

Selection of primer-template sequences that bind human immunodeficiency virus reverse transcriptase with high affinity

Jeffrey J. DeStefano* and Jason V. Cristofaro

Department of Cell Biology and Molecular Genetics, University of Maryland, Building 231, College Park, MD 20742, USA

Received October 26, 2005; Revised November 29, 2005; Accepted December 15, 2005

ABSTRACT

A SELEX (systematic evolution of ligands by exponential enrichment)-based approach was developed to determine whether HIV-RT showed preference for particular primer-template sequences. A 70 nt duplex DNA was designed with 20 nt fixed flanking sequences at the 3' and 5' ends and a randomized 30 nt internal sequence. The fixed sequence at the 5' end contained a BbsI site six bases removed from the randomized region. BbsI cuts downstream of its recognition site generating four base 5' overhangs with recessed 3' termini. Cleavage produced a 50 nt template and 46 nt primer with the 3' terminus within the randomized region. HIV-RT was incubated with this substrate and material that bound RT was isolated by gel-shift. The recovered material was treated to regenerate the BbsI site, amplified by PCR, cleaved with BbsI and selected with HIV-RT again. This was repeated for 12 rounds. Material from round 12 bound approximately 10-fold more tightly than starting material. All selected round 12 primer-templates had similar sequence configuration with a 6–8 base G run at the 3' primer terminus, similar to the HIV polypurine tract. Further modifications indicate that the Gs were necessary and sufficient for strong binding.

INTRODUCTION

Using an approach developed in the 1990s called SELEX (systematic evolution of ligands by exponential enrichment) (1–3), single-stranded nucleic acids that bind to human immunodeficiency virus reverse transcriptase (HIV-RT) with high affinity were identified (4–6). The SELEX method

is based on differential binding of nucleic acids to a substrate protein. Initially a large random pool of RNAs is incubated with a limiting amount of protein. Nucleic acids that bind with higher affinity will preferentially associate with the protein and can be isolated by gel-shift or nitrocellulose filter binding. The selected pool is expanded using PCR then RNA transcription and the new pool is subjected to another round of protein binding. After several rounds nucleic acids with high affinity for the protein, also referred to as aptamers, can be isolated. In the case of HIV-RT, these RNA aptamers were usually pseudoknot-type structures (7). RNA pseudoknot aptamers have been shown to interfere with primer-template binding and are potent inhibitors of reverse transcription (8,9). Single-stranded DNA aptamers with similar properties have also been selected (10). For both RNA and DNA, the main factor underlying tight binding to RT was the folded structure rather than the sequence of the nucleic acids. This method has since been used to isolate aptamers that can bind several different proteins, including several therapeutic targets [reviewed in (11,12)]. Many of these aptamers are currently being developed as potential treatments for diseases.

The natural substrate for RT and most other polymerases is duplex nucleic acid with a recessed 3' terminus (classical primer-template configuration). It is well established that RT binds DNA primers on an RNA template more tightly than on DNA templates; however, it is not known whether strong sequence preferences exist. Tight binding sequences could potentially be used to help design aptamers and aid in understanding how RT recognizes and binds to substrates. To determine whether tight binding DNA–DNA primer-template sequences could be isolated, we used a novel SELEX-based approach. Primer-template sequence that bound 10-fold tighter than random sequences were selected. All the recovered sequences had the same basic sequence motif characterized by a run of several G residues at the primer 3' terminus. In such, they mimicked the sequence of the polypurine tract (ppt) RNA primer used by HIV to initiate plus strand DNA synthesis.

*To whom correspondence should be addressed. Tel: +1 301 405 5449; Fax: +1 301 314 9489; Email: jdestefa@umd.edu

MATERIALS AND METHODS

Materials

Wild-type HIV-RT was from Worthington Biochemical Corporation. TY-3 RT was a kind gift from Dr Stuart Le Grice (HIV Drug Resistance Program, National Cancer Institute). Drug resistant HIV-RTs (D67N/K70R/T215F/K219Q and K65R) were a kind gift from Dr Michael Parniak, (University of Pittsburg). MuLV-RT was from United States Biochemical. *Taq* polymerase was from Eppendorf. Restriction enzymes, Klenow polymerase, and *Pfu* polymerase and T4 polynucleotide kinase were from New England Biolabs. Calf intestinal phosphatase and dNTPs were from Roche. Rapid DNA ligation kit was from Promega. Competent GC5 *Escherichia coli* cells were from Gene Choice. Miniprep DNA preparation kit was from Qiagen. Radiolabeled compounds were obtained from Amersham. Sephadex G-25 spin columns were from Amika Corp. All oligonucleotides were from Integrated DNA Technologies. All other chemicals were from Sigma or Fisher Scientific.

