256–268 *Nucleic Acids Research, 2007, Vol. 35, No. 1* *Published online 12 December 2006*
doi:10.1093/nar/gkl909

Purine analog substitution of the HIV-1 polypurine tract primer defines regions controlling initiation of plus-strand DNA synthesis

Jason W. Rausch and Stuart F. J. Le Grice*

RT Biochemistry Section, Resistance Mechanisms Laboratory, HIV Drug Resistance Program, National Cancer Institute at Frederick, Frederick, MD 21702, USA

Received September 8, 2006; Revised October 11, 2006; Accepted October 12, 2006

ABSTRACT

Despite extensive study, the mechanism by which retroviral reverse transciptases (RTs) specifically utilize polypurine tract (PPT) RNA for initiation of plus-strand DNA synthesis remains unclear. Three sequence motifs within or adjacent to the purine-rich elements are highly conserved, namely, a rU:dA tract region immediately 5′ to the PPT, an rA:dT-rich sequence constituting the upstream portion of the PPT and a downstream rG:dC tract. Using an *in vitro* HIV-1 model system, we determined that the former two elements define the 5′ terminus of the (+)-strand primer, whereas the rG:dC tract serves as the primary determinant of initiation specificity. Subsequent analysis demonstrated that G→A or A→G substitution at PPT positions −2, −4 and +1 (relative to the scissile phosphate) substantially reduces (+)-strand priming. We explored this observation further using PPT substrates substituted with a variety of nucleoside analogs [inosine (I), purine riboside (PR), 2-aminopurine (2-AP), 2,6-diaminopurine (2,6-DAP), isoguanine (iG)], or one of the naturally occurring bases at these positions. Our results demonstrate that for PPT positions −2 or +1, substituting position 2 of the purine was an important determinant of cleavage specificity. In addition, cleavage specificity was greatly affected by substituting −4G with an analog containing a 6-NH2 moiety.

INTRODUCTION

The polypurine tract (PPT[1]) of retroviruses and long terminal repeat-containing retrotransposons is a short RNA sequence from which second or plus (+) strand DNA synthesis initiates (1). Subsequent to this, the PPT primer must be accurately removed from nascent DNA to create a double-stranded, integration-competent DNA provirus. A stringent requirement is therefore that the PPT is precisely recognized by both the ribonuclease H (RNase H) and DNA polymerase domains of its cognate reverse transcriptase (RT). Using oligonucleotide-derived substrates, model systems accurately recapitulating (i) selection of the PPT 3′ terminus from within an RNA/DNA hybrid (ii) DNA-dependent DNA synthesis from the PPT 3′ terminus and (iii) removal from nascent DNA have been constructed. At the same time, several PPT mutations have now been studied *in vivo*, using recombinant viruses. Surprisingly, despite a wealth of crystallographic (2), biochemical (3–13) and cell culture data (7,14–16), the structural basis for PPT selection and utilization in (+)-strand DNA synthesis remains elusive.

The unusual pattern of base pairing observed in a co-crystal of HIV-1 RT and a PPT-containing RNA/DNA hybrid (2) suggested that local flexibility might promote sequestering and orienting of the retroviral polymerase. To explore this notion, we have inserted a variety of nucleoside analogs into PPT-containing RNA/DNA hybrids to alter their structure in a manner that minimally alters sequence context. Examples include pyrimidine isosteres, or shape mimics, which locally remove hydrogen bonding (12,17), abasic tetrahydrofuran linkages, which remove the nucleobase while preserving the sugar-phosphate backbone (18,19), pyrrolo-dC, a fluorescent cytosine analog providing information on local base pairing (20) and locked nucleic acid monomers, which locally stabilize the phosphate backbone. This strategy demonstrated (i) that the junction between the rG:dC and upstream rA:dT tract is particularly sensitive to altering hydrogen bonding (17) (ii) weakened base pairing at position −2 (20) and (iii) that DNA nucleobases on either side of the PPT-U3 junction are not required for accurate cleavage (18).

Although our studies to date have focused primarily on individual bases of the PPT and its DNA complement, the modular architecture of the RNA/DNA hybrid may also be significant. A comparison of several retroviral (+)-strand primers (7,21) indicated a conserved rU:dA tract immediately 5′ to the PPT, an rA:dT-rich sequence constituting the upstream portion of the PPT, and a downstream rG:dC tract, the 3′ terminus of which is almost always flanked by

*To whom correspondence should be addressed. Tel: +1 301 846 5256; Fax: +1 301 846 6013; Email: slegrice@ncifcrf.gov

from one to three A's. Classically, rA:dT tracts have been associated with minor groove compression and DNA bending (21–27), while rG:dC tracts are reported to induce major groove compression (28). While the complete structure of the PPT-containing RNA/DNA hybrid was not solved by Sarafianos *et al*. (2) these authors did observe minor groove compression in one of the two rA:dT tracts within the HIV-1 PPT. Thus, we cannot rule out that alterations in groove width or malleability induced by juxtaposed homopolymer tracts may be exploited by the retroviral polymerase as a means of correctly 'docking' onto the PPT as either a DNA polymerase or RNase H. The goal of the present study was 2-fold, namely (i) to investigate the significance of the modular architecture of the HIV-1 PPT by displacing or interrupting homopolymeric stretches and (ii) if particularly critical regions were identified, to investigate this in more detail by targeted insertion of purine analogs into the RNA strand.

Using a model system in which (+)-strand selection and initiation are recapitulated *in vitro*, we have determined that an important function of the rU:dA tract in HIV-1 is as a specificity determinant for 5′ terminal cleavage of the (+)-strand primer, which in turn assures that the primer is of a length sufficient for recognition and processing by HIV-1 RT. The rA:dT tracts likewise add length to the primer, and assures that only negligible internal cleavage of the PPT occurs. However, the element most critical to PPT function is the rG:dC tract, which both resists internal hydrolysis and, more importantly, directs cleavage to the 5′-G-A-3′ PPT-U3 junction. This occurs not only in the context of the wild-type PPT, but also when the rG:dC tract is relocated to alternative sites of the PPT. In keeping with previous data (29), −2G and −4G (defining −1G as the base 5′ to the scissile bond) and +1A were found to be essential for directing HIV-1 RT to cleave precisely at the PPT-U3 junction. The structural basis for these findings were investigated further using an oligonucleotide-based RNase H assay in which a variety of purine analogs were substituted for G or A at positions −4, −2 and +1. Our results show a strong correlation between the efficiency and precision of RNase H-mediated hydrolysis and the identity of the chemical moieties at the 2- and 6-positions of the purine ring in nucleosides at PPT positions −4, −2 and +1. Potential explanations for these findings, especially the possible involvement of RNase H primer grip residues Q475 and R448 (2,30), are discussed.

MATERIALS AND METHODS

Oligodeoxyribonucleotides and oligoribonucleotides

DNA oligonucleotides for PCR and the PPT selection assay were purchased from Integrated DNA Technologies (Coralville, IA). RNA oligonucleotides were synthesized at 1 μmol scale on a PerSeptive Biosystems Expedite synthesizer model 8909. Purine riboside (PR), 2-aminopurine (2-AP), 2,6-diaminopurine (2,6-DAP) and standard RNA phosphoramidites were purchased from Glen Research (Sterling, VA), while inosine (I) and isoguanosine (iG) were purchased from ChemGenes (Wilmington, MA). Stepwise coupling yields for incorporation of each amidite were determined by trityl cation monitoring. Base deprotection and cleavage of oligonucleotides from the CPG support were carried out by incubation with 37% methylamine in 50% ethanol for 6 h at room temperature. Subsequent 2′ deprotection was by incubation with tetrabutylammonium fluoride at 25°C for 24 h. Oligonucleotides were purified by preparative PAGE and quantified by ultraviolet (UV) spectroscopy.

Substrates for (+)-strand primer selection/extension assay

RNA/DNA hybrids used in the PPT selection/extension assay were created by annealing equimolar amounts of 100 nt RNA containing the 3′ PPT and flanking sequences of HIV-1 (Hxb2 strain) to complementary, single-stranded DNA (2 μM each) in 10 mM Tris–HCl (pH 7.6), 25 mM NaCl. The reaction was heated to 86°C for 2 min then cooled to 4°C at a rate of 5°C/min prior to use. RNAs were generated by *in vitro* transcription (MegaShortScript Kit, Ambion, Austin, TX) from nested PCR products created from four overlapping DNA oligonucleotides. Two flanking oligonucleotides, one of which contained the T7-promoter, were used to amplify internal oligonucleotides containing overlapping viral sequences. All PPT variants were contained exclusively within the downstream internal oligonucleotide, which was exchanged in each reaction to introduce the mutations summarized in Table 1. Single-stranded DNA was obtained from PCR products generated in the same manner, except that the flanking oligonucleotide containing the T7-promoter was replaced with a 5′-phosphorylated oligonucleotide containing only viral sequence. The hemi-phosphorylated amplification product was treated with T7 exonuclease (Epicentre Technologies, Madison, WI) to remove the phosphorylated strand, after which the remaining strand was purified by PAGE and quantified by UV spectroscopy.

