Chemistry & Biology
Article

High-Resolution NMR Analysis of the Conformations of Native and Base Analog Substituted Retroviral and LTR-Retrotransposon PPT Primers

Hye Young Yi-Brunozzi,[1] Robert G. Brinson,[1,2] Danielle M. Brabazon,[3] Daniela Lener,[1] Stuart F.J. Le Grice,[1,*] and John P. Marino[2,*]
[1]HIV Drug Resistance Program, NCI-Frederick National Institutes of Health, Frederick, MD, 21702
[2]Center for Advanced Research in Biotechnology of the University of Maryland Biotechnology Institute and the National Institute of Standards and Technology, Rockville, MD, 20850
[3]Department of Chemistry, Loyola College in Maryland, Baltimore, MD, 21210
*Correspondence: marino@umbi.umd.edu (J.P.M.), slegrice@ncifcrf.gov (S.F.J.L.G.)
DOI 10.1016/j.chembiol.2008.01.012

SUMMARY

A purine-rich region of the plus-strand RNA genome of retroviruses and long terminal repeat (LTR)-containing retrotransposons, known as the polypurine tract (PPT), is resistant to hydrolysis by the RNase H domain of reverse transcriptase (RT) and ultimately serves as a primer for plus-strand DNA synthesis. The mechanisms underlying PPT resistance and selective processing remain largely unknown. Here, two RNA/DNA hybrids derived from the PPTs of HIV-1 and Ty3 were probed using high-resolution NMR for preexisting structural distortions in the absence of RT. The PPTs were selectively modified through base-pair changes or by incorporation of the thymine isostere, 2,4-difluoro-5-methylbenzene (dF), into the DNA strand. Although both wild-type (WT) and mutated hybrids adopted global A-form-like helical geometries, observed structural perturbations in the base-pair and dF-modified hybrids suggested that the PPT hybrids may function as structurally coupled domains.

INTRODUCTION

Replication of retroviruses and LTR-containing such retrotransposons as HIV-1 and *Saccharomyces cerevisiae* Ty3, respectively, begins with minus-strand DNA synthesis from a host tRNA hybridized to the primer-binding site located near the 5′ end of the viral genome. As synthesis continues through minus-strand strong stop and strand transfer events, an RNA/DNA duplex is created that is subject to hydrolysis via the RNase H activity of the virally encoded RT. This occurs nonspecifically, with the exception of two highly conserved purine-rich segments of RNA referred to as the 5′ and central PPTs. These elements remain hybridized to nascent minus-strand DNA and serve to prime plus-strand DNA synthesis. The precise generation and removal of plus-strand primers is crucial for viral replication (Champoux, 1993; Telesnitsky and Goff, 1997).

Despite differences in sequence between the HIV-1 and Ty3 PPTs, both are accurately processed by their cognate RT. HIV-1 RT is a p66/p51 heterodimer with DNA polymerase and RNase H catalytic centers spaced a distance equivalent to about 17 bp apart, whereas the HIV-1 PPT sequence is 15 bp (Huang et al., 1998; Jacobo-Molina et al., 1993; Sarafianos et al., 2001). The *Saccharomyces cerevisiae* retrotransposon Ty3 RT is a 55 kDa monomer, for which a high resolution crystal structure is unavailable, with an identified PPT sequence that totals 12 bp (Lener et al., 2002; Rausch et al., 2000). Although no high-resolution structures are currently available for either the full-length HIV-1 or Ty3 RNA/DNA PPT hybrids, the crystal structure of HIV-1 RT, in complex with a PPT-containing RNA/DNA hybrid, has been solved at 3 Å resolution (Sarafianos et al., 2001). As with many sequence-independent nucleic acid-binding enzymes, the cocrystal structure showed contacts with the nucleic acid that were primarily restricted to the sugar-phosphate backbone, with very few base-specific interactions. One important and unexpected finding from the cocrystal structure was the observation of unpaired and mispaired bases, as well as weakened base pairing. On the basis of these observations, it was postulated that recognition and binding to these uniquely distorted structural elements in the PPT by the thumb subdomain could serve to position the RNase H catalytic center over the PPT/U3 junction for specific cleavage by RT. The PPT sequences of HIV-1 and Ty3, with regions proposed to be involved in contacts with the thumb and RNase H domains of RT highlighted, are shown in Figure 1.

Further studies that have probed the conformation of the HIV-1 PPT in solution have reported that structural anomalies in the PPT, which map to the sequence regions of the PPT that are predicted to interact with RT, may preexist in the absence of RT (Figure 1). In one study, $KMnO_4$ chemical probing analysis of the HIV-1 PPT in the absence of RT showed an increased solvent accessibility in the "unzipped" region between positions −10 and −15 (defining position −1 as the first base pair upstream of the PPT/U3 junction), corresponding to the p66 thumb binding site and the +1 position at the PPT/U3 junction (Kvaratskhelia et al., 2002). In a second study, the fluorescent cytosine analog, pyrrolo-deoxycytosine, was substituted in several positions in the PPT as a probe of local base stacking. The results from that study provided further evidence that the

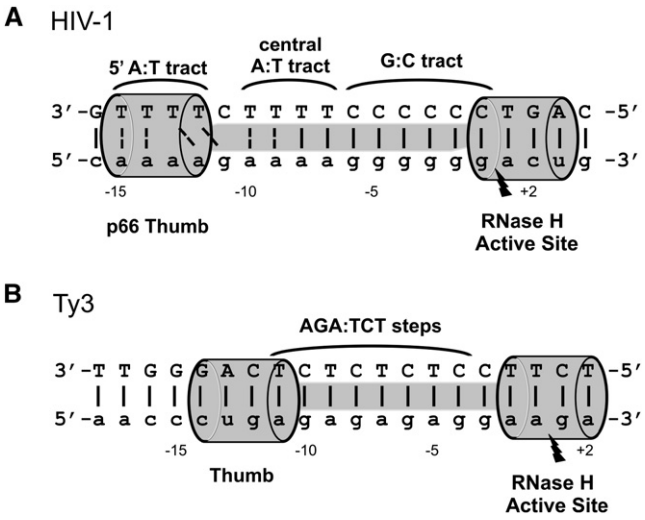

Figure 1. HIV-1 and Ty3 PPT Sequences

Schematic representation of HIV-1 (A) and Ty3 (B) PPT sequences with regions of contact with cognate RTs highlighted. DNA is shown in uppercase, and RNA is shown in lowercase. Weakly paired and unpaired bases are shown on the basis of inferences from the crystal structure of the HIV-1 RT bound PPT sequence (Sarafianos et al., 2001). The RNase H cleavage position is indicated on the RNA strand in each duplex by arrows. The numbering of bases is indicated 5′ to 3′ for each strand, and corresponding positions relative to the RNase H cleavage sites are indicated below the RNA strand. The $(rG)_6$:$(dC)_6$ tract and two $(rA)_4$:$(dT)_4$ tracts in the HIV-1 sequence and AGA:TCT steps in the Ty3 sequence are indicated and labeled. Graphical representations of the binding sites for RT are indicated by the shaded cylinders for RT thumbs and RNase H active sites.

cytosine base at position −11 is not well stacked in the hybrid and is possibly unpaired in the absence of RT (Dash et al., 2004).