Methods

Production of substrate for initial round of selection. About 2500 pm of template 5'-GCATGAATTCCTCGAAGACGC(N)₃₀TCTAGAGTCGACCTGCAGGC-3' (N-any nucleotide) was hybridized to ~3000 pm of primer 5'-GCCTGCAGGTC-GACTCTAGA-3' that was 5' end-labeled with P-32 using T4 polynucleotide kinase and the manufacturer's protocol. Hybridization was performed in 50 mM Tris-HCl (pH = 8), 1 mM DTT and 80 mM KCl in a volume of 100 µl. The mixture was heated to 80°C then slow cooled to room temperature. The material was divided into five tubes and the primer in each was extended with 15 U of Klenow polymerase using the following conditions: 50 mM Tris-HCl (pH = 8), 1 mM DTT, 25 mM KCl, 6 mM MgCl₂ and 500 µM dNTPs. Reactions were for 2–3 h at 37°C. The material was then combined, extracted with phenol:chloroform:isoamyl alcohol (25:24:1 v:v:v) and precipitated with 2 vol of ethanol and 1/10th vol of 3 M Na acetate (pH = 7). The recovered material was divided into four reaction tubes and digested with BbsI (50 U) in the appropriate buffer in a volume of 150 µl overnight at 37°C. Material was run on a 13% native polyacrylamide gel along with a small amount of uncut starting material. About 1/2 of the material was digested to the appropriate size as judged by UV shadowing. Digested material was recovered by eluting from crushed gel slices overnight in buffer containing 50 mM Tris-HCl (pH = 8), 1 mM DTT, 80 mM KCl and 0.1 mM EDTA (pH = 8) as described previously (13). Spectrophotometric analysis indicated that ~600 pm was recovered.

Selection of material with HIV-RT using gel-shift and subsequent processing. In the initial round, ~500 pm of substrate from above was incubated with 20 pm of HIV-RT in a buffer containing 50 mM Tris-HCl (pH = 8), 80 mM KCl, 6 mM MgCl₂, 1 mM DTT and 0.1 mM EDTA (pH = 8). Incubation was for 1 h at room temperature in a volume of 100 µl. For rounds 11 and 12 the KCl concentration was increased to 150 mM. Material was then run on a 6% native 1.5 mm polyacrylamide gel at 125 V as described previously (13). A control reaction containing 25 pm of material without

RT was also run. Shifted material was excised and eluted as described above. Approximately 2–4 pm of material was recovered as determined from specific activity. The recovered material was extended with Klenow (2.5 U) to produce blunt ends as described above except that 100 µM dNTPs were used, the reaction volume was 50 µl, and the incubation time was 15 min. The reaction was extracted and precipitated as described above, then resuspended in 10 µl water. The recovered material was ligated using a rapid ligation kit (Promega) to 5–10-fold excess of the following two primers that were hybridized together to form a duplex: 5'-ATAGCATGAATTCGCAGAAGACCC-3' and 5'-GGGTCTTCTGCGAATTCATGC-3'. Ligations were in a volume of 30 µl for round one and 10 µl for all subsequent rounds using the manufacturer's protocol except that reactions were for 2 h. Note that hybridization of the primers will form a duplex with a blunt end and an end with a 3 base 5' overhang. This was to help direct the ligations (see Figure 1). Also, the protocol was designed to direct ligations as shown in Figure 1 since the only 5' phosphorylated nucleic acid end in the reactions was on the blunt ended randomized region of the substrate ('N' containing region). A very small proportion of 5' ends would have been phosphorylated at the other end of the selected material due to 5' end labeling with P-32 (see above). After ligation, PCRs were prepared in 400 µl total volume using the *Taq* polymerase buffer provided by the manufacturer. Approximately 400 pm of each primer, (5'-GCCTGCAGGTCGACTCTAGA-3 (P-32 end-labeled) and 5'-GCATGAATTCGCAAGACCC-3'), were added to reactions along with the entire ligation mixture. The reactions were divided into four tubes (100 µl each) and PCR was performed at 94°C (1 min), 50°C (1 min) and 72°C (1 min) for a total of 12–14 cycles followed by an additional cycle of 5 min at 72°C. Material was combined, extracted and ethanol-precipitated before digestion with BbsI (50 U total in one 100 µl reaction as described above). The digested material was run on a 13% native polyacrylamide gel and cleaved material was recovered as described above and used in subsequent rounds of selection. A total of 12 rounds of selection were performed. In each round after round one, ~1/20th equivalent of HIV-RT (mole:mole) was incubated with the recovered substrate for the next round of selection. Also, after round 2, 1/5th of the recovered PCR material was saved as a source to regenerate the selected material from that round or for use in K_d determinations (see below).