(+)-Strand primer selection/extension assay

The method used to evaluate (+)-strand DNA synthesis from primers embedded within RNA–DNA hybrids has been described previously (6). Briefly, 100 bp hybrid substrates (50 nM; see above) were incubated with purified, recombinant HIV-1 RT (50 nM), 50 μM dNTPs and [α-^{32}P]dTTP in reaction buffer [25 mM Tris–HCl (pH 8.0), 80 mM NaCl, 6 mM $MgCl_2$ and 5 mM DTT] for 30 min at 37°C. Reactions were terminated by mixing with 10 mM EDTA, after which 50% of each mixture was incubated with 0.3 vol 1 N NaOH at 65°C for 20 min to achieve complete RNA hydrolysis. Reaction products were then precipitated (50% isopropanol and 3.5 M NH_4OAc), re-suspended in a urea-based gel-loading buffer, fractionated over a 15% polyacrylamide gel containing 7 M urea, and detected by phosphorimaging. Sites of (+)-strand initiation were established by comparing migration of DNA synthesis products to a sequencing ladder, while (+)-strand primer lengths were determined by counting bands resulting from low-level RNA degradation. The latter were only evident upon extreme exposure of the gels in question.

PPT selection assay

PPT selection was evaluated as described previously (17). Substrates consisted of a series of PPT-containing RNA/DNA

Table 1. Sequence of HIV-1 PPT variants containing rG rA and rA rG substitutions both upstream and downstream of the PPT-U3 junction

Wild type PPT

PPT Mutant	RNA Sequence	PPT Mutant	RNA Sequence
PS-1	5'UUUUUAAAAGGGGGGGAAAAA3'	PS-16	5'UUUUUAAAAGAAAAGGGAGAA3'
PS-2	5'UUUUUGGGGGGAAAAGAAAAA3'	PS-17	5'UUUUUAAAAGAAAAGAGAGGA3'
PS-3	5'UUUUUAAAAGAAAAAGAGAGA3'	PS-18	5'UUUUUAAAAGAAAAGAGAGAA3'
PS-4	5'UUUUUGAGAGAGAGGGGGGGA3'	PS-19	5'UUUUUAAAAGAAAAGGAGAGA3'
PS-5	5'UUUUUGAGAGAGAGAGAGAGA3'	PS-20	5'UUUUUAAAAGAAAAAGAGGGA3'
PS-6	5'UUUUUAAAAAAAAAAGGGGGGA3'	PS-21	5'UUUUUAAAAGAAAGAGGGGGA3'
PS-7	5'AAAAGAAAAGGGGGGA3'		

PPT Mutant	RNA Sequence	PPT Mutant	RNA Sequence
PS-8	5'UUUUUAAGAAAAGAAAAGGGGGGA3'	PS-22	5'UUUUUAAAAGAAAAGGGGGGG3'
PS-9	5'UUUUUAAAAGAAAAGAAAAGGGGGGA3'	PS-23	5'UUUUUAAAAGAAAAGGGGGAA3'
PS-10	5'AAAAGAAAAGAAAAGGGGGGA3'	PS-24	5'UUUUUAAAAGAAAAGGGGAGA3'
PS-11	5'UUUUUAGAAAAGGGGGGA3'	PS-25	5'UUUUUAAAAGAAAAGGGAGGA3'
PS-12	5'UUUUUAAAAGGGGGGA3'	PS-26	5'UUUUUAAAAGAAAAGGAGGGA3'
PS-13	5'UUUUUGGGGGGA3'	PS-27	5'UUUUUAAAAGAAAAGAGGGGA3'
PS-14	5'UUUUUAAAAGAAAAGGGGGGAAAA3'	PS-28	5'UUUUUAAAAGAAAAAGGGGGA3'
PS-15	5'UUUUUAAAAGAAAAGGGGGGAGGG3'	PS-29	5'UUUUUAAAAGAAAGGGGGGGA3'

The sequence of the wild-type PPT is indicated above the table. Note that, for these variants, the PPT sequence was embedded within a larger RNA/DNA hybrid (see Figure 1a).

hybrids comprised of a 29 nt, 5′ ^{32}P-end-labeled RNA hybridized to a 40 nt DNA such that both RNA termini are recessed. RNAs contained either the wild-type (+)-strand primer extended by 10 nt at the 3′ terminus (i.e. 5′-UUUUA-AAAGAAAAGGGGGGACTGGAAGGG-3′), or a mutant sequence in which one of several purine analogs were substituted for the native nucleoside at position −4, −2 or +1. Note that for each substitution in the RNA strand, an appropriate complementary substitution was made in the DNA strand to insure that appropriate inter-strand base pairing would occur (i.e. rG:dC, rA:dT, rI;dC, rPR:dT, r2-AP:dT, r2,6-DAP:dT or riG:diC). RNA/DNA hybrids (100 nM) were incubated with HIV-1 RT (50 nM) in reaction buffer at 37°C. Reactions were terminated after 1, 3, 10 or 30 min by mixing with an equal volume of a urea-based gel-loading buffer. Cleavage products were fractionated by denaturing PAGE detected by phosphorimaging and quantified using QuantityOne software (BioRad, Hercules, CA).

RESULTS

A model system for (+)-strand selection and extension

To investigate the importance of the rU:dA, rA:dT and rG:dC tracts, a model system was developed in which RNA/DNA hybrids containing variants of the 3′ HIV-1 PPT and flanking

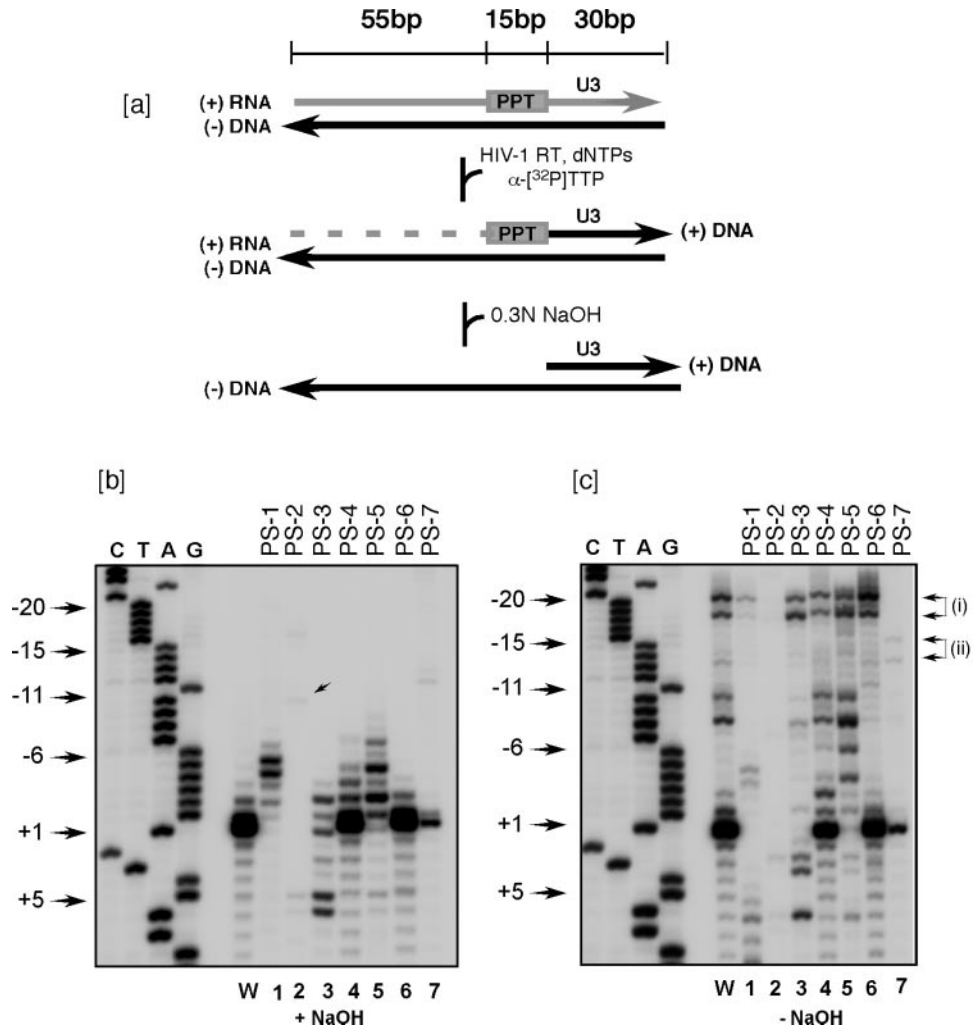


Figure 1. (**a**) Schematic representation of HIV-1 PPT cleavage assay. The substrate is a 100 bp RNA/DNA hybrid, within which the PPT is embedded. In the presence of HIV-1 RT and a dNTP cocktail, the hybrid is cleaved to release the PPT 3′ terminus, which is subsequently extended by the DNA polymerase activity of the same enzyme. Following DNA synthesis, RNase H activity releases nascent DNA. Alternatively, the PPT primer is removed by treatment with NaOH. Note that this scenario reflects those events anticipated to occur in the presence of the wild-type substrate. Different patterns of RNase H cleavage and/or plus-strand initiation may be observed in the presence of mutant substrates. (**b** and **c**) Effect of displacing and disrupting the homopolymeric tracts of the HIV-1 PPT on cleavage specificity. Lanes 1–7, PPT substrates PS1–PS7, respectively (Table 1), and Lane W, the unsubstituted PPT. For correct location of nascent DNA, a sequence ladder was run in parallel (Lane C, T, A and G). Base numbering −20 to +5 is with respect to the 5′ terminus of nascent (+) DNA, which we denote as +1, and the rG 5′ to this as −1. In (b), the PPT primer was removed by NaOH treatment, while (c) depicts the (+)-strand products arising from RT-associated RNase H. Arrows: (b), marks (+)-strand product derived from initiation of DNA synthesis from the 3′-terminus of the displaced rG:dC tract within substrate PS-2 (lane 2; position −10); (c), [i] and [ii] denote 5′ termini of plus-strand primers derived from wild-type substrate (WT; lane W) and substrate from which the rU:dA tract has been deleted (PS-7; lane 7), respectively.

sequences are incubated with purified HIV-1 RT in the presence of a full complement of deoxynucleotides, one of which is ^{32}P-radiolabeled (Figure 1a). This system monitors both cleavage of the PPT RNA strand and (+)-strand DNA synthesis from RNA fragment(s) remaining hybridized to the (−) DNA template. By comparing the migration of NaOH-treated and untreated (+) DNA synthesis products to that of a sequencing ladder derived from the wild-type substrate, (+)-strand initiation sites, as well as the length(s) of the (+)-strand primers, can be determined. The relative efficiency with which (+)-strand DNA synthesis occurs among various substrates can be likewise assessed. It is important to note that while numerous transition mutations, deletions or insertions were introduced into the PPT and rU:dA tract embedded within the 100 bp hybrid, flanking sequences are common to all substrates. A summary of the substrates utilized is presented in Table 1.

