectively, for conformation I$_{1-2}$ and 250.1, 197.0 and 45.6° for conformation II$_{1-2}$. It is evident that these angles are influenced by the value of ζ of the preceding nucleotide (u1), especially in the case of the second conformer.

The other region of increased local flexibility of the RNA is comprised of residues g6–a8. The sugar rings of adenosines 7 and 8 have a C2′-*exo* pucker in conformation I$_{6-8}$ ($P = 359.9°$ and 355.6°, respectively) and a C3′-*endo* pucker in conformation II$_{6-8}$ ($P = 9.7°$ and 7.6°, respectively). In the g6–a8 region, the backbone disorder in conformation I$_{6-8}$ is associated with a correlated change of the α and γ torsion angles of a8 from the usual *gauche*$^-$/*gauche*$^+$ conformation to the *trans*/*trans* region (α/γ flipping; Conn *et al.*, 1999). The increased deformability at a–g–a steps in RNA/DNA hybrid duplexes has been noted before, for instance by Kopka *et al.* (2003). However, these authors reported destacking of a5 at the ag step without any disorder of the backbone conformation. The present situation is quite different, since the disorder is manifested mainly in the backbone conformation, with only limited effect on base stacking.

Most of the Zp values, which describe the displacement of the P atom from the *xy* plane of the 'middle frame' between neighbouring base pairs (1.92–2.67 Å for both RNA conformations), also indicate the A-form of the hybrid helix. For conformation I$_{(1-2,\ 6-8)}$ of the RNA strand, the dinucleotide steps 1–3 (u1a2/T23A24–a3a4/T22T21) and 7 (a7a8/T18T17) have Zp values of 1.02, 0.76, 1.35 and 1.49 Å, respectively,

which all fall in the gap between the ranges characteristic of pure A-DNA (Zp > 1.5 Å) and B-DNA (Zp < 0.5 Å) (Lu & Olson, 2003). For the RNA conformations labelled II$_{(1-2,\ 6-8)}$ the Zp values for dinucleotide steps 1–3 also fall in the gap between the ranges of A-DNA and B-DNA and are 0.81, 0.76 and 1.36 Å, respectively. However, for dinucleotide step 7 the Zp value of 2.02 Å indicates the A-form.

The trajectory of the helical axis of the RNA/DNA hybrid complexed with HIV-RT shows a strong kink in the unzipped area, with a total axis bend of ∼40°. In contrast, in the present structure the double helix is nearly straight (maximum kink of ∼4.4° at the disordered segment a2·T23), with a total bend of ∼10.7°, which is even less than that in the RNA/DNA hybrid described by Kopka *et al.* (2003) (24.6° total; maximum of 4.0° at a7·T18), which has no backbone disorder in the middle of the duplex.

The width of the minor groove across the phosphate groups after subtraction of 5.8 Å for the radii of the two phosphate groups exhibits slight differences for conformations I$_{(1-2,\ 6-8)}$ (9.2–10.5 Å) and II$_{(1-2,\ 6-8)}$ (9.4–10.7 Å) of the RNA strand. The width of the major groove is 5.0–6.4 Å for conformation I$_{(1-2,\ 6-8)}$ and 5.5–6.2 Å for conformation II$_{(1-2,\ 6-8)}$.

3.3. Stacking interactions

The consecutive Watson–Crick base pairs show the typical extensive overlap, ranging from 2.87 to 4.49 Å^2 for the RNA chain and from 2.63 to 5.29 Å^2 for the DNA strand. The degree of overlap is only reduced in the agaa/TTCT fragment, which includes adenine 7 in two conformations. The major conformation (I) forms a proper Watson–Crick pair with T18. Owing to the altered backbone trace, however, adenine 7 in the minor conformation (II) is shifted in the base-pair plane, leading to increased ga/TC overlap (6.58 Å^2) and compromising the base-pair interactions, which are limited to one hydrogen bond [(a7)N1···O4(T18), 2.72 Å]. The reduction of overlap is seen at RNA step 5 (ag/CT, 1.86 Å^2), step 6 (ga/TC for a7 form I, 1.75 Å^2) and step 7 (aa/TT for a7 form II, 0.57 Å^2) (Fig. 4). Stacking interactions are also reduced to ∼1.5 Å^2 for the DNA bases at step 7 (aa/TT). The u1

nucleobase does not form any stacking with a2 on the 3′-side. The propeller, roll and buckle parameters can be used to characterize the degree of destacking, as illustrated in Table 3.