Determination of equilibrium dissociation constants (K_d). Selected material (1 nM) from various rounds of selection (see Results) or other designed primer-templates (see Figure 3) were end-labeled on the primer strand and mixed with various amounts of HIV-RT (2.5, 5, 10, 15, 20, 40, 60, 80 and 100 nM unless otherwise indicated) in 8 µl of buffer containing 50 mM Tris-HCl (pH = 8), 1 mM DTT, 80 mM KCl, 6 mM MgCl₂ and 0.1 µg/µl BSA for 5 min at room temperature. Reactions were initiated by the addition of 2 µl dNTPs (100 µM final in reactions) and heparin 'trap' (1 µg/µl final in reactions) in the same buffer as above. The trap was added to sequester RT molecules not bound to the substrate and those that dissociate. This limits extension to a single binding event between the substrate and enzyme (14). Samples were incubated for 2 min then stopped with an equal volume of 2× gel

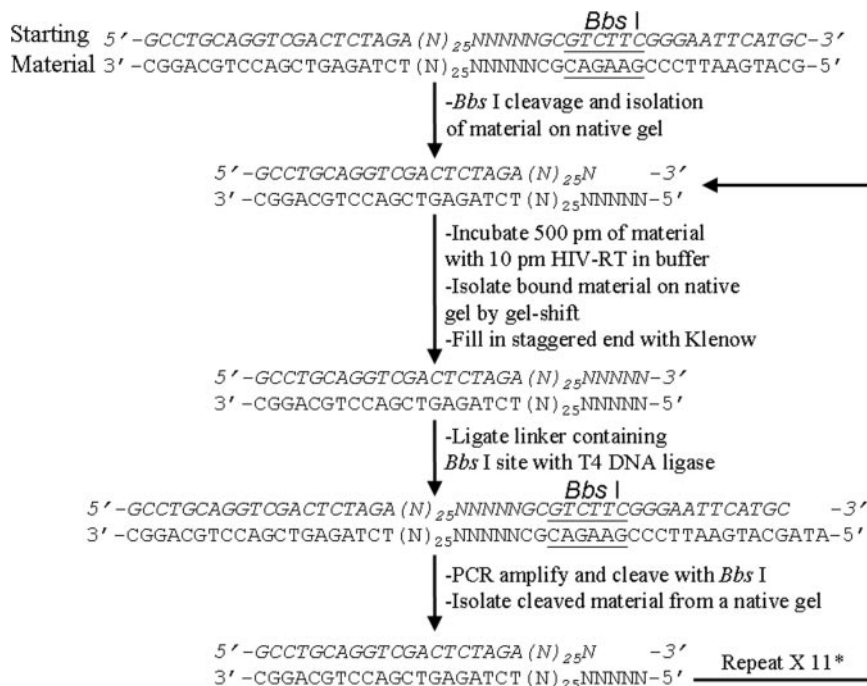


Figure 1. SELEX approach used to select tight binding sequences. The diagram illustrates the basic approach used to isolate primer-template sequences that bound HIV-RT with high affinity. The double-stranded DNA starting material contained a 30 nt random region flanked by two 20 nt fixed sequences. The BbsI cleavage site in the 5' fixed sequences was used to generate substrates with recessed 3' termini in the random region. These substrates (500 pm in the first round) were incubated with a limiting amount of HIV-RT and bound material was isolated by gel-shift. The recovered material was processed to essentially regenerate the starting material configuration. PCR was then used to amplify the selected material which was cleaved with BbsI and recovered by gel electrophoresis. This recovered material was then subjected to another round of selection for a total of 12 rounds. *, In rounds subsequent to the first enough RT to bind $\sim 1/20$ th of the total recovered material was added. See Methods for details.

loading buffer [90% formamide, 10 mM EDTA (pH = 8), 0.25% each bromphenol blue and xylene cyanol]. The reactions were run on a 12% denaturing polyacrylamide gel as described below, and dried gels were imaged using a BioRad FX phosphorimager. The amount of bound enzyme at the various concentrations of RT was determined from the level of extended products. Controls for the effectiveness of the trap and full extension of the substrate were also performed as described in the various figures. Values for K_d were determined by plotting the concentration of extended product (nM) versus the concentration of HIV-RT and fitting the data by nonlinear least square fit to the quadratic equation: $[ED] = 0.5([E]_t + [D]_t + K_d) - 0.5(([E]_t + [D]_t + K_d)^2 - 4[E]_t[D]_t)^{1/2}$, where $[E]_t$ is the total enzyme concentration and $[D]_t$ is the total primer-template concentration (15).

Sequences analysis of products recovered from rounds 10 and 12. Products selected from rounds 10 and 12 were sequenced by cloning into vector pBSM13+. A portion of the PCR products from each round were digested with HincII and EcoRI. Sites for these enzymes are present in the 5' and 3' flanking sequences, respectively (see Figure 1). The cleaved products were isolated from a polyacrylamide gel as described above and ligated into pBSM13+ that was previously cleaved with the same restriction enzymes, and then dephosphorylated with calf intestinal phosphatase. Ligation was performed using the Rapid ligation kit as described above. Ligated material was transformed into competent *E.coli* GC5 cells and DNA for

sequencing was prepared from bacterial colonies using a mini-prep kit.