U-, A- and G-tracts have distinct roles in (+)-strand primer selection

To examine the function(s) of individual homopolymeric PPT tracts in (+) DNA synthesis, substrates were generated in which the rU:dA, rA:dT and rG:dC tracts were disrupted, deleted and/or transposed (Table 1; WT, PS-1–PS-7). DNA synthesis profiles derived from these substrates are depicted in Figure 1b. Lane W illustrates specific and efficient initiation of (+)-strand synthesis from the 3′ terminus of the

wild-type PPT. Because the first nucleotide corresponds to +1A at (Table 1, top), the primary synthesis product co-migrates with the equivalent band of the sequencing ladder. Note also that residual RNA primers are not observed here, since reaction products of Figure 1b were subjected to alkaline hydrolysis. However, because this treatment occasionally leaves a nucleoside fragment(s) attached to the 5′ terminus of the nascent DNA, the primary reaction product in this case actually migrates as a doublet, which is evident in lighter gel exposures (data not shown).

DNA synthesis profiles of substrates PS-1 and PS-2, in which the position of the rG:dC tract is shifted within the PPT from its usual 3′ terminal location, are shown in Figure 1b, lanes 1 and 2. Remarkably, we observe a shift in the primary initiation site in register with the repositioned rG:dC tract to position −4/−5 in substrates PS-1, such that similar to the wild-type PPT, synthesis originates from the 3′ terminus of this tract. A similar shift is evident in lane 2 (substrate PS-2), although the hydrolysis product is barely detectable (arrow). These data demonstrate that the primary specificity determinants for initiation of (+)-strand DNA syntheses are contained within the rGdC tract. Furthermore, the functions of the rA:dT and rG:dC tracts are not interchangeable, since (+)-strand DNA synthesis is not efficiently initiated from the 3′ terminus of the modified PPTs.

PS-3, PS-4 and PS-5 contain PPTs in which the rG:dC tract, rA:dT tract or both, are disrupted, while their all-purine composition remains unchanged. (+)-DNA synthesis profiles derived from PS-3 and PS-5 demonstrate that interrupting the rG:dC tract reduces the efficiency and alters the specificity with which (+)-strand synthesis is initiated (Figure 1b, lanes 3 and 5). Note also that (+)-strand products derived from these substrates appear to initiate primarily from 5′-G-A-3′ rather than from 5′-A-G-3′ junctions, suggesting a potential conformational preference for HIV-1 RT-mediated cleavage within this purine tandem. Selective disruption of the rA:dT tracts (substrate PS-4, Figure 1a, lane 4) (31) on (+)-strand priming appear relatively minor, in agreement with previous data indicating that this sub-motif is relatively insensitive to mutation (29,32). Similarly, removing the single rG discontinuity from within the rA:dT tract has no affect on (+)-strand priming (Figure 1b, lane 6). Conservation of this discontinuity among retroviruses may prevent pausing (or slippage) that might occur during minus-strand DNA synthesis over an extended homopolymeric rA:dT tract, as was observed during *in vitro* analysis of minus-strand DNA synthesis over the RNA strand of the PS-6 substrate (data not shown). Finally, deleting the rU:dA tract (Figure 1b, lane 7; substrate PS-7) substantially reduces priming efficiency, perhaps due to 5′ terminal primer truncation (see below).

In order to gain insight into the nature of the PPT primer, (+)-strand DNA synthesis products of Figure 1c were not treated with NaOH. For wild-type RT (Figure 1c, lane W), ∼10% of the reaction products contain either one of two 'full-length' (+)-strand primers, or fragments thereof. The former, originally described by Huber and Richardson (33), derive from RT-mediated cleavage at positions −17 and −19, i.e. within the rU:dA tract. Products from mutant substrates PS-1–PS-5 indicate multiple cleavages within the PPT primer (Figure 1c, lanes 1–5). However, almost no internal cleavage is observed when the discontinuity in the A-tract is removed (Figure 1c, lane 6, substrate PS-6).

Despite evidence of variable internal cleavage of the PPT, hydrolysis is consistently observed at positions −19/−17 among WT and mutant substrates PS-1, PS-3, PS-4, PS-5 and PS-6, although the extent of cleavage varies with the amount of (+)-strand product generated [Figure 1c, arrows

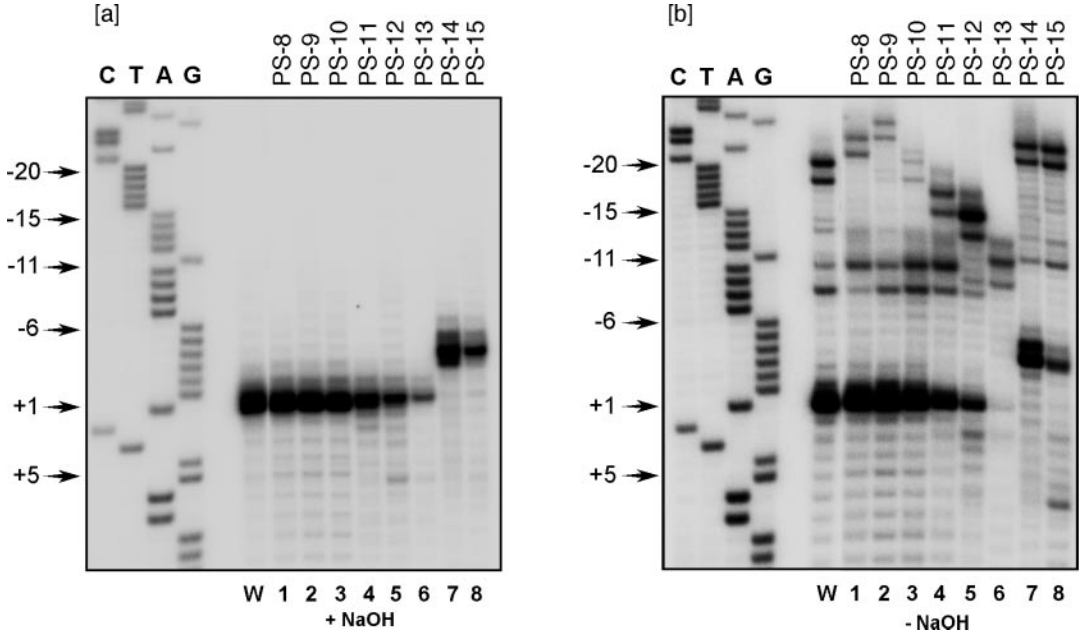


Figure 2. (**a** and **b**) Effect of altering the length of the HIV-1 PPT on cleavage specificity. Lanes 1–8, PPT substrates PS8–PS15, respectively (Table 1), and lane W, the unsubstituted PPT. Additional lane notations and migration positions are as in the Legend to Figure 1. (**a**) (+)-strand DNA products following PPT removal by NaOH treatment. (**b**) (+)-strand DNA products derived from RT/RNase H hydrolysis of the RNA/DNA hybrid.

(i)]. This suggests that neither A- nor G-tracts play a direct role in defining the 5′ terminus of the (+)-strand primer. In contrast, substrate that does not contain the U-tract (PS-7) is cleaved at positions −14/−12 [lane 7, arrows (ii)], generating PPT primers exactly 5 nt shorter than the respective equivalent cleavages (i.e. −19/−17) derived from the wild-type substrate. Coupled with the observation that considerably less (+)-strand DNA synthesis product is generated from substrate PS-7 compared to wild-type, this indicates that one of the functions of the U-tract may be to extend the length of the (+)-strand primer.

(+)-Strand priming efficiency is directly related to primer length

To better define relationships among PPT rU:dA tract sequence, primer length and primer function, a second set of substrates was prepared containing PPTs of various lengths and in which rU:dA and rA:dT tract sequences were exchanged (Table 1; PS-8–PS-15). DNA synthesis profiles derived from these substrates are shown in Figure 2a and b.