Two previous structural studies have been performed using model HIV-1 PPT sequences. In the first of these studies, Fedoroff et al., (1997) used solution state NMR spectroscopy to examine an octanucleotide hybrid duplex flanking the PPT/U3 junction, showing widening of minor groove width as the only anomalous feature that might distinguish the PPT RNA/DNA hybrid from random sequence hybrids to control the correct cleavage at the junction. Kopka et al., (2003) subsequently used an octanucleotide hybrid sequence derived from the 5′ end of the HIV-1 PPT to examine the transition between the two $(rA)_4$:$(dT)_4$ tracts by X-ray crystallography. Interestingly, a ribose sugar pucker switch was observed in the crystal structure of this hybrid duplex at position −15 in the sequence, along with changes in base stacking. However, the hybrid duplex in this crystal structure did not display any unzipping or narrowing of minor groove, as observed in the RT-hybrid duplex complex. A limitation of both of these studies was that only a portion of the PPT sequence was used and that sequence alterations were introduced within the PPT sequence, either to facilitate crystallization or to diminish overlap of NMR signals that results from the homopolymeric nature of the PPT sequence.

As a first step toward probing full-length PPT RNA/DNA hybrid duplexes for preexisting structural features that may confer resistance to the RNase H activity of RT within the PPT while conveying specific hydrolysis at the PPT/U3 cleavage junction, we recently demonstrated that high-resolution NMR techniques could be used to analyze a 20 mer RNA/DNA hybrid duplex, representing the full WT Ty3 PPT sequence (Yi-Brunozzi et al., 2005). The NMR data acquired for this duplex indicated that regular Watson-Crick base pairing and an A-form-like helical conformation were adopted by the Ty3 PPT. However, a structural anomaly was detected in the form of a sugar pucker switch at +1 position of the RNA strand from C3′-endo to mixed C3′/C2′-endo. In the present study, we have extended our investigation of the Ty3 PPT and its HIV-1 PPT counterpart by employing single base-pair changes or by incorporating of the thymine (dT) isostere dF into the DNA strand to probe for subtle structural perturbations (Lener et al., 2003; Rausch et al., 2003). Using WT, base-pair-exchanged, and dF-substituted PPTs, we report a comparative analysis of the dynamic characteristics and structural motifs inherent to both the PPTs in the absence of RT that may underlie their common resistance to RNase H hydrolysis and selective processing to serve as primers for plus-strand DNA synthesis.

RESULTS

Watson-Crick Base-Pairing in WT Ty3 and HIV-1 PPT Hybrids

The WT Ty3 and HIV-1 sequences used in this study are shown in Figure 2 with positions in the PPT where a base pair has been changed or the thymine isostere dF has been substituted, indicated above the DNA strand. dF has been widely used to probe the role of hydrogen bonding and steric effects in oligonucleotides (Guckian et al., 1998; Kool and Sintim, 2006; Schweitzer and Kool, 1995) and has been found to be relatively nonperturbing to the helical structure when inserted into DNA duplexes (Li et al., 2007; Silverman and Kool, 2007; Xia et al., 2006). All 20 mer hybrid duplex samples were first visualized by native polyacrylamide gel electrophoresis analysis to confirm 1:1 stoichiometry of the two strands and completeness of hybridization (data not shown). To ascertain whether the HIV-1 PPT displays any pre-existing structural anomalies in the absence of RT, the WT 20 bp hybrid duplex was analyzed using the same standard set of NOESY and *J*-correlated NMR experiments previously used to characterize the Ty3 PPT. This previous analysis indicated that the Ty3 WT PPT displayed standard Watson-Crick base-pairing, with no obvious distortions in the helix immediately evident from the NMR data, with the exception of the mixed sugar pucker for the guanine residue at the +1 position of the RNA (Yi-Brunozzi et al., 2005). In contrast, previous biochemical data acquired for the HIV-1 PPT in the absence of RT has suggested a structural preorganization at the PPT/U3 cleavage junction and the region predicted to be contacted by the p66 thumb (Kvaratskhelia et al., 2002; Rausch et al., 2003; Yi-Brunozzi et al., 2005). In addition, incorporation of the fluorescent base analog, pyrrolo-deoxycytosine, into the HIV-1 PPT to probe base pairing suggested that C(−11) may be unpaired (Dash et al., 2004).

Consistent with the previous observations made for the Ty3 PPT and somewhat in contrast to the biochemical data, NMR analysis indicates standard Watson-Crick pairing in the HIV-1 PPT in the absence of RT, with no clear evidence for unpairing or mispairing of bases in the hybrid duplex. Despite overlap in the $(rG)_6$:$(dC)_6$ tract (Figure 1), base pairing and stacking in the

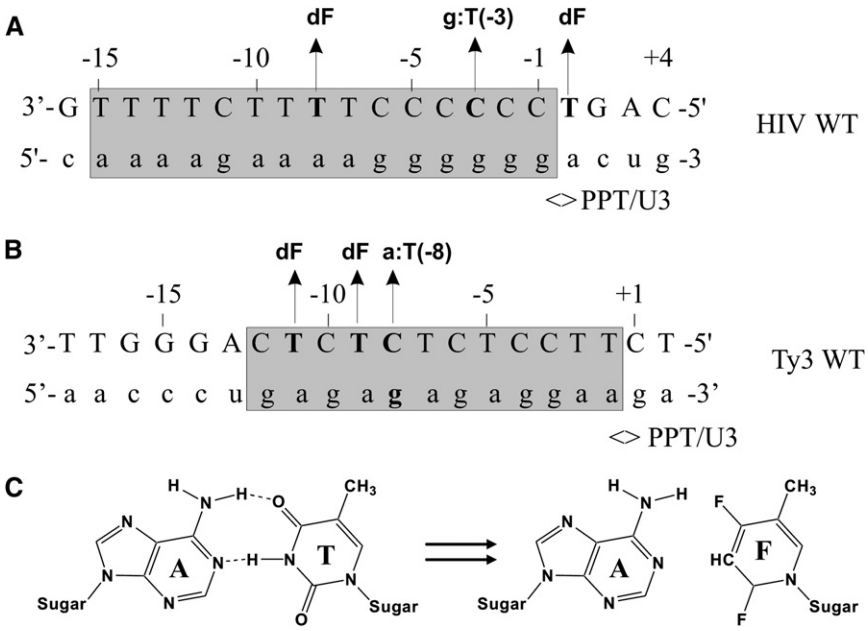

Figure 2. Model PPT Substrates

Model sequences used in this study for (A) the HIV-1 PPT-containing RNA/DNA hybrids and (B) the Ty3 PPT-containing RNA/DNA hybrids. (C) Schematic representation of 2,4-difluoro-5-methylbenzene. For both PPTs, the hybrid duplex is shown with the DNA strand in upper case and the RNA strand in lowercase. The consensus PPT sequences within the duplexes is boxed and shaded. The PPT/U3 junction is indicated on both WT duplexes, and positions relative to this junction are numbered. Single mutations made in each sequence are indicated with a vertical arrow. "dF" denotes the position at which dF was incorporated.