Dissociation rate constant (k_{off}) determinations. Substrate (5 nM final concentration) and HIV-RT (15 nM for substrate 12-1 and 100 nM for substrate 1-1) were mixed in 64 μ l of buffer containing 50 mM Tris-HCl (pH = 8), 1 mM DTT, 80 mM KCl, 6 mM MgCl₂ and 0.1 μ g/ μ l BSA and incubated for 3 min at 37°C. Fifteen μ l of trap solution in the same buffer with heparin (1 μ g/ μ l final concentration) was added at time '0'. Aliquots of 5.25 μ l were removed at 10 s, 20 s, 40 s, 1 min, 2 min, 4 min, 8 min, 12 min and 16 min and added to a tube containing 1 μ l of dNTP (100 μ M final concentration) solution in the same buffer. Incubations were continued for 2 min then stopped with 2 \times loading buffer and subjected to electrophoresis on 12% polyacrylamide denaturing gels as described below. A time '0' sample was prepared by adding 4.25 μ l of the enzyme template mix to a tube with 1 μ l each of trap and dNTP mix. Off-rates were determined by plotting the amount of extended product (determined using a phosphorimager) versus time. A nonlinear least-squares fit of the data to an equation for single-exponential decay [$f(x) = ae^{-bx}$, where a is the y intercept at time 0 and b is the dissociation rate] was used to graph the data using Sigma Plot (Jandel Corp).

Preparation of ddG terminated 12-1 loop-back substrate. Fifty pm of 12-1 loop-back DNA (see Figure 5A) that was labeled at the 5' end with P-32 at very low specific activity was incubated with 5 U of Klenow polymerase in 50 μ l of buffer

containing 50 mM Tris-HCl (pH = 8), 1 mM DTT, 50 mM KCl, 6 mM MgCl₂ and 25 μM ddGTP for 30 min at 37°C. The material was extracted and precipitated then run through a sephadex spin column to remove any remaining ddGTP. Recovered material co-migrated with 12-1 loop-back DNA and was not extendable by HIV-RT in the presence of dNTPs indicated that it was 3' terminated with ddG.

Competition binding assay. Reactions contained 10 nM (final concentration) 5' P-32 end-labeled 12-1 loop-back substrate (see Figure 5) and 2 nM (final concentration) HIV-RT in 8 μl of buffer containing 50 mM Tris-HCl (pH = 8), 1 mM DTT, 80 mM KCl, 6 mM MgCl₂ and 0.1 μg/μl BSA. Various amounts (0, 5, 10, 20, 30, 40, 60, 80 or 100 nM) of cold competitor were also included in the reactions. Samples were incubated at room temperature for 1 h. Reactions were initiated by the addition of 2 μl dNTPs (100 μM final in reactions) and heparin 'trap' (1 μg/μl final in reactions) in the same buffer as above. After 2 min reactions were terminated with 10 μl of 2× loading buffer and subjected to electrophoresis on 12% polyacrylamide denaturing gels as described below. The amount of extended product was determined with a BioRad FX phosphoimager. A graph of relative extension (the sample with no competitor added was assigned a value of 1 and all other samples were relative to this) versus amount of competitor was constructed.

Gel electrophoresis. Six or 13% native polyacrylamide (29:1 w:w acrylamide:bisacrylamide) or 12% denaturing polyacrylamide (19:1 w:w acrylamide:bisacrylamide, 7 M urea) gels were prepared and subjected to electrophoresis using TBE buffer as described previously (13).

RESULTS

Approach used to select tight binding sequences

The basic approach used for sequence selection is illustrated in Figure 1 and described under Methods. The 30 nt length in the randomized region was chosen based on the ~30 nt footprint of HIV-RT on primer-templates (16,17). Therefore, contacts along the entire enzyme could be probed. With a 30 nt region and 500 pm of starting material, this calculates to about 1 in every 3800 possible sequences being represented. Despite this small proportion, most 'sequence motifs' are likely to be present unless they are very long. For example, all possible 20 nt runs would be present more than 200 times each. Typically enough RT to bind about 1/20th of the total recovered substrate was used in each round (see Methods). With 12 rounds of selection this would in theory be enough to select single molecules from the starting material. The choice to stop the selection after 12 rounds was determined based on the K_d values for the various pools. Values were (in nM, average of three experiments ± SD) 82 ± 14, 52 ± 15, 10 ± 3 and 12 ± 2, for the starting pool and rounds 6, 10 and 12, respectively. Since rounds 10 and 12 had similar values, it was assumed that no further gain in binding affinity would be made by additional rounds. Also, rounds 11 and 12 were performed using 150 mM KCl in the selection as compared with 80 mM for the first 10 rounds. The fact that no further decrease in K_d occurred despite the more stringent conditions further indicated that maximal binding was reached.