In substrates PS-8 and PS-9, the PPT was lengthened by inserting 5′-rArArG-3′ and 5′-rArArArArG-3′ (with DNA complements), respectively, between the rU:dA and rA:dT tracts, elongating the (+)-strand primer by 3 or 5 nt relative to wild-type (Figure 2b, lanes 1 and 2). Such a result indicates that despite the insertions, cleavage defining the 5′ termini of these RNA fragments occurs within the rU:dA tract at positions equivalent to those of the wild-type substrate. The insertions of substrates PS-8 and PS-9 did not affect either the specificity of cleavage at the PPT 3′ terminus, nor the efficiency with which (+)-strand DNA synthesis was initiated (Figure 2a, lanes 1 and 2). Lane 3 in Figure 2a and b illustrate processing of substrate PS-10, in which a 5′-rArArArArG-3′ sequence was substituted for the U-tract to determine whether the two motifs are interchangeable. Despite this alteration, processing of the mutant substrate was unaffected, with equivalent cleavages occurring both at positions −19/−17 and at the PPT-U3 junction. The efficiency of (+)-strand initiation was also unaffected. The only significant difference between the two profiles is that cleavage internal to the PPT (positions −7 and −9) was more pronounced for the mutant substrate. Since internal PPT cleavage may adversely affect either (+)-strand initiation or primer removal, this difference may reflect a selective advantage of the wild-type sequence *in vivo*.

Truncating the PPT (substrates PS-11–PS-13; Figure 2a and b, lanes 4–6) both reduces the length of the (+)-strand primer (Figure 2b) and the efficiency of (+)-strand DNA synthesis (Figure 2a). The efficiency with which RT removes the PPT primer is also diminished. Taken together with the results of Figure 1b, lane 7, these data suggest that an important function of both the rU:dA and rA:dT tracts is to insure that the (+)-strand primer is of sufficient length to permit efficient processing by HIV-1 RT. Indeed, it appears that reducing primer length by 3 nt substantially affects priming efficiency (compare lanes W and 4 of Figure 2b), with the most severe effects observed upon complete removal of the rA:dT tract (Figure 2b, lane 6).

Finally, inserting 5′-rArArA-3′ or 5′-rGrGrG-3′ at the 3′ terminus of the PPT (substrates PS-14 and PS-15, respectively) shifts the 3′ terminus of the (+)-strand primer 3 nt upstream (Figure 2a and b, lanes 7 and 8). Although the 5′-rGrGrG-3′ insertion reduces the efficiency of (+)-strand initiation and perhaps primer removal, the specificity of initiation is still dictated by the native rG:dC tract.

PPT 3′ terminal sequences are variably sensitive to mutation

To this point, the PPT and sequences immediately upstream have been considered as a collection of rU:dA, rA:dT and rG:dC tracts, of which we have determined that the rG:dC tract defines the site of (+)-strand initiation. Figure 3 depicts (+)-strand DNA synthesis profiles derived from substrates in which multiple (substrates PS-16–PS-21, PS-3; Figure 3a) or point mutations (substrates PS-22–PS-29; Figure 3b) were introduced within the rG:dC tract and/or adjacent 5′-rA-rG-3′ or 5′-rG-rA-3′ junctions to determine which bases of these motif are most critical to PPT function. Remarkably, dual and triple rG→rA transition mutations at positions −1, −3 and −5, or tandem transitions at positions −6/−7, have almost no effect on (+)-strand priming (Figure 3a, lanes 1–3 and 7). In contrast, equivalent changes at positions −2, −4 and −6 reduced priming efficiency and altered initiation specificity (Figure 3a, lanes 4–6; Figure 1b, lane 3), indicating that specific nucleosides within an apparently homogeneous rG:dC tract are of particular importance to its overall function. This notion is strengthened by the data of Figure 3b. Again, rG→rA substitutions at positions −1 or −3 (lanes 2 and 4, respectively) have little effect on (+)-strand DNA synthesis, whereas the same change at positions −2 or −4 influences the efficiency and specificity of (+)-strand priming (Figure 3b, lanes 3 and 5). Similar changes were observed with a +1 rA→rG substitution, continuing a pattern where alternating nucleosides appear to be important. rG→rA or rA→rG changes at positions −5, −6 and −7 (Figure 3b, lanes 6–8, respectively) have moderate effect, with the greatest reduction in priming observed with a rG→rA transition at position −6. A similar phenotype was noted by Pullen *et al*. (29), although in their study, a +1 rA→rG change was notable primarily for a 1 nt shift in the site of (+)-strand initiation, rather than a reduction in efficiency.

Using nucleoside analogs to dissect G-tract function. The results of Figure 3 suggest that residues −4rG, −2rG and +1rA, have particular significance in terms of overall PPT function. A recent study further demonstrated that these sites may serve as important *cis*-acting signals directing RNase H-mediated cleavage by HIV-1 and Moloney murine leukemia virus RTs (34). To examine the structural basis for these observations, PPT-containing oligoribonucleotide substrates were generated in which purine analogs were incorporated at positions +1, −2 or −4. A summary of the chemical structures and properties of these analogs is presented in Table 2. Synthetic RNAs were 5′ end-labeled and annealed to complementary DNA oligonucleotides, and the resulting hybrids subjected to RNase H-mediated hydrolysis by HIV-1 RT. Cleavage profiles derived from substrates containing purine analog substitutions for +1rA are depicted in Figure 4. As previously shown, HIV-1 RT principally cleaves at the −1rG/+1rA phosphodiester bond defining the PPT-U3

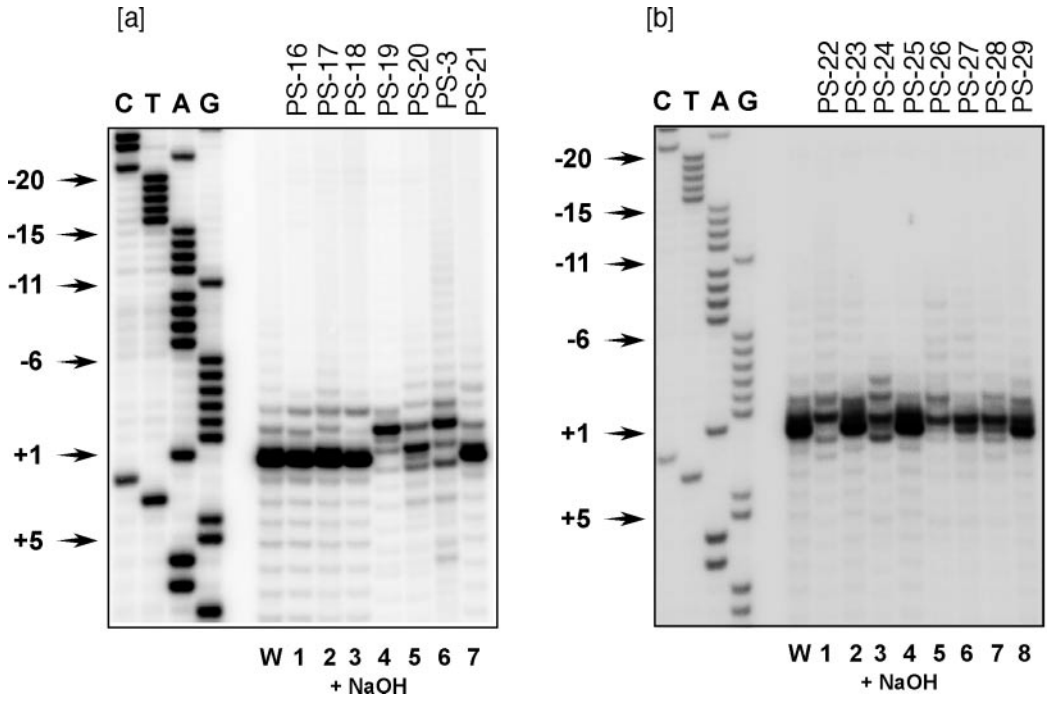
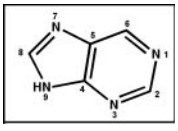


Figure 3. (**a**) Effect of individual and multiple rG→rA substitutions within the PPT G-tract on PPT cleavage specificity. Lanes 1–7, mutants PS16–PS22, respectively. (**b**), Effect of rG→rA scanning of the PPT G-tract on cleavage specificity. Lanes 1–8, PPT mutants PS23–PS30, respectively. In (a) and (b), the (+)-strand DNA products have been treated with NaOH. Additional lane notations are as in the legend to Figure 1.

Table 2. Properties of purine analogs



	Purine analog						
	G	A	I	PR	2-AP	2,6-DAP	iG
Pairs with	C	T	C	T	T	T	iC
Interstrand H-bonds	3	2	2	1	2	3	3
2-position[a]	NH_2	H	H	H	NH_2	NH_2	O
6-position[a]	O	NH_2	O	H	H	NH_2	NH_2
N1 protonated?[b]	Yes	No	Yes	No	No	No	Yes

Numbering conventions for purines are indicated (upper left).
[a]Exocyclic moiety bonded to the carbon at this position.
[b]At pH 8.0.

junction (Figure 4a and Table 1). After 30 min, this cleavage constitutes 80% of the total hydrolysis product [Figure 4a, lane 5, position −1). In contrast, only 25% of substrate containing a +1rG is cleaved at the PPT-U3 junction (Figure 4b, lane 5). This result is consistent with the DNA synthesis profile from similarly mutated substrate PS-23 (Figure 3b, lanes 1), suggesting reduced levels of (+) strand synthesis derived from this hybrid duplex may be the indirect result of reduced cleavage at the PPT-U3 junction.