HIV-1 PPT could be established by assigning an imino NOE "walk" (see Figure S1 in the Supplemental Data available with this article online) through the entire duplex, with the exception of the terminal base pairs where imino protons on G(−16), u(+3), and g(+4) could not be observed because of fraying. Assignment of the imino resonances was facilitated using imino $^{1}H_{N}$-^{15}N correlations obtained from natural abundance ^{15}N-HMQC spectra. These data indicated that all imino proton resonances in the HIV-1 hybrid duplex are found within the chemical shift ranges that are consistent with canonical Watson-Crick base pairing (Figure 3, trace A). In particular, no imino protons were upfield shifted, as would be expected if a stable g(−11):T(−12) mismatch was found in the PPT in the absence of RT. To demonstrate the observed upfield shift in the imino proton chemical shift pattern that would be expected if such a mismatched base pair occurred, a g:T(−3) mispair was created in the HIV-1 PPT (Figure 3, trace B). Indeed, in this construct both the g(−3) and T(−3) imino protons were shifted upfield to 11.4 and 10.4 ppm, respectively, into chemical shift ranges indicative of a G-T mismatched base pair. In the WT HIV-1 sample, no imino resonance signals were observed in this region, further suggesting that the major conformation for the HIV-1 PPT observed on the NMR time scale is regularly base paired and stacked.

Resonance Assignment and Structural Probing Using dF and Base Pair Substitutions

dF or single base-pair substitutions were introduced at various positions to serve to disperse overlapping NMR resonances that result from the redundancy of the purine tracts. These dF substitutions served to facilitate assignment by shifting a subset of ^{1}H and ^{13}C resonances that were assigned to residues nearby the insertion site. These local chemical shift perturbations served to reveal overlapped signals that did not shift and could be related back to the WT spectra. In this way, a comparative analysis of multiple dF substituted PPT hybrid samples allowed common resonance assignment pathways to be determined. The approach of using dF to disperse and aid assignment of purine rich oligos will be described in more detail elsewhere (unpublished data). An initial analysis of 1D ^{1}H imino overlay for both Ty3 and HIV-1 PPTs showed the resolving power of single dF substitutions in the DNA strand. For the Ty3 PPT, single dF substitutions within d(C•T•C) repeats resulted in resolution of the thymines at and around 13.5 ppm (Figure 4). Interestingly, substitution of the base pair g:C(−8) for a:T(−8) had a similar dispersing effect on the thymine imino resonances. In an analogous manner, substituting dF for thymine in the DNA strand of the HIV-1 PPT likewise had a dispersing effect of the imino protons (Figure 3, traces C and D). The G imino protons in HIV dF(+1) were better resolved, whereas in HIV dF(−8), the thymine imino protons were dramatically dispersed.

To determine the effect of dF(−8) substitution on imino line widths, we calculated and compared the line widths for the imino proton signals from the HIV-1 WT and dF samples (Table S1). Broader imino line widths correlate with an increased rate of solvent exchange, which, in turn, may correlate with an increased lifetime in the open state (Snoussi and Leroy, 2001) or, as in the case of dF, increased solvent accessibility (Kool et al., 2000). The WT HIV-1 PPT had similar line widths of 23.5 Hz throughout the helix, with four notable exceptions: g(−1) (35.2 Hz), T(−7) (29.3 Hz), g(−11) (29.3 Hz), and T(−15) (29.3 Hz). All four broadened peaks represent a transition in the sequence between the $(rA)_4$:$(dT)_4$ tract and the $(rG)_6$:$(dC)_6$ tract. The line widths of HIV dF(−8) and HIV dF(+1) imino protons, 24.7 Hz and 29.3 Hz, respectively, were also fairly uniform throughout the helix, with a few notable exceptions. Broadening of T(−7) (30.8 Hz) and T(−9) (37.0 Hz) in HIV dF(−8) reflect greater solvent access, a local consequence of the dF substitution (Kool et al., 2000). There also remains in HIV dF(−8) a hint of the helical distortion around the junction of the two $(rA)_4$:$(dT)_4$ tracts, at T(−10) (30.8 Hz) and g(−11) (30.8 Hz). Additional resonances broadened in HIV dF(−8) were g(−2) (30.8 Hz) and T(−14)

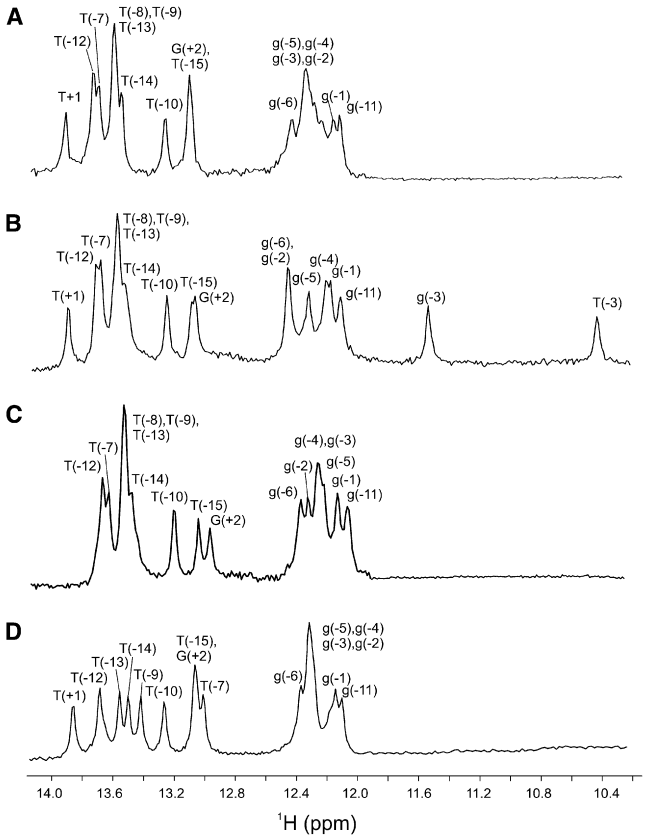

Figure 3. HIV-1 PPT Imino Proton Spectra

Comparison of the imino regions of the 1D water flip-back watergate ^{1}H NMR spectra of (A) HIV WT; (B) HIV C(−3)T; (C) HIV dF(+1); and (D) HIV dF(−8) PPT RNA/DNA hybrid duplexes at 10°C. Imino proton assignments are indicated on each spectrum. The proton carrier frequency was set to 4.69 ppm. The data were acquired on a Bruker DMX600 equipped with a TXI probe and with a sweep width of 12,000 Hz and 32 scans.

(30.8 Hz). In HIV dF(+1), there were only three notable line widths; resonances G(+2) (23.5 Hz), T(−7) (23.5 Hz), and T(−15) (23.5 Hz) exhibited line narrowing rather than line broadening.

Further resonance assignment and characterization of the PPTs was performed using standard analysis of 2D NMR experiments: NOESY, DQF-COSY, and natural abundance ^{1}H-^{13}C HMQC (Rance et al., 1983; Varani et al., 1996; Wijmenga and van Buuren, 1998). Using these methods, a nearly complete assignment could be made for both sets of duplexes for imino, H1′, H2′ (RNA), aromatic (H2/H6/H8), and H5 resonances (Tables S2 and S3). However, as a result of persistent spectral overlap and intermediate exchange processes on the NMR timescale for the HIV-1 PPT, C(−3) and C(−4) of the C tract could not be unambiguously assigned. Despite these exchange processes, both the Ty3 and HIV-1 duplexes showed every indication of adopting a stable hybrid duplex structure that was primarily A-form in helical geometry. In particular, cross-strand H1′-H2 NOEs were observed in the D_2O NOESY spectra, indicative of minor groove widths found in A-form helical geometry. In addition, cross-strand NOEs in ^{1}H-^{19}F NOESY spectra acquired for the dF-containing duplexes also indicate that the nucleoside analog is well stacked (Figures S2 and S3).