3.4. Hydration and metal cations

A total of 76 water molecules have been found to hydrate the hybrid duplex in the asymmetric unit. All water molecules are fully occupied, except for four waters that form close contacts with symmetry-related or double-conformation fragments. A greater number of water molecules are involved in interactions with the RNA chain than with the DNA chain. 25 water molecules hydrate the backbone phosphates, 17 hydrate the sugar atoms O2′, O3′, O4′ and O5′, and 27 hydrate the base moieties, while 11 are found in the coordination spheres of magnesium ions. The purine bases have betweeen one and three hydrogen-bonded water molecules, of which one can form a bridge with the phosphates of a4, a10 and g12. The pyrimidine bases also have between one and three associated water molecules, which can form bridges between two consecutive pyrimidine rings of the DNA strand. The O2′ atoms can be water-bridged to O3′ or to O4′ of 3′-adjacent residues. Water molecules bridging adjacent phosphate groups in the same strand are also observed. These hydration patterns are typical for double-helical nucleic acids in the A-form (Fig. 5; Schneider *et al.*, 1992; Egli *et al.*, 1996). Detailed information about the hydration of this hybrid duplex can be found in Supplementary Table S1[1].

Hydrated magnesium ions are known to have a preference for guanine binding at gg or gu steps in the major groove of RNA pseudoknots and also in the major groove at GN steps of A-DNA and B-DNA (Egli, 2002). In the present structure, three hydrated magnesium ions with two, four or five co-ordinated water molecules have been found. Mg1 binds at the g11–g12 step in the major groove and is engaged in water-mediated contacts to O6 and N7 of the guanine rings (Fig. 6). Mg2 forms water-mediated contacts to a phosphate group of the DNA strand and makes lattice contacts with a symmetry-related duplex. Wat66 forms a bridge between Wat45 from the coordination sphere of Mg1 and Wat64 coordinated to Mg3. Wat64 also creates a weak hydrogen bond to O2P of T21 from a symmetry-related molecule.

3.5. Crystal packing

The hybrid crystal displays a typical A-DNA packing, with the termini of one molecule abutting the shallow grooves of symmetry-related molecules. Each molecule of the hybrid dodecamer forms contacts with five symmetry-related molecules by means of stacking interactions between terminal base pairs and minor-groove sugar rings, direct lattice contacts involving the O2′ groups of the RNA strand and *via* water-mediated intermolecular interactions (Supplementary Fig. S1[1]). G12 is strongly involved in a hydrogen-bonding interaction with symmetry-related copies of g6, a7(I) and C19,

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: DZ5239). Services for accessing this material are described at the back of the journal.
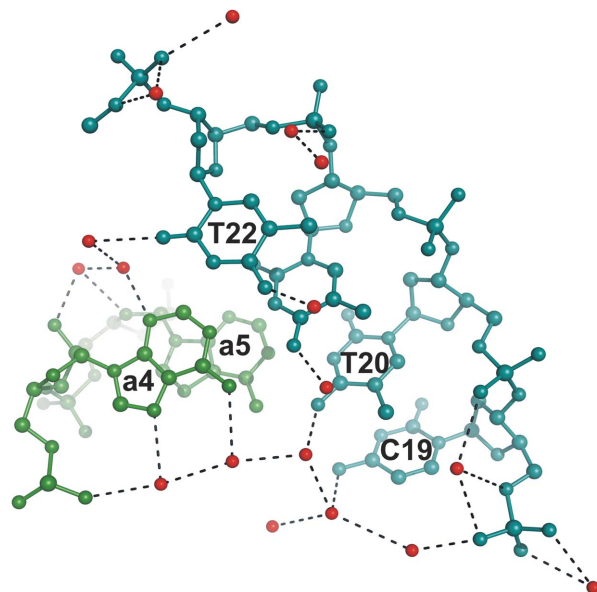