Round 12

1. AAACGAAGAGGCCACCGAAGGGGGGGTATT-3'
2. ~AAGAGATTAAGCGCTGAAGGGGGGGTATT-3'
3. TCACTTCATGCGCCTGGACGGGGGGTATT-3'
4. **T**CTGGGTCGTTAGACCGTAGGGGGGGTATC-3'
5. ~AAACGATATACGCGAAAGGGGGGGTATT-3'
6. ~GCCTGTGTAGTCGCCTAAGGGGGGGTATT-3'
7. TAACCTCGTAAGCCCCGAAGGGGGGGTATT-3'
8. ~AAGATCGCAGAGCATGAAGGGGGGGTATT-3'
9. ~GTCCGGCGAGTGCATAAGGGGGGGTATT-3'
10. NGAATCTTTAAGCCCAATGGGGGGGTATG-3'
11. ~CTCCGTGTTGTGCCTGAAGGGGGGGTATT-3'
12. ~AATTACAAACGTCCCTACGGGGGGTATT-3'
13. ~GAAATGTCTATGCCTAAAGGGGGGGTAAT-3'
14. ~CGAGTCCAGCGCCGAAGGGGGGGTATT-3'
15. NNCNNCCGCNATNGCTGAAGGGGGGGTATT-3'
16. AACATGCGTCGCTCCCTAAGGGGGGCAGCT-3'
17. TCGCATTANAAGCGCCTAAGGGGGGGTATT-3'
18. ~TATCCAAGTTACTCCCTAGGGGGGGTATT-3'

Round 10

1. ~TATCCAAGTTACTCCCTAGGGGGGGTATT-3'
2. ~GATGCAAGTAGTCCGCGACAGGGGGCAAC-3'
3. TCATGCCCTACGCCATGAGGGGGGCACAAG-3'
4. ~CCGACGACTTAACCTGAGGGGGGCAGCA-3'
5. **G**TAAAGTTTGTCCCTGAAGGGGGGCACAAG-3'
6. ~TATAGTCGAAGCGGGAATGGGGGGCATCT-3'
7. AACTAATTGAAGCCCTGCGGGGGGCACAAG-3'

Figure 2. Sequence of the material recovered from rounds 10 and 12. Sequences of the round 10 and 12 material in the 30 nt random region of the primer strand are shown. All sequences are aligned at the 3' end and the first six bases of the primer strand that would be present after cleavage with BbsI are underlined. Many recovered sequences were truncated by 1 nt and lost bases (~) are denoted at the 5' end. Two sequences had a 1 nt addition (shown in bold at the 5' end) resulting in a 31 base random region. 'N' denotes a nucleotide that was not clearly identified during sequencing.

Tight binding substrates all contain a run of 6–8 G residues at the primer 3' end

Material from the round 10 and 12 selections was recovered, cloned, and sequences as described in Methods. Shown in Figure 2 are sequences from 18 clones from round 12 and 7 from round 10. Only the sequence of the primer strand in the random region is shown. All the sequences contained the additional 20 nt at the 5' end derived from the PCR primer (see Figure 1). In theory, the primer strand should be 30 nt in the random region. The recovered products were 29–31 nt with some having one base additions or deletions in the random region, presumably resulting from PCR errors. Although the round 10 and 12 sequences were clearly related, only one was recovered in both pools (round 12 #18 and round 10 #1). None of the sequences from round 12 was identical; however, 17 had a 6–8 nt run of Gs at what would be the primer terminus after BbsI cleavage (underlined), while the other one (number 16) had a 3' terminal C residue preceded by a run of 6 Gs. There were also other common features including most sequences having a 2–3 nt A run preceding the Gs. The As were typically preceded by a G-C rich region. Since these substrates were derived by selection and PCR, common features contributing to tight binding, as well as other non-relevant common features might be expected. This was explored below.

The run of G residues is necessary and sufficient for tight binding

In order to determine what motif(s) of the recovered substrates was important for tight binding several changes to the sequences were made. The sequence labeled number 1 in Figure 2 was used as a model sequence because it retained the 30 nt length in the random region and represented a reasonable 'consensus' of all the recovered sequences (determined using BioEdit). This substrate is labeled 12-1 in Figure 3. Two random sequences recovered from the starting material were also used in these experiments (labeled 1-1 and 1-2). Because of the strong related nature of the round 12 sequences it was not deemed necessary to analyze several sequences.

The K_d values for each substrate in Figure 3A are listed in parentheses near the substrate. An autoradiogram from a typical K_d determination experiment with 1-1 and 12-1 is shown in Figure 3B and graphed in Figure 3C. The measured value for 12-1 was at least an order of magnitude lower than the values for 1-1 and 1-2. To determine why, several chimeric substrates were tested that included sequences derived from both 12-1 and 1-1 as well as other changes. In the '2 Gs' substrate the run of 7 G's in 12-1 was reduced to two at the 3' terminus with the other nucleotides of the run replaced by mostly 1-1 nt. This substrate bound comparably with 1-1 indicating that a run of Gs greater than two is required for tight binding. Substrates '12 Gs' and '26 Gs' increased the length of the G run in 12-1 to 12 and 26 nt, respectively. The latter represents a complete G-C primer-template sequence in the random region of the substrate. Both substrates bound at least as tightly as 12-1, with 26 Gs showing slightly tighter binding. These results indicate that the longer runs of Gs do not prevent tight binding and there may even be slightly tighter binding with very long runs. Two additional substrates were tested. In one (7 Gs template) the 7 Gs in the primer strand were switched into the template and the C's in the template switched to the primer. This substrate lost the ability to bind tightly indicating that the G residues must be in the primer strand for tight binding. Finally, a substrate that retained the 7 Gs from 12-1 but had all other nucleotides in the random region replaced by 1-1 nt was tested (7 Gs only). This substrate retained tight binding, although there was some increase in the average K_d value (from 5 ± 1 with 12-1 to 12 ± 7 with 7 Gs only). In this case, the increase was not great enough to be significant.