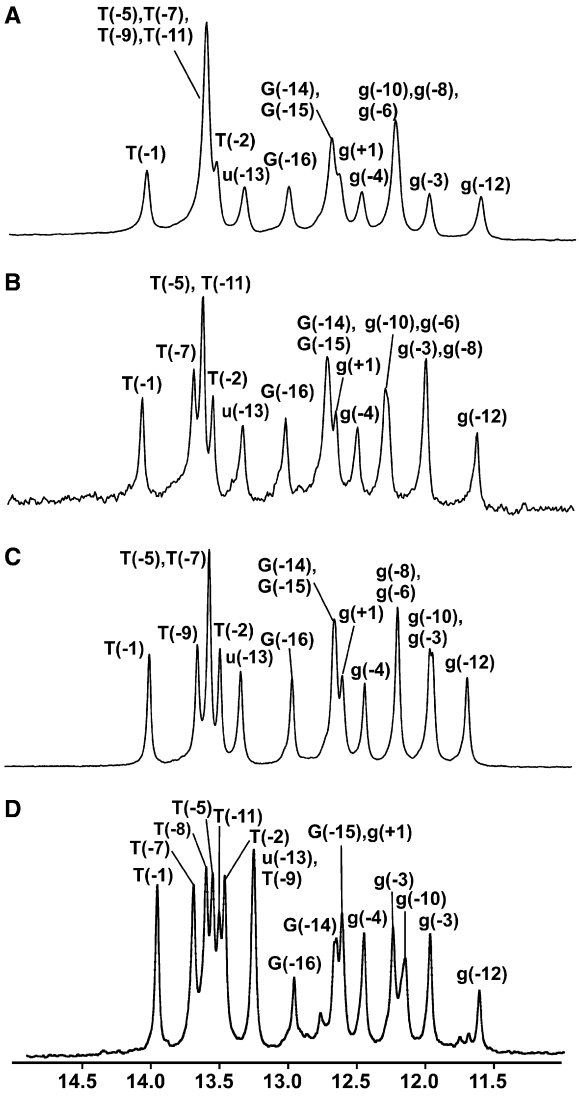

Figure 4. Ty3 PPT Imino Proton Spectra

Comparison of the imino regions of the 1D water flip-back watergate ^{1}H NMR spectra of (A) Ty3 WT; (B) Ty3 dF(−9); (C) Ty3 dF(−11); and (D) Ty3 C(−8)T Ty3 PPT RNA/DNA hybrid duplexes at 10°C. Imino proton assignments are indicated on each spectrum. The proton carrier frequency was set to 4.69 ppm. Data were acquired as indicated in Figure 3.

Long-Range Structural Coupling in the PPTs

Our previous work on Ty3 WT showed a ribose sugar pucker change at a(−1) and g(+1) from C3′-endo to mixed C3′/C2′-endo, as detected by observation of a sizable cross-peak in H1′-H2′ region of the DQF-COSY spectrum (Figure 5A; Yi-Brunozzi et al, 2005). Analysis of the same region in Ty3 dF(−11) and Ty3 dF(−9)-substituted PPTs showed additional cross-peaks in this same region of the DQF-COSY spectrum that are assigned to other RNA ribose sugars, which switch from predominately C3′-endo to mixed C3′/C2′-endo puckers (Figures 5B and 5C). An additional sample with a base pair substitution from g:C(−8) to a:T(−8) (designated Ty3 C(−8)T) was synthesized to confirm that this effect was not specific to samples with dF substitution (Figure 5D) and to break an r(a•g•a)

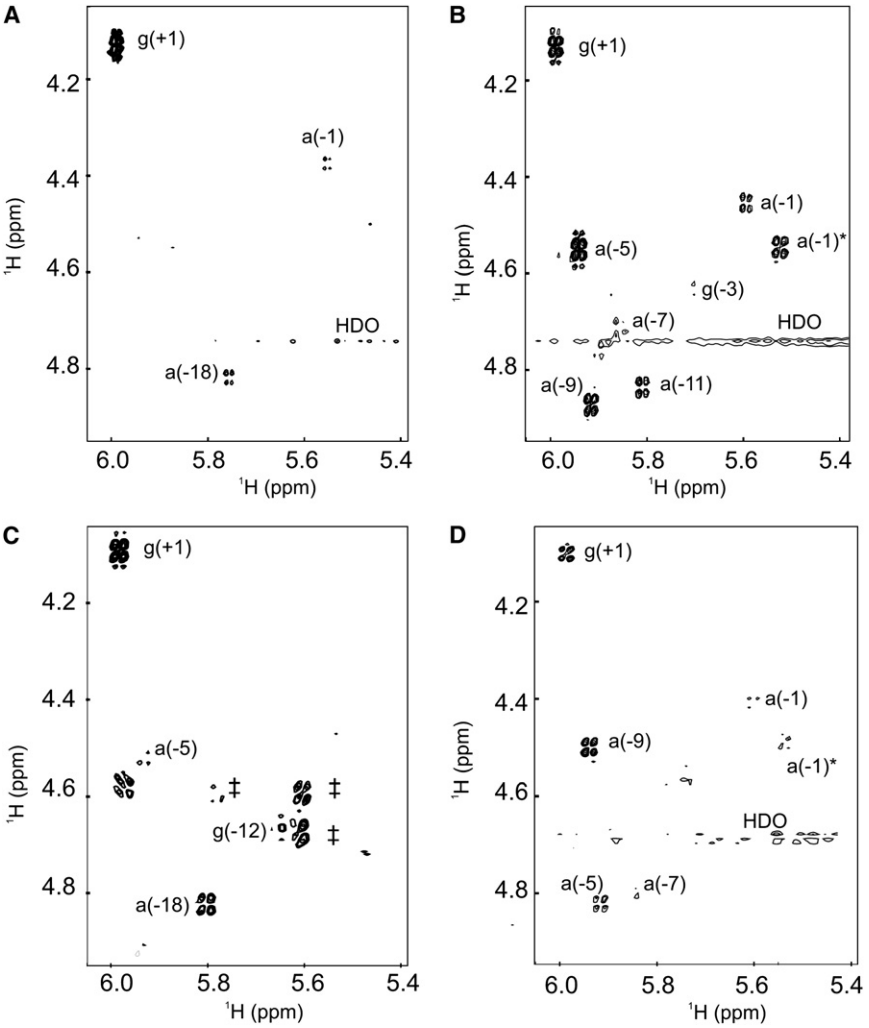

Figure 5. Ty3 PPT H1′-H2′ Ribose Sugar Correlations

Expansion of the H1′-H2′ ribose sugar correlated region of DQF-COSY experiments from the (A) Ty3 WT; (B) Ty3 dF(−11); (C) Ty3 dF(−9); and (D) Ty3 C(−8)T PPT RNA/DNA hybrid duplexes at 30°C. The proton carrier frequency was set to 4.70 ppm. The data were acquired with sweep widths of 5,000 Hz in both dimensions, 4096 by 800 complex points, and 48 scans per increment. ‡ = unassigned peaks.

