Structural Studies of RNA/DNA Polypurine Tracts

Megan E. Fitzgerald[1] and Alexander C. Drohat[1,2,*]
[1]Department of Biochemistry and Molecular Biology
[2]Greenebaum Cancer Center
University of Maryland School of Medicine, Baltimore, MD 21201, USA
*Correspondence: adroh001@umaryland.edu
DOI 10.1016/j.chembiol.2008.03.001

During reverse transcription, an RNA polypurine tract (PPT) resists digestion by reverse transcriptase (RT) and primes plus-strand DNA synthesis. In this issue of *Chemistry & Biology*, Yi-Brunozzi et al. (2008) report structural studies of PPTs, illuminating how they are recognized by RT.

During the life cycle of retroviruses including HIV-1, reverse transcriptase (RT) converts single-stranded viral RNA into double stranded DNA for integration into the host genome. While minus-strand DNA synthesis is primed using a host-derived tRNA annealed to the RNA template, plus-strand DNA synthesis is initiated from a purine-rich segment of viral RNA known as the polypurine tract (PPT). In a remarkable and still poorly understood process, RT nonspecifically degrades viral RNA from the nascent RNA/DNA hybrid while selectively preserving the PPT, and the RNase H domain of RT cleaves precisely at the 3′ end of the PPT to create the primer needed for proper initiation of plus-strand DNA synthesis and subsequent integration (Rausch and Le Grice, 2004).

A crystal structure of HIV-1 RT bound to an RNA/DNA hybrid containing the HIV-1 PPT showed the vast majority of interactions involve the sugar-phosphate backbone, with very few contacts to the nucleobases of the PPT (Sarafianos et al., 2001). Although not surprising, this observation underscores the challenge of determining exactly how RT recognizes the PPT. In the crystal structure, the RNase H active site was not positioned at the target PPT-U3 junction; rather, it was centered 9 nucleotides upstream, in the central a·T tract (Figure 1). Thus, the structure may be more relevant to understanding how RT avoids cleavage within the PPT region than how it targets the PPT-U3 junction. One striking observation was a region of unpaired, mispaired (i.e., g·T), and weakly paired bases at positions −9 to −15 with respect to the PPT-U3 junction (Figure 1). The altered base pairing in this "unzipped" region may be important for PPT recognition, and raised the immediate question of whether these anomalies are imposed by RT or if they are an inherent property of the PPT.

To address this question, a number of biochemical studies have probed the HIV-1 PPT structure in the absence of RT. As indicated in Figure 1, chemical footprinting studies ($KMnO_4$) showed increased reactivity for thymines at positions −10, and −12 through −15, indicating greater solvent accessibility for these bases (Kvaratskhelia et al., 2002). Enhanced solvent accessibility was also indicated for the +1 thymine, suggesting a structural perturbation at the RNase H cleavage site. This is consistent with NMR studies of an 8 bp RNA/DNA hybrid derived from residues at the PPT-U3 junction (Fedoroff et al., 1997). In another study, the stability of base pairing in the g·C tract was probed by substituting dC nucleotides with pyrrolo-dC (pdC), a cytosine analog whose fluorescence is quenched upon forming base pairs with guanine (Dash et al., 2004). The observation that pdC at DNA position −11 exhibited nearly the same fluorescence in free DNA as for the RNA/DNA hybrid is consistent with the possibility that C(−11) could be unpaired, as it is in the RT-PPT crystal structure. Alternatively, it could indicate weakened base pairing and/or bending of the helical axis at this site, which joins the two a·T tracts (Figure 1). Consistent with the latter possibility, the crystal structure (1.1 Å) of a 10 bp RNA/DNA hybrid derived from the 5′ end of HIV-1 PPT revealed a 15° helical bend across the a-g-a junction of the two straight a·T tracts (Kopka et al., 2003). Importantly, this structure did not exhibit the altered base pairing interactions seen in the RT-PPT structure. The somewhat mixed message resulting from these biochemical and structural studies indicated that additional work was needed, using the full-length PPT domain, to determine whether altered base pairing exists in the absence of RT.