Figure 5
Hydration patterns in the major groove of the present RNA/DNA hybrid. The a4–a5 RNA atoms are shown in green and the C19–T23 DNA atoms are shown in cyan. Water molecules are represented by red spheres. Hydrogen bonds are indicated by dashed lines.
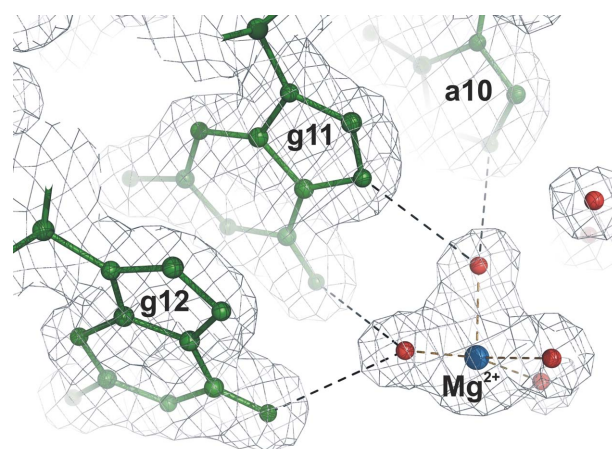

Figure 6
The hydration sphere around the Mg1 Mg^{2+} ion (blue sphere) near the g11–g12 step. The $2F_o − F_c$ map is contoured at the 1σ level.

forming a base-paired pentaplex (Supplementary Fig. S2[1]). The intermolecular contacts with neighbouring molecules are listed in Supplementary Table S2[1].

4. Discussion

In this work, we have studied the structure of an RNA/DNA hybrid duplex with the RNA sequence of the polypurine tract of the HIV genome. The double helix adopts the A-form conformation and is similar in a number of respects to previously determined crystal structures of RNA/DNA duplexes. However, in spite of the similarity (75%) of the nucleic acid sequences used in the crystallization experiments, the present structure and the previous model of the PPT (Kopka *et al.*, 2003) differ in a number of significant ways. The most impor-

tant difference concerns the total absence of any RNA sugar in the C2′-*endo* conformation. This is in contrast to previous findings, which postulated that C2′-*endo* ribose was essential for the biological properties of the PPT, namely its resistance to hydrolysis by the retroviral RNase H domain. In addition, the two structures differ in several other conformational characteristics of the DNA and RNA chains. In particular, the present model has only a single conformation of the DNA strand and there is no dual conformation at nucleotide a3. However, our structure does show increased flexibility of the RNA/DNA hybrid with the PPT sequence, as seen at several places along the RNA strand. In particular, our model reveals dual conformation of the RNA backbone at nucleotides u1–a2 and g6–a8. It is remarkable that, with the exception of a7, the nucleobases of these fragments are clearly defined in only one orientation and are evidently not influenced by the increased flexibility of the RNA ribose-phosphate backbone. It should be emphasized that the g6–a8 sequence is part of the 'unzipping' tract identified in the crystal structure of the RNA/DNA hybrid in complex with HIV-1 reverse transcriptase (Sarafianos *et al.*, 2001). At the disordered segment near the 5′-end of the RNA, the u1 and A24 bases are connected by just one (noncanonical) hydrogen bond, forming a 'pair' which involves an unusual *syn* orientation of the adenosine glycosidic bond. Despite the disordered backbone, the nucleobases of this segment display a single well ordered conformation. The u1·A24 nucleotides reflect the lability of the terminal fragment of the hybrid duplex and have a clear effect on the torsion angles of the adjacent nucleotides. The conformation of this fragment suggests trends for changes in the helix parameters of the duplex which may occur in the vicinity of weakly paired bases. The presence of such drastic changes of the geometry of the RNA strand as observed in this structure and as reported by others (*e.g.* local A-to-B conversion) clearly highlights the deformability of the PPT sequence. Such structural rearrangements can occur at a number of places along the RNA chain and seem to be a prerequisite of the nucleic acid substrate for proper action of HIV-1 reverse transcriptase during synthesis of the retroviral DNA.