symmetric step within the PPT. Although all Ty3 samples had a ribose sugar pucker switch at g(+1), a shift from a predominantly A-form WT helix to one with more B-form character is immediately apparent. Ty3 dF(−9) PPT mutant exhibited additional H1′-H2′ correlations at a(−5) and g(−12) as well as few unassigned peaks. Ty3 dF(−11) showed a more interesting periodic pattern. The RNA nucleotide directly across from Ty3 dF(−11) exhibited a ribose sugar pucker switch. The switches were observed to varying degrees for every other RNA nucleotide to the 3′ end—that is, a(−9), a(−7), a(−5), g(−3), and a(−1). For the base-substituted Ty3 C(−8)T sample, a periodic ribose sugar pucker switch, albeit weaker than that in dF(−11), was also observed. Interestingly, a(−9), 5′ of the substitution, gave a H1′-H2′ correlation. The propagation then proceeded to the 3′ end of the RNA strand, through a(−7), a(−5), and a(−1).

The HIV-1 ribose sugar pucker switches observed for WT and dF-substituted PPTs followed a slightly different pattern than for the Ty3 mutants (Figure 6). All samples exhibited a greater degree of fraying at the ends of the helix, particularly at g(+4), which is four bases downstream of the cleavage junction. Unlike the Ty3 samples, no ribose sugar pucker switch was observed at the PPT/U3 cleavage junction. However, when T(+1) was substituted with dF, short- and long-range structural effects were observed (Figure 6D). Base g(−2) was the only base in the $(rG)_6$:$(dC)_6$ tract to exhibit a ribose sugar pucker switch. Significantly, a slight ribose sugar pucker switch was observed at a(−10), a region that is known to exhibit structural anomalies (Conn et al., 1999; Sugimoto et al., 1995, 2000). HIV C(−3)T mutant more closely resembled the HIV WT and HIV dF(−8) duplexes, with ribose sugar puckering probably due to end effects (Figure 6B).

Thermal Denaturing Studies

Melting temperatures of 69.0°C and 69.1°C were measured for Ty3 WT duplex and HIV WT duplex, respectively (Table S4). For Ty3 substitutions, dF(−9), dF(−11), and C(−8)T, the T_m was reduced to 60.0°C, 62.0°C, and 65.0°C, respectively. A similar effect was observed upon substituting the HIV-1 PPT at positions dF(+1), dF(−8), and C(−3)T. Melting temperatures of 61.1°C, 65.0°C, 64.0°C, respectively, were measured. For both the Ty3 C(−8)T and HIV C(−3)T samples, their T_ms reflect expected helical destabilization for nearest-neighbor model predictions (Sugimoto et al., 1995, 2000). The dF-associated melting temperatures reflect the effect of the loss of hydrogen bonding partly compensated by stronger stacking, resulting in a net destabilization of the helix. As stated earlier, NOEs observed in ^1H-^{19}F NOESY experiments showed that in all cases dF was well stacked within the hybrid helix (Figure S2; Guckian et al., 1998).

DISCUSSION

The exact mechanism by which HIV-1 or Ty3 RT recognizes its cognate PPT remains largely unknown despite extensive studies. Although there are significant differences between these sequences, a general mechanism of PPT recognition for both retroviruses and LTR-retrotransposons would not be unexpected. However, the only apparent similarities between the two PPT sequences are specific positioning of the RNase H domain over the PPT/U3 cleavage junction (Lener et al., 2003; Rausch and Le Grice, 2004; Wilhelm and Wilhelm, 2001). The structural motifs used in recognition may differ, but it is expected

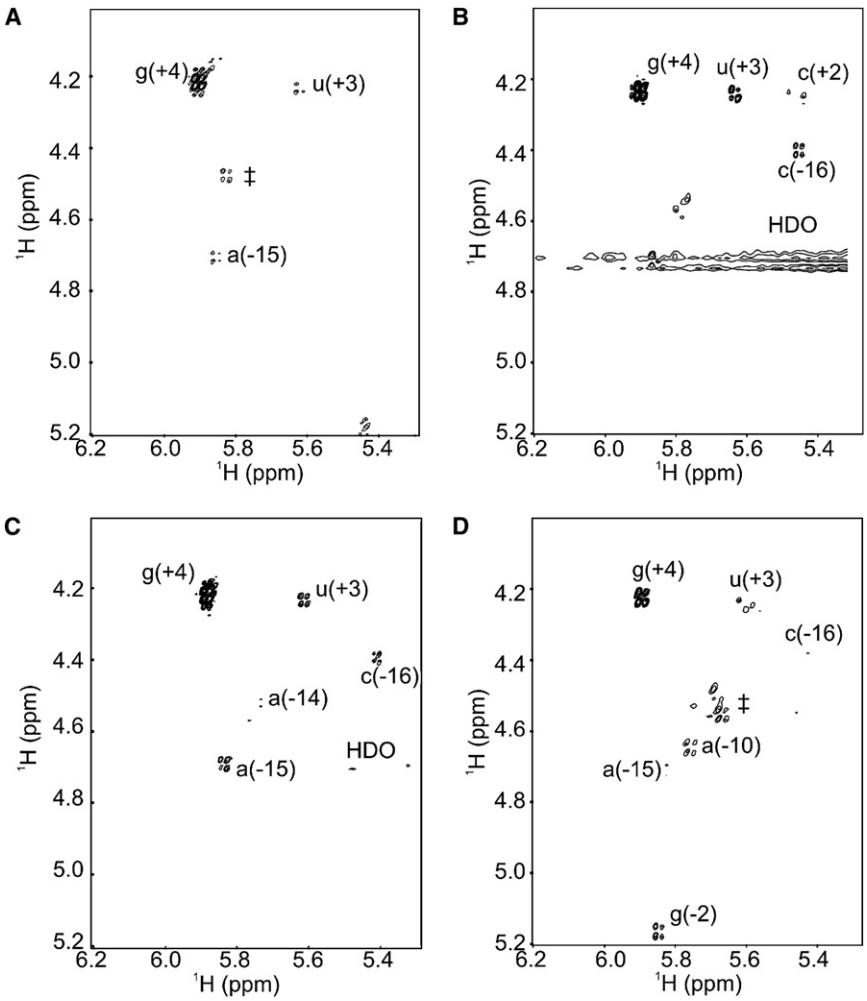

Figure 6. HIV-1 PPT H1′-H2′ Ribose Sugar Correlations

Expansion of the H1′-H2′ ribose sugar correlated region of DQF-COSY experiments from the (A) HIV WT; (B) HIV C(−3)T; (C) HIV dF(−8); and (D) HIV dF(+1) RNA/DNA hybrid duplexes at 30°C. The proton carrier frequency was set to 4.70 ppm. The data were acquired with sweep widths of 5,000 Hz in both dimensions, 4096 by 800 complex points, and 48 scans per increment. Note that unassigned cross peaks are indicated with a ‡. For HIV C(−3)T, the reduced signal-to-noise is due to a sample concentration of approximately 0.5 mM. All other samples had a concentration of approximately 1.0 mM.

that there are similar regions of distortion favoring RT binding in the appropriate orientation. Indeed, it has been shown that the HIV-1 RT nonspecifically degrades the Ty3 PPT; interestingly, the Ty3 RT apparently recognized a structural anomaly in the HIV-1 PPT in the upstream (rA):(dT) tracts and specifically hydrolyzed 12–13 bp downstream from this distortion (Lener et al., 2003).