To further verify the tight binding observed with 12-1, an off-rate determination was performed using 1-1 and 12-1 (Figure 4). An autoradiogram is shown in Figure 4A and graphed in Figure 4B. The k_{off} values determined in this experiment were 0.0009/s and 0.0093/s for 12-1 and 1-1, respectively. This was consistent with the 10-fold lower K_d observed for 12-1. The result also indicates that the tighter binding was caused by a slower dissociation rate from the substrate rather than a change in the association rate. This experiment was repeated with similar results.

The affinity of MuLV-RT and TY-3 RT for 1-1 and 12-1 was also measured. MuLV-RT showed ~3-fold tighter binding to 12-1 compared with 1-1 while no difference was observed with TY-3 RT. Two drug resistant forms of HIV-RT, and D67N/K70R/T215F/K219Q and K65R (18,19), both showed a strong preference for 12-1 similar to wild-type RT (data not shown).

Overall, the results show that a run of several G residues at the 3' primer terminus on a DNA-DNA primer-template leads to tight binding by RT. The tight binding results from slow dissociation of RT from the substrate and the Gs must be in the primer strand.

A loop-back DNA but not RNA substrate designed based on 12-1 binds very tightly to RT and binding is not inhibited by a terminal dideoxy G residue

The above results demonstrated that a run of Gs, similar to that observed on the HIV ppt RNA, was responsible for the tight binding of the selected primer-template sequences. To determine whether an RNA version of the high affinity substrate could also bind tightly, a competition assay was developed. First, single-stranded 60 nt versions of 1-1 and 12-1 that formed a loop-back primer-templates were made (Figure 5). The loop-back substrates retained only the sequences in the random region plus an additional four A residues that formed the loop. Values for K_d s of 26 ± 12 and 2.2 ± 0.4 nM were measured for the 1-1 and 12-1 loop-backs, respectively. In each case, this was about 1/2 the value of the original substrate indicating slightly tighter binding to the loop-backs. There was an ~10-fold tighter binding for the 12-1 loop-back substrate compared with 1-1 loop-back, consistent with what was observed with the two stranded substrates.

Other loop-back substrates were also tested using a competition assay. In this assay, increasing amounts of cold competitor were added to a reaction containing a fixed amount of the radiolabeled 12-1 loop-back substrate. This mixture was incubated with a fixed amount of RT for 1 h, followed by addition of dNTPs and heparin trap as described in Methods. The presence of competitor reduces the amount of RT that bound to and extended the labeled 12-1 loop-back depending on the amount of competitor and its affinity for RT. The competitors used were the 1-1 and 12-1 loop-back substrates, an RNA version of the 12-1 loop-back and a version of 12-1 loop-back in which the 3' terminal G residue was a dideoxy nucleotide. The RNA substrate would have more potential as an RT inhibitor because it can be expressed in cells after transfection or infection with virus-based vectors as has been carried out with other RT aptamer inhibitors (4). An autoradiogram from a typical competition assay is shown in Figure 6A while a graph is presented in Figure 6B. As expected, cold 12-1 loop-back was able to effectively compete against radiolabeled 12-1 loop-back, leading to a decrease in the amount of the latter that was extended in the assay. The dideoxy version of 12-1 loop-back was also an effective competitor, indicating that the presence of a non-extendable G residue at the 3' terminal position does not significantly affect the affinity of RT for the substrates. In contrast, 1-1 loop-back and the RNA version of 12-1 loop-back (data not shown) showed no substantial inhibition of extension even when present at 10-fold greater concentration than 12-1 loop-back.

DISCUSSION

In this article, we show for the first time that specific primer-template sequences bind HIV-RT with greater affinity than random sequences. Tight binding in the recovered sequences resulted from a run of 6-8 G residues at the 3' primer terminus.