Hydrolysis profiles of substrates containing 2-AP or 2,6-DAP at position +1 were similarly altered, with the extent of cleavage at position +2 exceeding that at the PPT-U3 junction (Figure 4e and f). Conversely, the hydrolysis profiles of RNA containing +1I or +1PR closely resemble that of the wild-type substrate. Finally, while hydrolysis at positions +2 and +3 is virtually abolished, the iG-containing substrate is cleaved primarily at the PPT-U3 junction.

As indicated in Table 2, analogs adversely affecting cleavage at the PPT-U3 junction when present at position +1 (i.e. G, 2-AP and 2,6-DAP) each contain a 2-amino substitution. In contrast, the same site in purines tolerated at +1 is occupied either by hydrogen (A, I and PR) or oxygen (iG), indicating that efficient hydrolysis at the PPT-U3 junction is facilitated by the absence of a 2-NH_2 group of the purine immediately 3′ to the scissile phosphate. Although this effect is likely mediated by steric hindrance in the vicinity of the RNase H active center (see Discussion), the fact that the oxygen of iG is tolerated suggests that the partial charge of the group at this position may also play a role in substrate recognition.

A complementary conclusion can be derived from the results of substituting PPT −2rG (Figure 5). Introducing rA at this position (Figure 5b) broadens cleavage specificity to include positions +1, −2 and −3 as well as the PPT-U3 junction. In addition, although hydrolysis at the PPT-U3 junction still represents the primary cleavage product, it constitutes only 32% of total cleavage, in contrast to 80% for the wild-type substrate. Similar effects were observed upon substitution of I, PR and iG for −2rG (Figure 5c, d and g), except that in these cases, the PPT-U3 junction is no longer the primary cleavage site. Note that in each of these profiles, significant cleavage is observed in the vicinity of the analog substitution (i.e. at position −2 and −3), indicating that analog-induced local structural changes have rendered the rG:dC tract susceptible to internal hydrolysis. Only the profiles of substrates containing 2-AP or 2,6-DAP at position −2 are unaffected by analog substitution (Figure 5e and f),

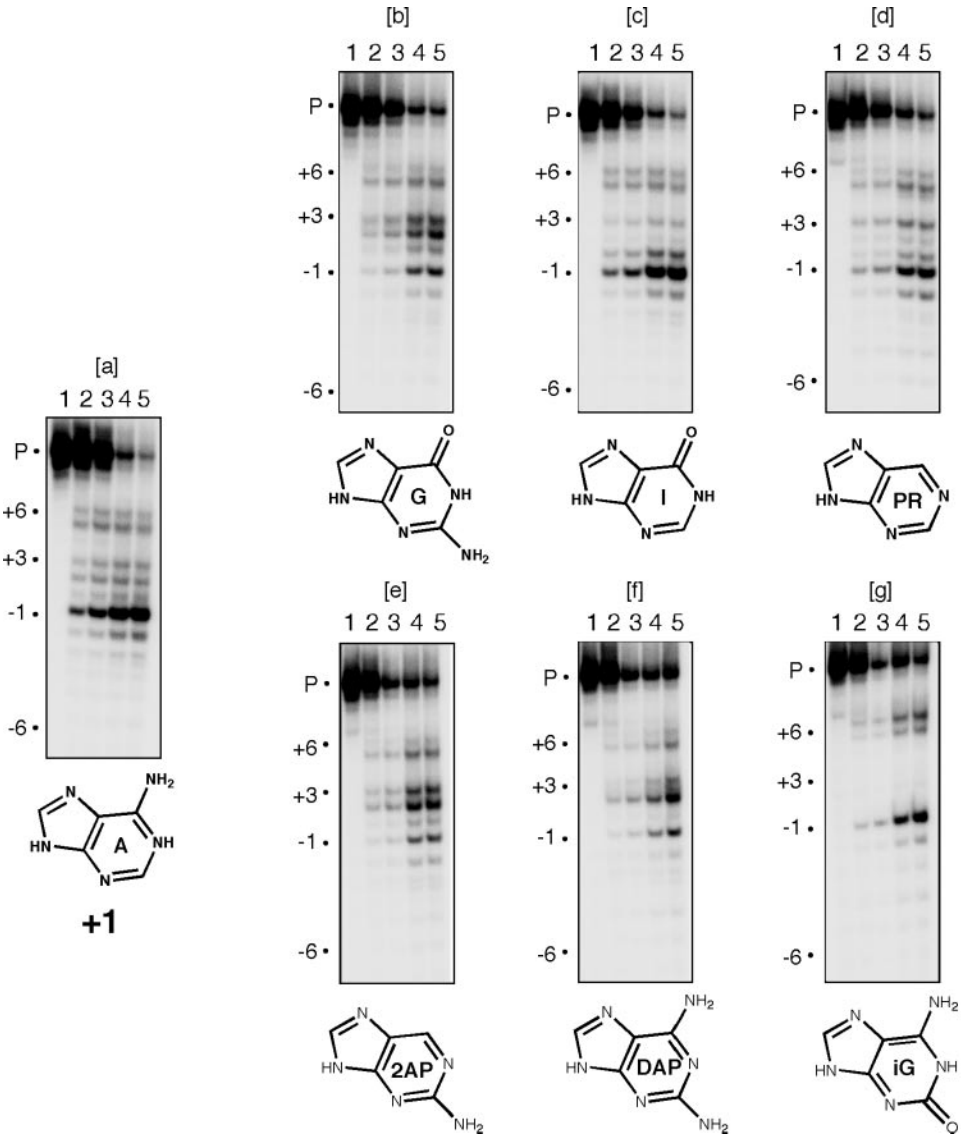


Figure 4. Effect of purine analog substitutions immediately 3′ to the PPT-U3 junction (i.e. +1rA) on cleavage specificity. For these experiments, the PPT RNA primer was synthesized chemically (**a**) wild-type PPT, i.e. +1A; (**b**) +1G; (**c**) +1 PR; (**d**) +1 2-AP; (**e**) +1 2,6-DAP; (**f**) +1 iso-G (iG). The structures of the +1 purine the analogs with which it is substituted, are indicated under (a–f). In each panel, Lanes 1–5 denote samples removed after 0, 1, 3, 10 and 30 min, respectively for analysis.

demonstrating that a purine 2-amino group is essential for resistance to hydrolysis within the G-tract and efficient cleavage at the PPT-U3 junction.

The results of analog substitution at position −4rG suggest the chemical requirements for the purine at this position differ from those of the purine at position −2. The hydrolysis profile resulting from a −4I substitution is unaffected (Figure 6c), suggesting the identity of the exocyclic group at position 2 of the purine is not critical in this context. In contrast, cleavage of substrate containing a −4A substitution is less precise (Figure 6b). Although the primary cleavage occurs at the PPT-U3 junction after 30 min, this constitutes only 30% of the total cleavage product. In addition, internal PPT cleavage is observed at position −6, and to a lesser extent at position −5, indicating a local susceptibility within the rG:dC tract similar to that observed upon introducing rA at position −2.

The cleavage patterns of substrates containing PR or 2-AP (Figure 6d and e) are intermediate between those of I- and A-containing substrates. With respect to the percentage of total hydrolysis that occurs at the PPT-U3 junction after 30 min, substrates containing analogs at position −4 rank as follows: G(WT) > PR > 2-AP > A. Finally, 2,6-DAP and iG are poor substitutes for −4G , despite the fact that each participates in three inter-strand hydrogen bonds. The precision with which substrate containing 2,6-DAP is cleaved is greatly reduced relative to wild-type, with equivalent hydrolysis observed between positions +3 and −3 after 30 min (Figure 6f, lane 5). Even more remarkably, cleavage of substrate containing a −4 iG substitution occurs primarily at position +3, i.e. 6 nt from the site of analog substitution (Figure 6g), with reduced hydrolysis detected as far as position −3. Internal PPT cleavage of this substrate also occurs at position −6.