What is immediately apparent from our data is that both the Ty3 and HIV-1 helixes, as a whole, contain normal Watson-Crick interactions in the absence of RT (Figure S1; Yi-Brunozzi et al., 2005). An imino NOE walk, which allows assignment of base pairing and stacking, could be traced for both Ty3 WT and HIV-1 WT duplexes. If there was any stable mispairing of bases in the PPT, the imino proton resonances of these mispaired bases would have been expect to be shifted upfield into a range from 10 to 11 ppm (Allawi and SantaLucia, 1998; Quignard et al., 1987). The ^{1}H spectrum for the HIV C(−3)T sample, in which the imino proton of T(−3) from the g:T(−3) mismatch is observed at 10.5 ppm (Figure 3B), highlights where an imino proton signal would be expected if the WT contained a G-T mismatch akin to what was observed in the RT•PPT crystal structure. As can be seen from the data, no such shifted imino proton signals are observed in the WT HIV-1 spectrum (Figure 3).

Line width analysis of the imino proton revealed broader line widths for g(−1), T(−7), and g(−11) (Table S1), supporting previous data suggesting a distortion in these regions of transition between the two $(rA)_4$:$(dT)_4$ tract as well as the transition from the central $(rA)_4$:$(dT)_4$ tract and the $(rG)_6$:$(dC)_6$ tract (Rausch et al., 2003). In HIV dF(−8), broader line widths of the imino protons of T(−7) and T(−9) serve as a control, because more solvent is known to penetrate the helix at the site of a dF substitution (Kool et al., 2000). Thus, it should be noted that, in the HIV WT, broadening of the T(−7) and g(−11) imino line widths is on the order of the broadening induced by a dF substitution. Although previous data suggested greater accessibility of thymines in the upstream $(rA)_4$:$(dT)_4$ tract to $KMnO_4$ oxidation (Kvaratskhelia et al., 2002), the imino proton NMR data do not indicate either mispairing or unpairing of bases in the HIV WT, as observed in the crystallographic data of the PPT-RT complex (Sarafianos et al., 2001). Nonetheless, the line width analysis corroborates the $KMnO_4$ data with the broadening of the imino line widths indicative of a change in the lifetime of base-pair opening and a potentially greater solvent accessibility for these bases (Guckian et al., 1996).

Another feature of our data is that both the HIV-1 and Ty3 PPTs appear dynamically coupled in nature. It is known that rR:dY hybrids primarily adopt a global A-form geometry, although local conformational changes in ribose sugar pucker to a more B-form like geometry have been observed (Fedoroff et al., 1997; Gyi et al., 1996). RNA strands in A-form geometry normally do exhibit H1′-H2′ correlations in a COSY spectrum, because the torsion angle between these protons in a C3′-endo sugar pucker has a negligible ^{3}J(H1′,H2′) couplings (<2Hz). Conversely, a helix that has B-form geometry and a C2′-endo sugar pucker has a relatively large ^{3}J(H1′,H2′) coupling of 6–8 Hz; correlations between H1′ and H2′ are observed in the COSY spectrum. We have previously reported a ribose sugar pucker switch in the RNA strand of the Ty3 PPT at position +1 relative to the PPT/U3 cleavage junction, from a C3′-endo to a mixed C3′/C2′-endo

conformation (Yi-Brunozzi et al., 2005). We postulated that the ribose sugar pucker switch at position +1 in the Ty3 PPT may contribute an additional "distortable" feature that ensures proper alignment of the RNase H active site for accurate cleavage.

In the current study, we sought to gain insight into the properties of both the Ty3 and HIV-1 PPTs as a structural motif by using dF as a structural probe to investigate the conformational state of the PPT. Previous biochemical studies incorporated nonnatural non-hydrogen-bonding nucleobases, such as dF, to probe the effects of hydrogen bonding on the processing of the PPTs. These studies, in general, found a decrease in the fidelity of RNase H processing for single dF substitutions and, for double dF substitutions, the RNase H domain was redirected to new sites (Lener et al., 2003; Rausch et al., 2003). A close correlation was also observed between the distance separating the dF insertion site and new RNase H cleavage position and the distance between the primer grip region and the RNase H catalytic domain of RT. The current study sought to probe for any structural perturbations resulting from single-site dF substitutions that might correlate with the previously reported biochemical data. As Figure 5 shows, significant alterations in the ribose sugar puckering patterns were observed for both Ty3 dF(−9) and Ty3 dF(−11). Both mutations resulted in an internal shift in the RNA strand to a more B-form-like geometry. Previous CD measurements on both Ty3 WT and dF-substituted hybrids have confirmed an overall intermediate conformation between A-form and B-form (Lener et al., 2003). What is especially intriguing about this phenomenon is that a mutation in a DNA base resulted in structural changes downstream in the RNA strand ribose sugars. This observation was not anticipated and cannot necessarily be correlated with the previous biochemical results; however, the effects clearly demonstrate a long-range structural coupling in the PPTs. In the most dramatic case, a dF substitution at position −11 of the Ty3 PPT induced a periodic propagation of altered sugar puckering in every other ribose to the 3′ end of the helix (Figure 5B). We suspect that this propagation may be a result of the 5′-a•g•a-3′ symmetric steps within the Ty3 PPT, whereas, in the HIV-1 PPT, the long-range structural effects were more limited. Indeed, only the HIV dF(+1) hybrid exhibited a long-range ribose sugar pucker switch at ribonucleotide a(−10). The asymmetry between the $(rG)_6$:$(dC)_6$ tract and two $(rA)_4$:$(dT)_4$ tracts may lessen the propagation of structural effects along the HIV-1 PPT.

The base-substituted samples, Ty3 C(−8)T and HIV C(−3)T, exhibited structural effects similar to those of the dF-substituted samples, suggesting that the observed structural effects in the latter samples are not solely the result of the incorporation of a fluorinated base. Although not as pronounced as that in Ty3 dF(−11), Ty3 C(−8)T showed a propagation of ribose sugar pucker switches to the 3′ end of the RNA strand. On the other hand, HIV C(−3)T exhibited ribose sugar puckering consistent with both HIV dF(−8) and HIV WT, where only end-fraying was observed. The G-tract is highly conserved across lentiviruses (Ilyinskii and Desrosiers, 1998), mutation of which tends to decrease the fidelity of RT-mediated processing (Rausch and Le Grice, 2007). However, no striking differences in the global structure of the HIV C(−3)T PPT mutant hybrid could be detected.