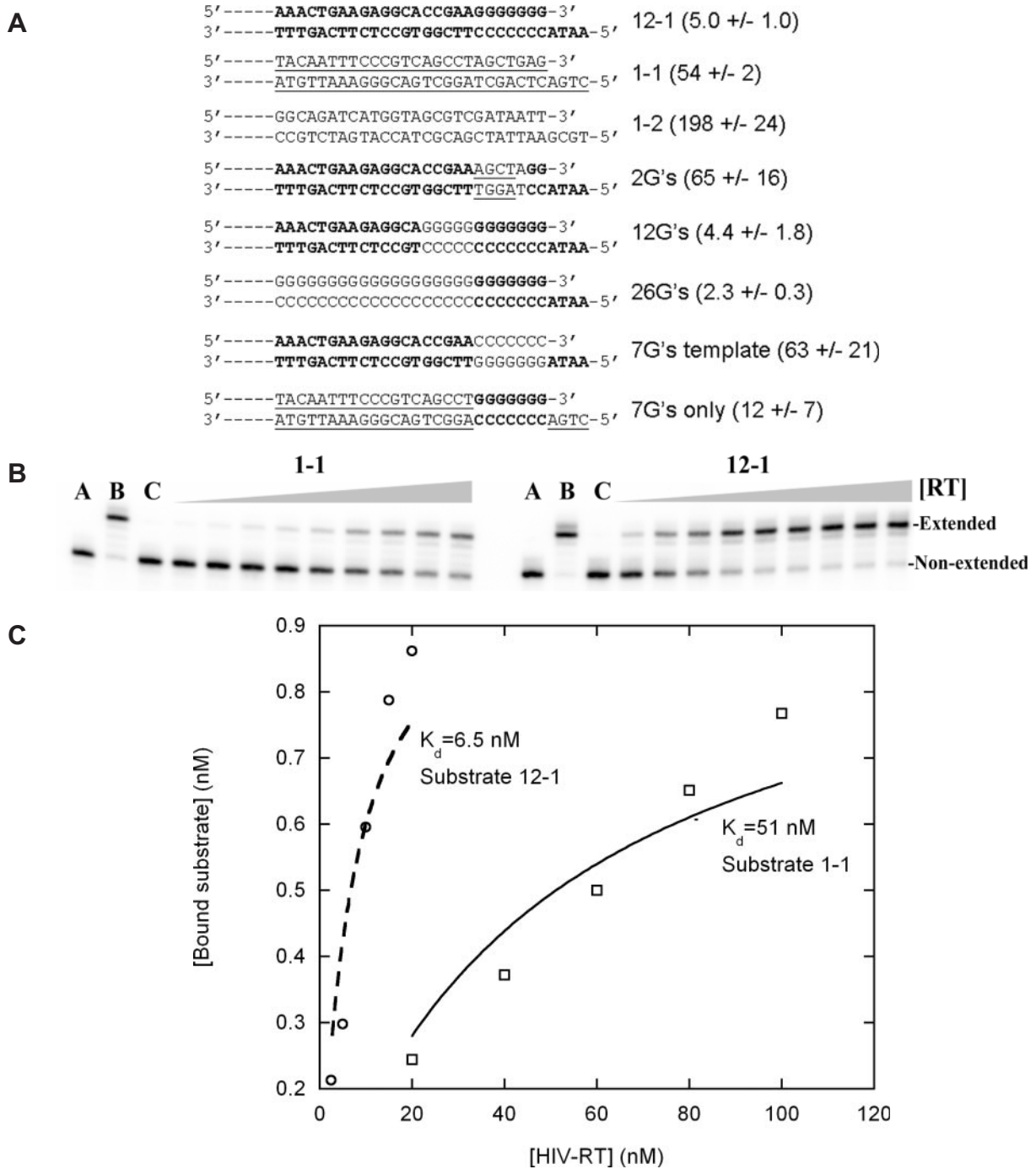


Figure 3. (A–C). Substrates measured for binding affinity to HIV-RT. (A) Shown are diagrams of 8 different substrates tested for affinity to HIV-RT. The 20 base pair fixed region is denoted by dashes (---) and was identical for all substrates (see Figure 1). Sequences corresponding to the random region from the first clone shown in Figure 2 (round 12 #1) are shown in bold in substrate 12-1 (bold letters). Substrates 1-1 (underlined nt) and 1-2 were randomly recovered from the starting material to serve as controls. Other substrates were as shown. When bases derived from 12-1 or 1-1 were used to construct chimeric substrates they are denoted in bold or underlined letters, respectively. The name and K_d value for binding to HIV-RT (in parentheses) is shown to the right of each substrate. Values are from an average of at least three experiments \pm SD. (B) An autoradiogram from a typical experiment using substrates 1-1 and 12-1 is shown. The substrate (1 nM) was 5' P-32 end-labeled on the primer strand and mixed with various amounts of HIV-RT as described in Methods. Reactions were initiated by the addition of dNTPs and heparin 'trap' to sequester RT molecules not bound to the substrate and those that dissociate. This limits extension to a single binding event between the substrate and enzyme. The concentrations of RT used were from left to right: 2.5, 5, 10, 15, 20, 40, 60, 80 and 100 nM. Extended and non-extended primer positions are indicated. The amount of bound enzyme at the various concentrations of RT was determined from the level of extended products using a phosphoimager. Lane A, no enzyme added; lane B, full extension of substrate in the absence of trap with 40 nM RT; lane C, RT (40 nM) was added to a reaction containing trap, dNTPs and substrate and incubated for 3 min (trap control). (C) Example of a graph used to determine the equilibrium dissociation constant (K_d). A plot of [Bound substrate] versus [HIV-RT] for an experiment with 1-1 and 12-1 is shown. The concentrations of RT used in the assays depended on RT's affinity for the particular substrate and were chosen to span the range of the approximate K_d value. For 12-1 RT concentrations were 1.3, 2.5, 5, 7.5 and 10 nM and for 1-1, 20, 40, 60, 80 and 100 nM were used. The plotted values were fitted to an equation for determining the K_d value (see Methods). The determined values for this particular experiment are listed on the graph.