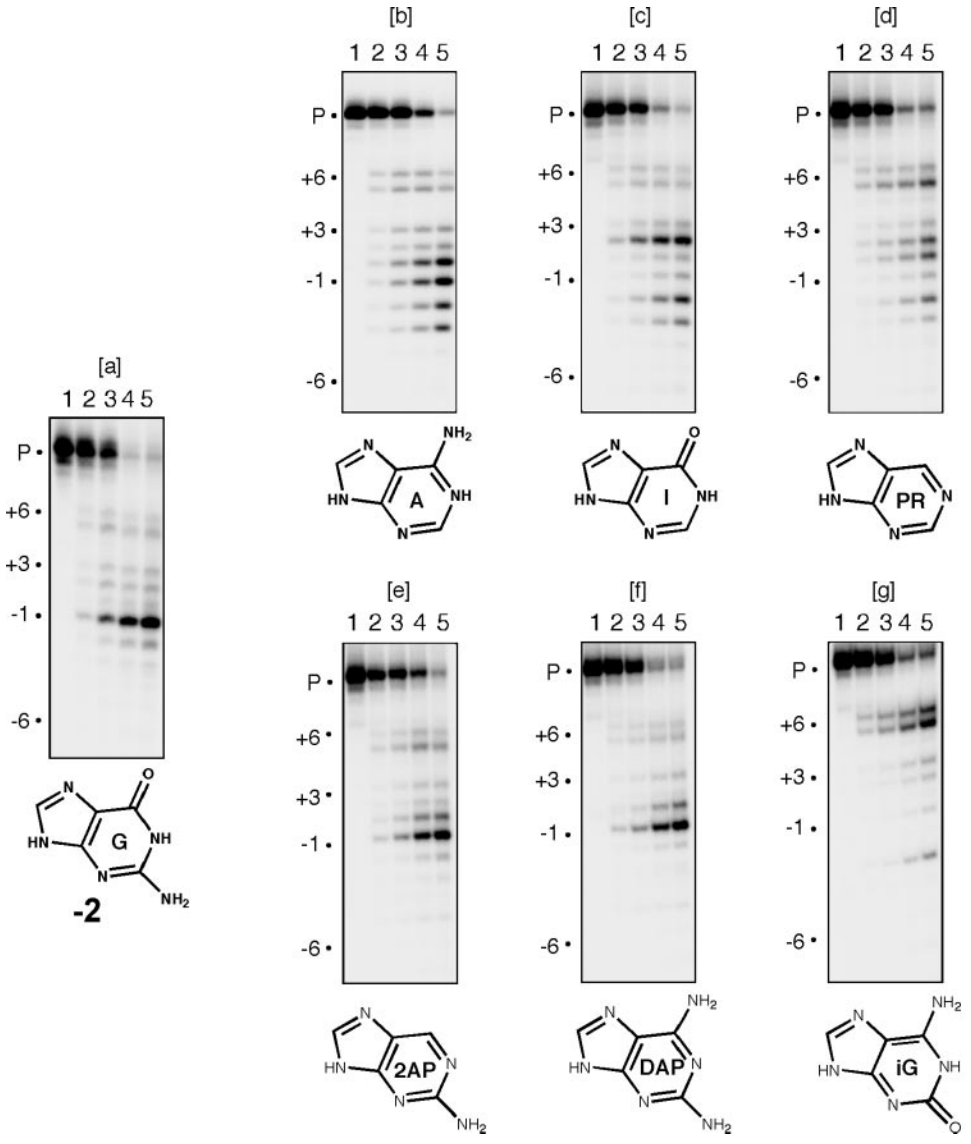


Figure 5. Effect of purine analog substitutions at −2rG of the HIV-1 PPT on cleavage specificity. Lane notations and product identification are as in the legend to Figure 4.

DISCUSSION

In the present study, we have (i) assessed specificity determinants for cleavage at the 3′ and 5′ termini of the (+)-strand primer, (ii) determined the relationships among PPT sequence, (+)-strand primer length and primer function, and (iii) examined the functions of rG:dC-, rA:dT- and rU:dA-tracts within and immediately upstream of the PPT. Each of the homopolymeric tracts is highly conserved among retroviral PPTs (7,21). In HIV, our data suggests the rU:dA tract helps define the 5′ terminus of the (+)-strand primer, thereby ensuring that it is the appropriate length for processing by RT (Figure 1a and b, *cf*. lanes W and 7). *In vivo* studies of simian immunodeficiency virus and Moloney murine leukemia virus have demonstrated this conserved sub-motif is critical for virus replication (7,14,35,36). However, although removing the rU:dA tract drastically reduced the levels of reverse transcription in these studies, it is not clear that this effect was directly related to PPT-processing or initiation of (+)-strand DNA synthesis. It is interesting to note, however, that after extensive serial passaging of the defective virus, a number of viral variants emerged in which the deleted rU:dA tract was replaced by an extended rA:dT tract. The result, invariably, was 5′ terminal extension of the PPT. For example, in one remarkable instance in SIV, poly(A) insertion resulted in a 36 nt PPT (7). These findings support conclusions derived from the current work, namely, (i) that rU:dA and rA:dT tracts adjacent to and within the PPT, respectively, are to some extent functionally interchangeable (Figure 2a and b, lanes 3) and (ii) that a minimum (+)-strand primer length of 15–17 nt is required for optimal PPT function. Finally, it must be noted that our results contrast with a previous biochemical analysis indicating that the rU:dA tract does not play a role in PPT primer selection or initiation of (+) strand DNA synthesis (32). This apparent discrepancy

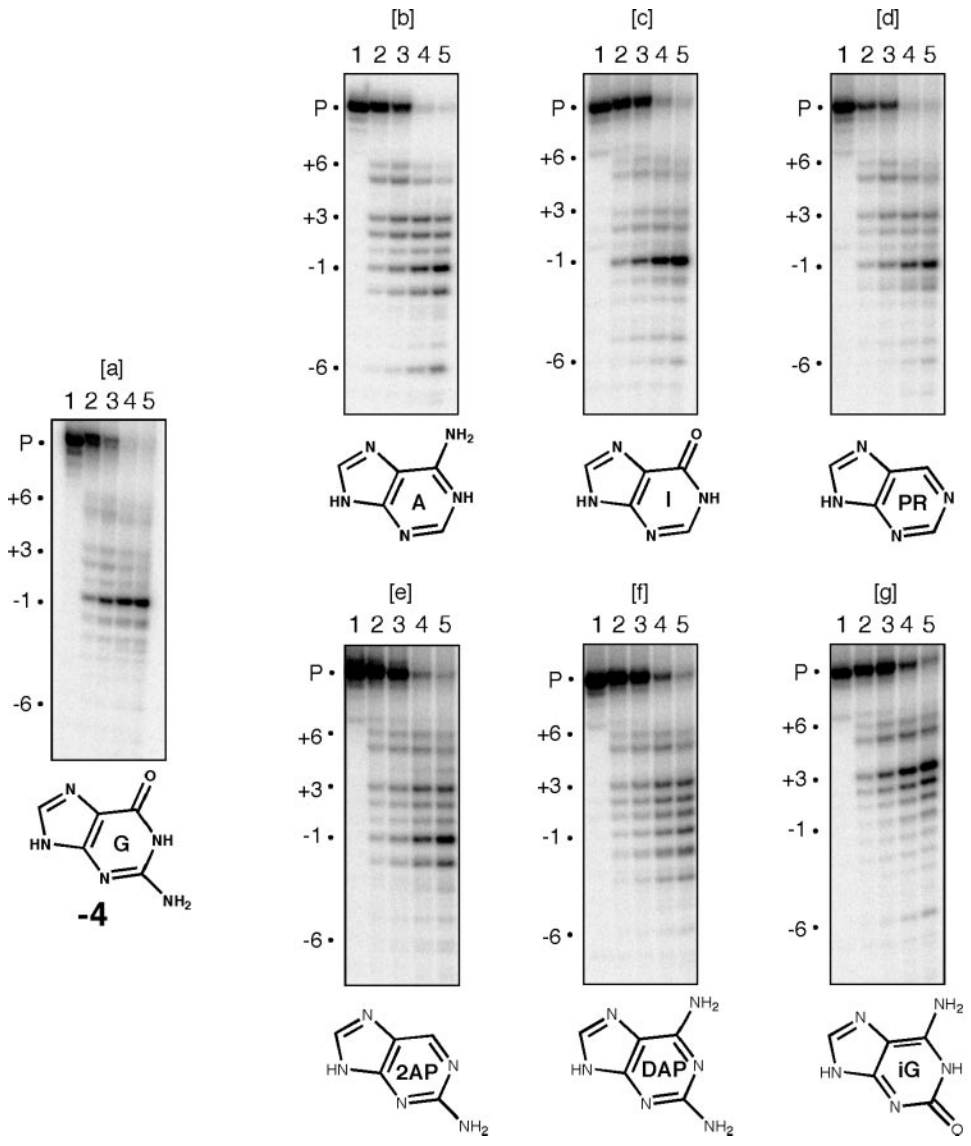


Figure 6. Effect of purine analog substitutions at −4rG of the HIV-1 PPT on cleavage specificity. Lane notations and product identification are as in the legend to Figure 4.

may be due to differences in experimental procedure, most notably, the higher stringency of our assay with respect to enzyme and salt concentration.

The interrupted rA:dT tract constituting the upstream portion of the HIV-1 PPT does not appear to contain determinants for cleavage at the 5′ or 3′ terminus of the (+)-strand primer (Figure 1b and c, lane 6; Figure 2). Instead, the role of this element is to add length to the primer while minimizing internal cleavage (Figure 2). Of the three sub-motifs, the rA:dT tract sequence is least conserved among retroviruses, and is essentially absent in bovine leukemia virus and avian myeloblastoma virus (7). Furthermore, the length of this element is highly variable, ranging from as few as 4 to as many as 17 nt, and it is usually discontinuous, being interrupted once or twice by one or more Gs or Cs. Despite this, interrupted rA:dT tracts usually contain a stretch of at least four consecutive rA:dTs, suggesting minor groove narrowing, helical rigidity and other structural features associated with homopolymeric A/T tracts may render the PPT resistant to RNase H. Altered inter-strand base pairing, noted to occur near the center of the interrupted rA:dT tract of the HIV-1 PPT (2,37), may also play a role in rA:dT tract function. However, this sub-motif has proven relatively insensitive to mutation *in vivo* (38) as well as *in vitro* (29,32), and in the current study, complete disruption of the rA:dT tract via multiple A→G mutations had little effect on (+)-strand initiation (Figure 1b and c, lane 4). Amino acid coding requirements may also play a role in determining rA:dT tract composition. The HIV-1 3′ PPT, for example, is embedded within the nef open reading frame, while the central PPT lies within the coding sequence for integrase (39). Finally, accessory proteins, such as NC may alter the sequence requirements of PPT selection/utilization *in vivo* in a manner not predicted by our simple assays.