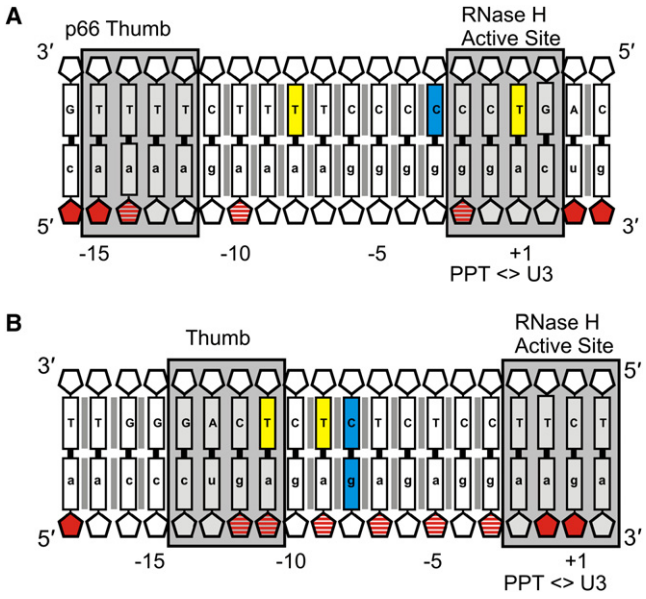

Figure 7. Ribose Sugar Pucker Switches Represented on the HIV-1 and Ty3 PPT Hybrids

A schematic representation of the (A) HIV PPT and (B) Ty3 PPT hybrid sequences. dF substitutions are in yellow. Base substitutions are in cyan. Nucleotides enclosed by a gray box denote the proposed RT binding site. WT ribose sugars in mixed C3′/C2′-endo conformation are shown in red. Highlighted collectively in red stripes are ribose sugars for which a C3-endo to C3′/C2′-endo sugar pucker switch is observed when the PPT hybrids are either substituted with a dF base or a base pair change is made.

Figure 7 summarizes these detected ribose sugar pucker switches, with red indicating those observed in the WT duplexes and red stripes for samples with substitutions. The disruption or alteration of the WT structure and/or allosteric effects within the PPT by dF substitution may be correlated with the observed reduction in the fidelity of PPT processing in the previous biochemical studies (Lener et al., 2003; Rausch et al., 2003). Overall, a greater helical flexibility is exhibited for the Ty3 PPT. Although this may be due to the differences in sequences as described above, this is also unexpected in light of the mispaired and unpaired bases observed in the HIV-1 PPT•RT crystal structure. To determine whether the mispaired and unpaired bases observed in the HIV-1 PPT•RT were induced by the binding of RT, an attempt was made to examine the effect of RT binding on the imino proton spectra of the HIV-1 PPT by solution state NMR (Turner et al., 2008). In this experiment, general broadening of all imino proton signals was observed as might be expected because of the size of the complex, while more significant signal broadening and some minor chemical shift perturbations were observed for resonances associated with bases in the 5′$(rA)_4$:$(dT)_4$ and the PPT-U3 junction where RT is predicted to make direct contacts. No evidence for a G•T mismatch, however, was observed in this spectrum.

SIGNIFICANCE

These studies represent the first, to our knowledge, solution NMR studies of the entire PPT sequence of the RNA/DNA

hybrid duplex from the retrovirus HIV-1 and our ongoing investigation of the LTR-retrotransposon Ty3. We found that, when the thymine isostere dF was substituted at single positions, subtle changes were observed in the conformational states. The inherent PPT flexibility suggests that upon RT binding, structural perturbations throughout the sequence promote contacts with the appropriate protein motifs, leading to productive binding and cleavage at the correct site. Together, the dF and base substitution data suggest that the PPT has a coupled dynamic structure that is likely tied through specific base stacking in the purine tracts that governs long-range structural signaling through alteration of helical geometry. More generally, the long-range structural coupling observed in the HIV-1 and Ty3 PPT hybrids is quite unique and rarely seen in nucleic acid helical structures. On a technical note, our ability to assign a majority of the carbon and proton resonances, despite the severe challenge posed by high redundancy in sequence, demonstrates a crucial step toward high-resolution solution structures of these PPTs.

EXPERIMENTAL PROCEDURES

Sample Preparation

All RNA and DNA phosphoramidites, including dF, were purchased from Glen Research Corp. (Sterling, VA). Substituted and unsubstituted 20 nt DNA and 20 nt RNA oligonucleotides were synthesized at a 1 μmol scale using standard phosphoramidite chemistry on a PE Biosystems Expedite 8909 nucleic acid synthesizer. The individual strands were purified using preparative polyacrylamide gel electrophoresis, were electroeluted, and were dialyzed, first against sterile ddH_2O and then against NMR buffer (80 mM NaCl and 10 mM phosphate [pH 7.0]). The RNA and DNA single strands were annealed by mixing equamolar amounts, heating the solution to 90°C for 3 min, and allowing the samples to cool slowly to form the duplex. Duplex concentrations were determined from absorbance at 260 nm (CARY300 UV-Vis Spectrophotometer), using a molar extinction coefficient that is equal to the sum of the ribo- and deoxyribonucleotides (384,900 cm^{-1} M^{-1} and 379,700 cm^{-1} M^{-1} for Ty3 and HIV-1 PPT duplexes, respectively). Final duplex concentrations for NMR measurements were typically 1.0 mM, except for HIV C(−3)T, which was measured at 0.5 mM.

NMR Spectroscopy

NMR samples were examined in 80 mM NaCl and 10 mM phosphate (pH 7.0) that were dissolved in 90% H_2O/10% D_2O or 99.96% D_2O. Data were collected on a Bruker Biospin DMX600 spectrometer equipped with an actively shielded triple-axis gradient, triple resonance 600 MHz TXI probe, or a three-channel Bruker Biospin *Avance* 600-MHz spectrometer equipped with an actively shielded z-axis gradient triple resonance TXI cryoprobe. ^{19}F experiments were performed on a Bruker Biospin DMX600 spectrometer equipped with a PH SEF 600SB F-H-D-05 Z probe. All spectra were processed using NMRPipe (Delaglio et al., 1995) and were analyzed using Sparky on a PC/Linux workstation. Assignments were made using DQF-COSY, NOESY, TOCSY, ^{1}H-^{13}C (natural abundance) HMQC, ^{1}H-^{19}F HSQC, and ^{1}H-^{19}F NOESY. Other experimental details and acquisition parameters are given in the figure legends.

Thermal Denaturation

Thermal denaturation of the RNA/DNA hybrids (1.5 μM) was measured using a Cary 300 UV-Visible Spectrophotometer (Varian) equipped with peltier thermostattable multicell holder. Absorbance at 260 nm was monitored as a function of temperature at 0.5°C intervals from 10°C to 90°C. The T_m was determined from the first derivative function provided by the manufacturer.

SUPPLEMENTAL DATA

Supplemental Data include three figures and four tables and can be found with this article online at http://www.chembiol.com/cgi/content/full/15/3/254/DC1/.

ACKNOWLEDGMENTS

This work was supported in part by the National Institutes of Health (grant GM 59107 to J.P.M.). S.F.J.L.G. is supported by the intramural research program of the National Cancer Institute, National Institutes of Health. NMR instrumentation was purchased in part with support from the W.M. Keck Foundation, the NIH/NCRR, and NIST. R.G.B. is a NIST/NIH NRC postdoctoral fellow. Certain commercial equipment, instruments, and materials are identified in this paper in order to specify the experimental procedure. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the material or equipment identified is necessarily the best available for the purpose.

Received: November 26, 2007
Revised: January 24, 2008
Accepted: January 30, 2008
Published: March 21, 2008

REFERENCES

Allawi, H.T., and SantaLucia, J., Jr. (1998). NMR solution structure of a DNA dodecamer containing single G-T mismatches. Nucleic Acids Res. *26*, 4925–4934.