In this issue of *Chemistry & Biology*, Yi-Brunozzi et al. (2008) report studies employing high resolution NMR and selective base substitutions to investigate

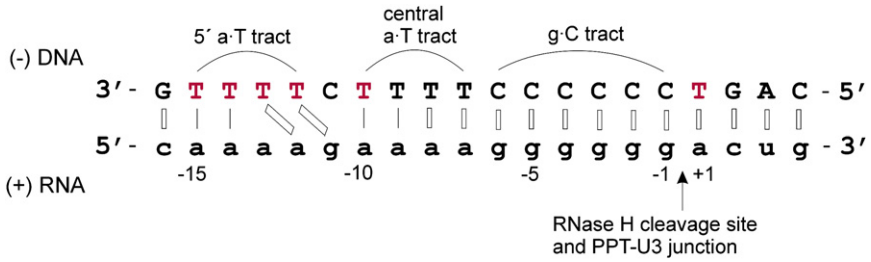

Figure 1. Model of the HIV-1 PPT Domain
A model of the HIV-1 PPT domain that shows the altered base pairing interactions observed in a crystal structure of RT bound to an RNA/DNA hybrid containing the PPT domain (Sarafianos et al., 2001) (thin lines indicate weakened base pairing). Also indicated are thymine bases (red) that exhibited enhanced solvent accessibility in chemical ($KMnO_4$) footprinting studies (Kvaratskhelia et al., 2002). RNA nucleotides are numbered with respect to the RNase H cleavage site (i.e., the PPT-U3 junction). The PPT domain includes −1 through −15. The entire 20 bp hybrid shown is that used for the NMR studies reported by Yi-Brunozzi et al. (2008). Their findings indicate that the altered base pairing observed in the RT-PPT structure (shown above) is not present in the PPT in the absence of RT.

the structure of PPT domains in solution. They examined two 20 bp RNA/DNA hybrids, one containing the full-length PPT from HIV-1, and the other containing the intact PPT from *Saccharomyces cerevisiae* Ty3, a long terminal repeat (LTR)-containing retrotransposon. As is the case for retroviruses, recognition and selective processing of the PPT is required for reverse transcription in Ty3 and other LTR-containing retrotransposons. Previous NMR studies by Le Grice, Marino, and coworkers (Yi-Brunozzi et al., 2005) revealed normal Watson-Crick base pairing throughout the Ty3 PPT, and a sugar pucker switch (C3′-*endo* to mixed C3′/C2′-*endo*) at the RNA +1 site, which may contribute to recognition of the cleavage site by RT (Yi-Brunozzi et al., 2005). In the current study, Yi-Brunozzi et al. find normal Watson-Crick type base pairing throughout the HIV-1 PPT domain (Yi-Brunozzi et al., 2008), in contrast to suggestions of pre-existing structural anomalies in the PPT from some biochemical studies and the RT-PPT crystal structure (Sarafianos et al., 2001). Moreover, they find no evidence for a g·T mispair, which was seen in the RT-PPT structure. Convincingly, they demonstrate an ability to detect a g·T mispair by introducing one in the 5′ region of the PPT, giving the expected decrease in chemical shift for the imino protons of the guanine and thymine bases. Thus, Yi-Brunozzi et al. find no evidence that the unpaired and mispaired bases observed in the RT-PPT crystal structure are an inherent property of the PPT. However, they do find evidence of weakened base pairing for some regions of the HIV-1 PPT, as suggested by increased linewidths for the imino protons of some guanine and thymine bases. The linewidth of an imino proton increases with its rate of solvent exchange, thus increased linewidth reflects, among other things, a greater propensity for the base to be extrahelical. Increased imino linewidths were observed at transitions between the a·T tracts and the g·C tract of the HIV-1 PPT, including g(−11), which is consistent with structural and biochemical studies suggesting a bend at this location (Dash et al., 2004; Kopka et al., 2003).