The results of Figure 2 suggest that the optimum length of the plus-strand primer is ∼17 nt, the length of which is

roughly equivalent to the distance between the DNA polymerase and RNase H active sites of HIV-1 RT (40,41). This suggests that positioning of RT over the hybrid duplex may play an important role in hydrolysis at the PPT-U3 junction. Normally, HIV-1 RT binds a hybrid duplex containing two recessed RNA termini such that the polymerase active site is located over the DNA strand opposite the RNA 5′ terminus (42). In the case of the (+)-strand primer/DNA hybrid, this would place the RNase H domain over the RNA strand near the PPT-U3 junction. This preferred mode of binding may explain why G-tract mutations (i.e. rG→rA) are more likely to result in internal PPT cleavage than the inverse substitutions (rA→rG) within the rA:dT tract, since the RNase H domain of HIV-1 RT (located at or near PPT position −1) would be expected to be in much closer proximity to the former sub-motif. It also appears, however, that in contrast to an RNA/DNA hybrid of random sequence, positioning of RT on a PPT/DNA hybrid is largely independent of the location of the primer 5′ terminus, since (+)-strand initiation is unaffected by the 5′-AAAAG-3′ insertion in substrate PS-9- (Figure 2a and b, lanes W and 2). This also argues that stepwise digestion of the substrate from the RNA 5′ terminus (43,44) may not be essential for PPT selection. However, it is clear that shortening the primer adversely affects both (+)-strand initiation and primer removal (Figure 2), most likely due to reduced affinity of RT for the smaller hybrid. Dissociation of the hybrid is probably not an important factor, since the T_m of even the shortest primer evaluated (substrate PS-13; Figure 2a and b, lane 6) would be expected to exceed 42°C (45).

Finally, our results indicate that in HIV-1, determinants for cleavage at, and initiation from the PPT-U3 junction reside within and immediately adjacent to the rG:dC tract, even when this is displaced to other locations within the PPT (Figure 1b, lanes 1 and 2). Furthermore, in agreement with the recent findings of Schultz *et al*. (34), −4rG, −2rG and +1rA were found to be of particular significance in defining cleavage specificity. Using nucleoside analogs, we were able to modulate the chemical composition of purines at these positions to determine which exocyclic moieties were critical in this regard. The presence of a 2-NH_2 group at position −2 and the absence of this group at position +1, proved essential for directing cleavage at the PPT-U3 junction. In the former case, the 2-NH_2 group also plays a role in preventing internal cleavage of the PPT at or near position −2. Intriguingly, only two residues directly contact RNA bases in the vicinity of the RNase H domain in the HIV-1 RT-PPT/DNA co-crystal structure, namely Q475 and R448 of the RNase H primer grip at positions −2 and +1, respectively, relative to the scissile phosphate identified in that structure (2). Furthermore, contacts occur at or near position 2 of the purines in these locations. This raises the possibility that steric interference and/or disruption of specific hydrogen bonds at positions −2 and +1 may directly influence whether or not cleavage occurs at the PPT-U3 junction. For example, a +1rA→rG substitution would be expected to create a steric clash with residue R448 of HIV-1 RT by introducing a relatively bulky amino group at position 2 of the the purine. In contrast, a −2rA→rI substitution, which would not change the constituent at the purine 2-position, would have no apparent effect. A model depicting such a scenario is shown in Figure 7. Alternatively, the importance of −2rG and +1rA may stem from unusual base pairing at these positions (20,37), wherein the 2-NH_2-substituted purine may disrupt important *cis*-acting signals directing RT-binding and cleavage, and/or enhanced helical flexibility noted to occur within 5′-rArGrA-3′ steps (31).

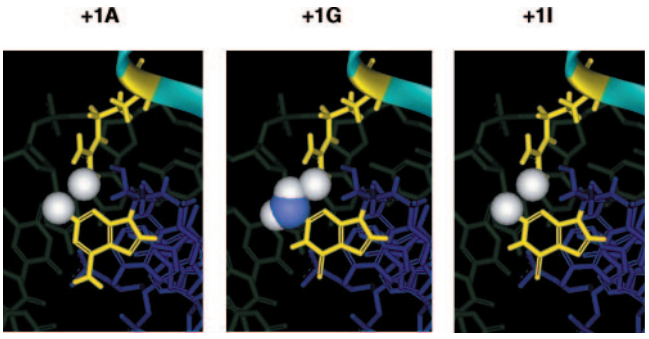


Figure 7. Steric interference upon rA→rG substitution at position +1. Panels reflect changes to the RT–nucleic acid interface when A, G or I is present 1 nt 3′ of the scissile phosphate. R448 of the RNase H primer grip and the base at position +1 are shown in yellow. Space-filled atoms include the 2-H and 2-NH_2 groups of A/I and G, respectively, as well as side-chain amine hydrogen of R448. Models are derived from the HIV-1 RT-PPT/DNA crystal structure (2), and nucleoside variants constructed using DS Viewer (Accelyrs, San Diego, CA).

Although there appears to be no single structural feature to explain the effects of each analog substitution at position −4rG (Figure 6), it is clear that a 6-NH_2-substituted purine is poorly tolerated at that position, as is the reversal in exocyclic group orientation resulting from rG→riG substitution. Since no direct contact between HIV-1 RT and the RNA strand of an RNA–DNA hybrid is predicted to occur at PPT position −4 (relative to the scissile phosphate), local steric interference with RT-binding is an unlikely explanation for our observations. However, it is possible that curvature of the helix is affected by substitution of an amino group for the 6-oxygen moiety of −4rG, which could indirectly affect the manner in which the duplex is contacted by HIV-1 RT. This might manifest itself via electronic effects on base stacking (46,47), altered coordination of solvent cations (48), or perhaps most intriguingly, by direct electronic (or steric) interaction between/among exocyclic groups on adjacent purines. For example, within the interrupted A-tract of the HIV-1 PPT the interatomic distances between 6-NH_2 nitrogens of adjacent purines are as little as 2.9 Å, suggesting the possibility of direct electronic interaction between bases (49). One could envision, therefore, that inserting an amino group into a string of adjacent 6-oxygen moieties within a G-tract could alter the trajectory of the helix. This is particularly plausible since elements bonded at the purine 6-position comprise one wall of the helical major groove, which like the minor groove in dA:dT and rA:dT tracts, is compressed in homopolymeric dG:dC stretches (48). However, since a homopolymeric rG:dC tract has never been resolved by X-ray crystallography or NMR, the spatial relationship between 6-oxygen groups on adjacent purines in this context remains speculative.

ACKNOWLEDGEMENTS

This work was supported by the Intramural Research Program of the NIH, NCI and Center for Cancer Research. Funding to pay the Open Access publication charges for this article was provided by the Intramural Research Program of the NIH, NCI and Center for Cancer Research.

Conflict of interest statement. None declared.