Champoux, J.J. (1993). Roles of ribonuclease H in reverses transcription. In Reverse Transcriptase, A.M. Skalka and S.P. Goff, eds. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), pp. 103–118.

Conn, G.L., Brown, T., and Leonard, G.A. (1999). The crystal structure of the RNA/DNA hybrid r(GAAGAGAAGC):d(GCTTCTCTTC) shows significant differences to that found in solution. Nucleic Acids Res. *27*, 555–561.

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.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR *6*, 277–293.

Fedoroff, O.Y., Ge, Y., and 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.

Guckian, K.M., Schweitzer, B.A., Ren, R.X.-F., Sheils, C.J., Paris, P.L., Tahmassebi, D.C., and Kool, E.T. (1996). Experimental measurement of aromatic stacking affinities in the context of duplex DNA. J. Am. Chem. Soc. *118*, 8182–8183.

Guckian, K.M., Krugh, T.R., and Kool, E.T. (1998). Solution structure of a DNA duplex containing a replicable difluorotoluene-adenine pair. Nat. Struct. Biol. *5*, 954–959.

Gyi, J.I., Conn, G.L., Lane, A.N., and Brown, T. (1996). Comparison of the thermodynamic stabilities and solution conformations of DNARNA hybrids containing purine-rich and pyrimidine-rich. Biochemistry *35*, 12538–12548.

Huang, H., Chopra, R., Verdine, G.L., and Harrison, S.C. (1998). Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. Science *282*, 1669–1675.

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.

Jacobo-Molina, A., Ding, J., Nanni, R.G., Clark, A.D., Jr., Lu, X., Tantillo, C., Williams, R.L., Kamer, G., Ferris, A.L., Clark, P., et al. (1993). Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 A resolution shows bent DNA. Proc. Natl. Acad. Sci. USA *90*, 6320–6324.

Kool, E.T., and Sintim, H.O. (2006). The difluorotoluene debate—a decade later. Chem. Commun. (Camb). *35*, 3665–3675.

Kool, E.T., Morales, J.C., and Guckian, K.M. (2000). Mimicking the structure and function of DNA: insights into DNA stability and replication. Angew. Chem. Int. Ed. Engl. *39*, 990–1009.

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.

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.

Lener, D., Budihas, S.R., and Le Grice, S.F. (2002). Mutating conserved residues in the ribonuclease H domain of Ty3 reverse transcriptase affects specialized cleavage events. J. Biol. Chem. *277*, 26486–26495.

Lener, D., Kvaratskhelia, M., and Le Grice, S.F.J. (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.

Li, F., Pallan, P.S., Maier, M.A., Rajeev, K.G., Mathieu, S.L., Kreutz, C., Fan, Y., Sanghvi, J., Micura, R., Rozners, E., et al. (2007). Crystal structure, stability and in vitro RNAi activity of oligoribonucleotides containing the ribo-difluorotoluyl nucleotide: insights into substrate requirements by the human RISC Ago2 enzyme. Nucleic Acids Res. *35*, 6424–6438.

Quignard, E., Fazakerley, G.V., van der Marel, G., van Boom, J.H., and Guschlbauer, W. (1987). Comparison of the conformation of an oligonucleotide containing a central G-T base pair with the non-mismatch sequence by proton NMR. Nucleic Acids Res. *15*, 3397–3409.

Rance, M., Sørensen, O.W., Bodenhausen, G., Wagner, G., Ernst, R.R., and Wüthrich, K. (1983). Improved spectral resolution in cosy 1H NMR spectra of proteins via double quantum filtering. Biochem. Biophys. Res. Commun. *117*, 479–485.

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.

Rausch, J.W., and Le Grice, S.F. (2007). Purine analog substitution of the HIV-1 polypurine tract primer deines regions controlling initiation of plus-strand DNA synthesis. Nucleic Acids Res. *35*, 256–268.

Rausch, J.W., Grice, M.K., Nymark-McMahon, M.H.,, Miller, J.T., and Le Grice, S.F. (2000). Interaction of p55 reverse transcriptase from the *Saccharomyces cerevisiae* retrotransposon Ty3 with conformationally distinct nucleic acid duplexes. J. Biol. Chem. *275*, 13879–13887.

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 pyrmidine isosteres defines regions essential for HIV type 1 polypurine tract selection. Proc. Natl. Acad. Sci. USA *100*, 11279–11284.

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.

Schweitzer, B.A., and Kool, E.T. (1995). Hydrophobic, non-hydrogen-bonding bases and base pairs in DNA. J. Am. Chem. Soc. *117*, 1863–1872.

Silverman, A.P., and Kool, E.T. (2007). RNA probes of steric effects in active sites: high flexibility of HIV-1 reverse transcriptase. J. Am. Chem. Soc. *129*, 10626–10627.

Snoussi, K., and Leroy, J.L. (2001). Imino proton exchange and base-pair kinetics in RNA duplexes. Biochemistry *40*, 8898–8904.

Sugimoto, N., Nakano, S., Katoh, M., Matsumura, A., Nakaumta, H., Ohmichi, T., Yoneyama, M., and Sasaki, M. (1995). Thermodynamic parameters to predict stability of RNA/DNA hybrid duplexes. Biochemistry *34*, 11211–11216.

Sugimoto, N., Nakano, M., and Nakano, S. (2000). Thermodynamics—structure relationship of single mismatches in RNA/DNA duplexes. Biochemistry *39*, 11270–11281.

Telesnitsky, A., and Goff, S.P. (1997). Reverse Transcriptase and the Generation of Retroviral DNA (Plainview, NY: Cold Spring Harbor Laboratory Press).

Turner, K.B., Brinson, R.G., Yi-Brunozzi, H.Y., Miller, J.T., Rausch, J.W., Le Grice, S.F.J., Marino, J.P., and Fabris, D. (2008). Structural probing of the HIV-1 polypurine tract RNA:DNA hybrid using classic nucleic acid ligands. Nucleic Acids Res., in press.

Varani, G., Aboul-ela, F., and Allain, F.H.-T. (1996). NMR investigation of RNA structure. Prog. Nucleic Mag. Res. Spect. *29*, 51–127.

Wijmenga, S.S., and van Buuren, B.N.M. (1998). The use of NMR methods for conformational studies of nucleic acids. Prog. Nucleic Mag. Res. Spect. *32*, 287–387.

Wilhelm, M., and Wilhelm, F.-X. (2001). Reverse transcription of retroviruses and LTR retrotransposons. Cell. Mol. Life Sci. *58*, 1246–1262.

Xia, J., Noronha, A., Toudjarska, I., Li, F., Akinc, A.,, Braich, R.,, Frank-Kamenetsky, M.,, Rajeev, K.G.,, Egli, M.,, and Manoharn, M. (2006). Gene silencing activity of siRNAs with ribo-difluorotoluyl nucleotide. ACS Chem. Biol. *1*, 176–183.

Yi-Brunozzi, H.Y., Brabazon, D.M., Lener, D., Le Grice, S.F., and Marino, J.P. (2005). A ribose sugar conformational switch in the LTR-retrotransposon Ty3 polypurine tract-containing RNA/DNA hybrid. J. Am. Chem. Soc. *127*, 16344–16345.