Another remarkable aspect of the report from Yi-Brunozzi et al. is their success in obtaining nearly complete chemical shift assignments for the 20 bp RNA/DNA hybrids, and the approach used to accomplish this task. The chemical shifts of nucleic acids can be very difficult to assign due to overlap of the NMR signals, and this problem is even worse for large and repetitive duplexes like the HIV-1 and Ty3 PPTs. To overcome this problem, they selectively replaced dT nucleotides with an analog known as 2,4-difluoro-5-methylbenzene (dF), which mimics the size and shape of thymine, and stacks with neighboring bases, but does not form significant hydrogen bond interactions (Kool, 2002). The dF substitutions in HIV-1 and Ty3 PPTs served to disperse the NMR signals of neighboring bases, enabling their assignment. In a similar manner, shift assignments were facilitated by selective base pair substitutions, e.g., they replaced a g·C pair with a·T in the Ty3 PPT. These methods may be useful to other investigators studying the structure of large and/or repetitive nucleic acids. In addition, the chemical shift assignments constitute the critical first step toward solving the solution structure of the intact PPTs from HIV-1 and Ty3.

Finally, the dF substitutions employed by Yi-Brunozzi et al. provide insight into the unique structural properties of these PPTs, and how RTs may recognize them. Single-site dF substitutions in both the Ty3 and HIV-1 PPTs resulted in altered sugar puckering for nucleotides well removed from the site of dF substitution. For example, the dF substitution at DNA position −11 of the Ty3 PPT resulted in a periodic switch in sugar pucker for upstream RNA nucleotides (positions −9, −7, −5, −3, and −1), indicating some degree of long-range structural coupling within the PPT. The effects of dF substitutions were more limited for the HIV-1 PPT, which the authors suggest may be explained by its asymmetrical structure; the g·C and a·T tracts may attenuate the propagation of structural changes. Nevertheless, dF substitution at the DNA +1 site resulted in a long-range effect at RNA site −10, consistent with the suggestion from previous biochemical studies of long-range effects between these regions of the PPT (Kvaratskhelia et al., 2002).

In summary, the work reported by Yi-Brunozzi et al. (2008) is remarkable for (1) providing significant insight into the structure of full-length PPT domains from HIV-1 and Ty3, informing the mechanism by which they are recognized by RT, (2) providing a stepping stone for determining the solution structure of the intact PPTs and for future NMR studies designed to address specific questions regarding PPT recognition, and (3) suggesting a productive approach for assigning chemical shifts in large and/or repetitive nucleic acids. This work may also be beneficial for the development of compounds designed to inhibit HIV-1 replication, by targeting recognition and/or processing of the PPT domain by RT.

REFERENCES

Dash, C., Rausch, J.W., and Le Grice, S.F.J. (2004). Nucleic Acids Res. *32*, 1539–1547.

Fedoroff, O.Y., Ge, Y., and Reid, B.R. (1997). J. Mol. Biol. *269*, 225–239.

Kool, E.T. (2002). Acc. Chem. Res. *35*, 936–943.

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

Kvaratskhelia, M., Budihas, S.R., and Le Grice, S.F. (2002). J. Biol. Chem. *277*, 16689–16696.

Rausch, J.W., and Le Grice, S.F. (2004). Int. J. Biochem. Cell Biol. *36*, 1752–1766.

Sarafianos, S., Das, K., Ding, J., Whitcomb, J., Boyer, P., Hughes, S., and Arnold, E. (2001). EMBO J. *20*, 1449–1461.

Yi-Brunozzi, H.Y., Brabazon, D.M., Lener, D., Le Grice, S.F., and Marino, J.P. (2005). J. Am. Chem. Soc. *127*, 16344–16345.

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