REFERENCES

1. Champoux,J.J. (1993) Roles of ribonuclease H in reverse transcription. In Skalka,A.M and Goff,S.P. (eds), *Reverse Transcriptase*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 103–118.
2. Sarafianos,S.G., Das,K., Tantillo,C., Clark,A.D.,Jr, Ding,J., Whitcomb,J.M., Boyer,P.L., Hughes,S.H. and Arnold,E. (2001) Crystal structure of HIV-1 reverse transcriptase in complex with a polypurine tract RNA:DNA. *EMBO J*., **20**, 1449–1461.
3. Rattray,A.J. and Champoux,J.J. (1989) Plus-strand priming by Moloney murine leukemia virus. The sequence features important for cleavage by RNase H. *J. Mol. Biol*., **208**, 445–456.
4. Pullen,K.A., Rattray,A.J. and Champoux,J.J. (1993) The sequence features important for plus strand priming by human immunodeficiency virus type 1 reverse transcriptase. *J. Biol. Chem*., **268**, 6221–6227.
5. Powell,M.D. and 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.
6. Rausch,J.W. and Le Grice,S.F. (1997) Substituting a conserved residue of the ribonuclease H domain alters substrate hydrolysis by retroviral reverse transcriptase. *J. Biol. Chem*., **272**, 8602–8610.
7. Ilyinskii,P.O. and 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.
8. Gotte,M., Maier,G., Onori,A.M., Cellai,L., Wainberg,M.A. and Heumann,H. (1999) Temporal coordination between initiation of HIV (+)-strand DNA synthesis and primer removal. *J. Biol. Chem*., **274**, 11159–11169.
9. Schultz,S.J., Zhang,M., Kelleher,C.D. and Champoux,J.J. (1999) Polypurine tract primer generation and utilization by Moloney murine leukemia virus reverse transcriptase. *J. Biol. Chem*., **274**, 34547–34555.
10. Schultz,S.J., Zhang,M., Kelleher,C.D. and Champoux,J.J. (2000) Analysis of plus-strand primer selection, removal, and reutilization by retroviral reverse transcriptases. *J. Biol. Chem*., **275**, 32299–32309.
11. Lim,D., Orlova,M. and Goff,S.P. (2002) Mutations of the RNase H C helix of the Moloney murine leukemia virus reverse transcriptase reveal defects in polypurine tract recognition. *J. Virol*., **76**, 8360–8373.
12. Lener,D., Kvaratskhelia,M. and Le Grice,S.F. (2003) Nonpolar thymine isosteres in the Ty3 polypurine tract DNA template modulate processing and provide a model for its recognition by Ty3 reverse transcriptase. *J. Biol. Chem*., **278**, 26526–26532.
13. Schultz,S.J., Zhang,M. and Champoux,J.J. (2003) Specific cleavages by RNase H facilitate initiation of plus-strand RNA synthesis by Moloney murine leukemia virus. *J. Virol*., **77**, 5275–5285.
14. Robson,N.D. and Telesnitsky,A. (2000) Selection of optimal polypurine tract region sequences during Moloney murine leukemia virus replication. *J. Virol*., **74**, 10293–10303.
15. Julias,J.G., McWilliams,M.J., Sarafianos,S.G., Alvord,W.G., Arnold,E. and Hughes,S.H. (2004) Effects of mutations in the G tract of the human immunodeficiency virus type 1 polypurine tract on virus replication and RNase H cleavage. *J. Virol*., **78**, 13315–13324.
16. Miles,L.R., Agresta,B.E., Khan,M.B., Tang,S., Levin,J.G. and Powell,M.D. (2005) Effect of polypurine tract (PPT) mutations on human immunodeficiency virus type 1 replication: a virus with a completely randomized PPT retains low infectivity. *J. Virol*., **79**, 6859–6867.
17. Rausch,J.W., Qu,J., Yi-Brunozzi,H.Y., Kool,E.T. and Le Grice,S.F. (2003) Hydrolysis of RNA/DNA hybrids containing nonpolar pyrimidine isosteres defines regions essential for HIV type 1 polypurine tract selection. *Proc. Natl Acad. Sci. USA*, **100**, 11279–11284.
18. Yi-Brunozzi,H.Y. and Le Grice,S.F. (2005) Investigating HIV-1 polypurine tract geometry via targeted insertion of abasic lesions in the (−)-DNA template and (+)-RNA primer. *J. Biol. Chem*., **280**, 20154–20162.
19. Dash,C., Marino,J.P. and Le Grice,S.F. (2006) Examining Ty3 polypurine tract structure and function by nucleoside analog interference. *J. Biol. Chem*., **281**, 2773–2783.
20. Dash,C., Rausch,J.W. and Le Grice,S.F. (2004) Using pyrrolo-deoxycytosine to probe RNA/DNA hybrids containing the human immunodeficiency virus type-1 3′ polypurine tract. *Nucleic Acids Res*., **32**, 1539–1547.
21. Rausch,J.W. and Le Grice,S.F. (2004) 'Binding, bending and bonding': polypurine tract-primed initiation of plus-strand DNA synthesis in human immunodeficiency virus. *Int. J. Biochem. Cell. Biol*., **36**, 1752–1766.
22. Koo,H.S., Drak,J., Rice,J.A. and Crothers,D.M. (1990) Determination of the extent of DNA bending by an adenine-thymine tract. *Biochemistry*, **29**, 4227–4234.
23. McCarthy,J.G. and Rich,A. (1991) Detection of an unusual distortion in A-tract DNA using KMnO4: effect of temperature and distamycin on the altered conformation. *Nucleic Acids Res*., **19**, 3421–3429.
24. Strahs,D. and Schlick,T. (2000) A-Tract bending: insights into experimental structures by computational models. *J. Mol. Biol*., **301**, 643–663.
25. Young,M.A. and Beveridge,D.L. (1998) Molecular dynamics simulations of an oligonucleotide duplex with adenine tracts phased by a full helix turn. *J. Mol. Biol*., **281**, 675–687.
26. Stefl,R., Wu,H., Ravindranathan,S., Sklenar,V. and Feigon,J. (2004) DNA A-tract bending in three dimensions: solving the dA4T4 vs. dT4A4. conundrum. *Proc. Natl Acad. Sci. USA*, **101**, 1177–1182.
27. Mollegaard,N.E., Lindemose,S. and Nielsen,P.E. (2005) Uranyl photoprobing of nonbent A/T- and bent A-tracts. A difference of flexibility? *Biochemistry*, **44**, 7855–7863.
28. Nussinov,R. (1991) Signals in DNA sequences and their potential properties. *Comput. Appl. Biosci*., **7**, 295–299.
29. Pullen,K.A., Rattray,A.J. and Champoux,J.J. (1993) The sequence features important for plus strand priming by human immunodeficiency virus type 1 reverse transcriptase. *J. Biol. Chem*., **268**, 6221–6227.
30. Rausch,J.W., Lener,D., Miller,J.T., Julias,J.G., Hughes,S.H. and Le Grice,S.F. (2002) Altering the RNase H primer grip of human immunodeficiency virus reverse transcriptase modifies cleavage specificity. *Biochemistry*, **41**, 4856–4865.
31. Kopka,M.L., Lavelle,L., Han,G.W., Ng,H.L. and Dickerson,R.E. (2003) An unusual sugar conformation in the structure of an RNA/DNA decamer of the polypurine tract may affect recognition by RNase H. *J. Mol. Biol*., **334**, 653–665.
32. Powell,M.D. and 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.
33. Huber,H.E. and Richardson,C.C. (1990) Processing of the primer for plus strand DNA synthesis by human immunodeficiency virus 1 reverse transcriptase. *J. Biol. Chem*., **265**, 10565–10573.
34. Schultz,S.J., Zhang,M. and Champoux,J.J. (2006) Sequence, distance, and accessibility are determinants of 5′ end-directed cleavages by retroviral RNases H. *J. Biol. Chem*., **281**, 1943–1955.
35. Bacharach,E., Gonsky,J., Lim,D. and Goff,S.P. (2000) Deletion of a short, untranslated region adjacent to the polypurine tract in Moloney murine leukemia virus leads to formation of aberrant 5′ plus-strand DNA ends *in vivo*. *J. Virol*., **74**, 4755–4764.
36. Robson,N.D. and Telesnitsky,A. (1999) Effects of 3′ untranslated region mutations on plus-strand priming during moloney murine leukemia virus replication. *J. Virol*., **73**, 948–957.
37. Kvaratskhelia,M., Budihas,S.R. and Le Grice,S.F. (2002) Pre-existing distortions in nucleic acid structure aid polypurine tract selection by HIV-1 reverse transcriptase. *J. Biol. Chem*., **277**, 16689–16696.
38. McWilliams,M.J., Julias,J.G., Sarafianos,S.G., Alvord,W.G., Arnold,E. and Hughes,S.H. (2003) Mutations in the 5′ end of the human immunodeficiency virus type 1 polypurine tract affect RNase H cleavage specificity and virus titer. *J. Virol*., **77**, 11150–11157.

39. Ratner,L., Fisher,A., Jagodzinski,L.L., Mitsuya,H., Liou,R.S., Gallo,R.C. and Wong-Staal,F. (1987) Complete nucleotide sequences of functional clones of the AIDS virus. *AIDS Res. Hum. Retroviruses*, **3**, 57–69.
40. Jacobo-Molina,A., Clark,A.D.,Jr, Williams,R.L., Nanni,R.G., Clark,P., Ferris,A.L., Hughes,S.H. and Arnold,E. (1991) Crystals of a ternary complex of human immunodeficiency virus type 1 reverse transcriptase with a monoclonal antibody Fab fragment and double-stranded DNA diffract X-rays to 3.5 Å resolution. *Proc. Natl Acad. Sci. USA*, **88**, 10895–10899.
41. Arnold,E., Jacobo-Molina,A., Nanni,R.G., Williams,R.L., Lu,X., Ding,J., Clark,A.D.,Jr, Zhang,A., Ferris,A.L., Clark,P. *et al.* (1992) Structure of HIV-1 reverse transcriptase/DNA complex at 7 Å resolution showing active site locations. *Nature*, **357**, 85–89.
42. Palaniappan,C., Fuentes,G.M., Rodriguez-Rodriguez,L., Fay,P.J. and Bambara,R.A. (1996) Helix structure and ends of RNA/DNA hybrids direct the cleavage specificity of HIV-1 reverse transcriptase RNase H. *J. Biol. Chem.*, **271**, 2063–2070.
43. Wisniewski,M., Balakrishnan,M., Palaniappan,C., Fay,P.J. and Bambara,R.A. (2000) Unique progressive cleavage mechanism of HIV reverse transcriptase RNase H. *Proc. Natl Acad. Sci. USA*, **97**, 11978–11983.
44. Wisniewski,M., Balakrishnan,M., Palaniappan,C., Fay,P.J. and Bambara,R.A. (2000) The sequential mechanism of HIV reverse transcriptase RNase H. *J. Biol. Chem.*, **275**, 37664–37671.
45. Sugimoto,N., Nakano,S., Katoh,M., Matsumura,A., Nakamuta,H., Ohmichi,T., Yoneyama,M. and Sasaki,M. (1995) Thermodynamic parameters to predict stability of RNA/DNA hybrid duplexes. *Biochemistry*, **34**, 11211–11216.
46. Zhang,Q. and Chen,E.C. (1995) The experimental hardness and electronegativity of the purines and pyrimidines in DNA and RNA supported by the AM1 calculation of the electron affinities and ionization potentials. *Biochem. Biophys. Res. Commun.*, **217**, 755–760.
47. Mignon,P., Loverix,S., Steyaert,J. and Geerlings,P. (2005) Influence of the pi-pi interaction on the hydrogen bonding capacity of stacked DNA/RNA bases. *Nucleic Acids Res.*, **33**, 1779–1789.
48. Hud,N.V. and Plavec,J. (2003) A unified model for the origin of DNA sequence-directed curvature. *Biopolymers*, **69**, 144–158.
49. Vinogradov,S.N. (1979) Hydrogen bonds in crystal structures of amino acids, peptides and related molecules. *Int. J. Pept. Protein. Res.*, **14**, 281–289.