References

Arnott, S. (1999). *Oxford Handbook of Nucleic Acid Structure*, edited by S. Neidle, pp. 1–38. Oxford University Press.

Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.

Conn, G. L., Brown, T. & Leonard, G. A. (1999). *Nucleic Acids Res.* **27**, 555–561.

Cross, C. W., Rice, J. S. & Gao, X. (1997). *Biochemistry*, **36**, 4096–4107.

DeLano, W. L. (2002). *PyMOL*. http://www.pymol.org.

Egli, M. (2002). *Chem. Biol.* **9**, 277–286.

Egli, M., Portmann, S. & Usman, N. (1996). *Biochemistry*, **35**, 8489–8494.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.

Fedoroff, O. Y., Salazar, M. & Reid, B. R. (1993). *J. Mol. Biol.* **233**, 509–523.

Feng, Z., Westbrook, J. & Berman, H. M. (1998). Report NDB-407. Rutgers University, New Brunswick, New Jersey, USA.

Heinemann, U., Lauble, H., Frank, R. & Blöcker, H. (1987). *Nucleic Acids Res.* **15**, 9531–9550.

Ilyinskii, P. O. & Desrosiers, R. C. (1998). *EMBO J.* **17**, 3766–3774.

Kennard, O. & Hunter, W. N. (1991). *Angew. Chem. Int. Ed. Engl.* **30**, 1254–1277.

Kopka, M. L., Lavelle, L., Han, G. W., Ng, H. L. & Dickerson, R. E. (2003). *J. Mol. Biol.* **334**, 653–665.

Lavery, R., Moakher, M., Maddocks, J. H., Petkeviciute, D. & Zakrzewska, K. (2009). *Nucleic Acids Res.* **37**, 5917–5929.

Lu, X.-J. & Olson, W. K. (2003). *Nucleic Acids Res.* **31**, 5108–5121.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.

Neidle, S. (2008). *Principles of Nucleic Acid Structure*, p. 33. London: Academic Press.

Olson, W. K. *et al.* (2001). *J. Mol. Biol.* **313**, 229–237.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Painter, J. & Merritt, E. A. (2006). *J. Appl. Cryst.* **39**, 109–111.

Rattray, A. J. & Champoux, J. J. (1989). *J. Mol. Biol.* **208**, 445–456.

Sarafianos, S. G., Das, K., Tantillo, C., Clark, A. D., Ding, J., Whitcomb, J. M., Boyer, P. L., Hughes, S. H. & Arnold, E. (2001). *EMBO J.* **20**, 1449–1461.

Sarafianos, S. G., Marchand, B., Das, K., Himmel, D. M., Parniak, M. A., Hughes, S. H. & Arnold, E. (2009). *J. Mol. Biol.* **385**, 693–713.

Schneider, B., Cohen, D. & Berman, H. M. (1992). *Biopolymers*, **32**, 725–750.

Schneider, B., Morávek, Z. & Berman, H. M. (2004). *Nucleic Acids Res.* **32**, 1666–1677.

Shaw, N. N. & Arya, D. P. (2008). *Biochimie*, **90**, 1026–1039.

Svozil, D., Kalina, J., Omelka, M. & Schneider, B. (2008). *Nucleic Acids Res.* **36**, 3690–3706.

Tonelli, M., Ulyanov, N. B., Billeci, T. M., Karwowski, B., Guga, P., Stec, W. J. & James, T. L. (2003). *Biophys. J.* **85**, 2525–2538.

Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.

Xia, T., SantaLucia, J., Burkard, M. E., Kierzek, R., Schroeder, S. J., Jiao, X., Cox, C. & Turner, D. H. (1998). *Biochemistry*, **37**, 14719–14735.

Xiong, Y. & Sundaralingam, M. (1998). *Structure*, **6**, 1493–1501.

Yi-Brunozzi, H. Y., Brinson, R. G., Brabazon, D. M., Lener, D., Le Grice, S. F. & Marino, J. P. (2008). *Chem. Biol.* **15**, 254–262.