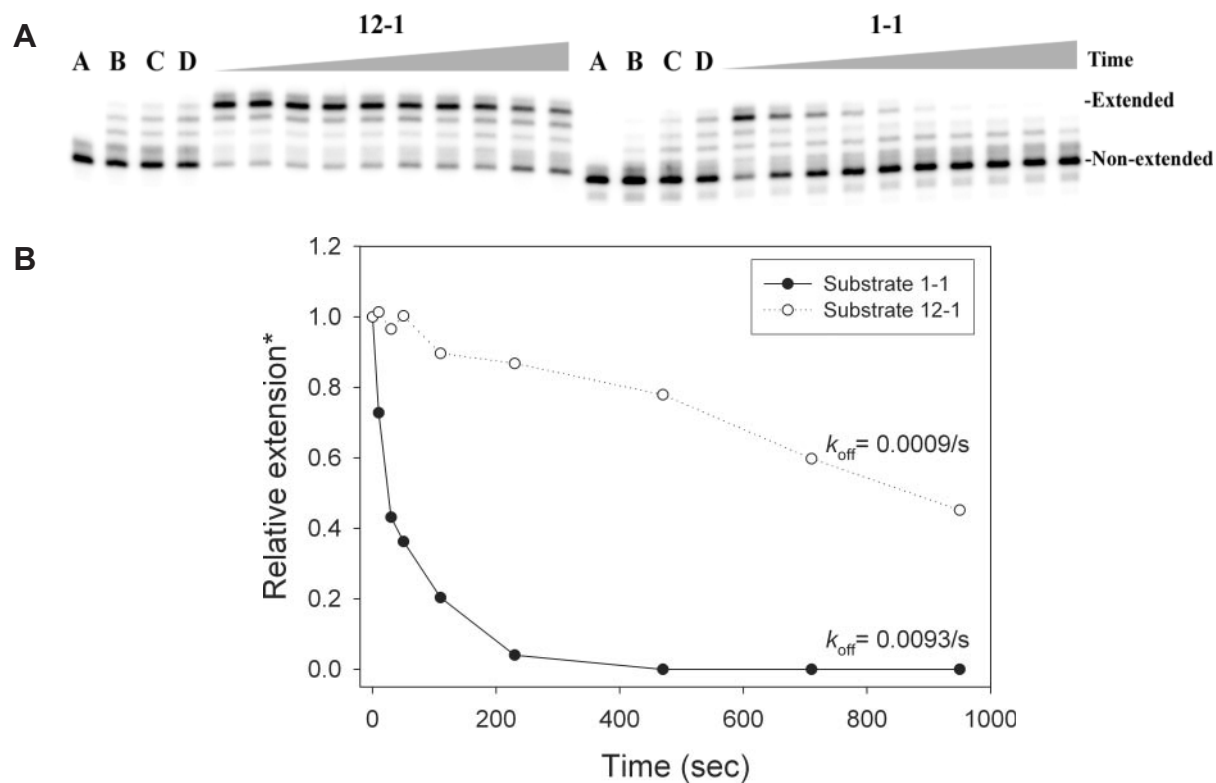


Figure 4. (A and B). Off-rate determination for substrates 12-1 and 1-1. (A) Substrates 1-1 or 12-1 were preincubated with 100 or 15 nM HIV-RT, respectively. A mixture containing heparin trap was added to the reactions and aliquots were removed at 10 s, 20 s, 40 s, 1 min, 2 min, 4 min, 8 min, 12 min and 16 min, mixed with a solution containing dNTPs and incubated for 3 min. Extended products represent the amount of RT remaining bound to the substrate at the various time points. For time 0 (reactions after lane D for each substrate), a mixture containing dNTPs and heparin trap was mixed with the RT-substrate preincubation solution. Lane A, no enzyme; lanes B–D, RT (100 or 15 nM for 1-1 or 12-1, respectively) was added to a reactions containing heparin trap, dNTPs, and substrate and incubated for 1 (A), 8 (B) or 16 (C) min (trap controls). (B) Plot of relative extension versus time for the experiment shown in (A). The amount of extended substrate for the time 0 point was assigned a value of 1 and all other points are relative to this. Off-rates (k_{off}) were determined by fitting the data to an equation for exponential decay (see Methods) and are listed on the graph for the two substrates. This experiment was repeated with similar results.

The exact length of the G run was not important as runs greater than 6–8 also bound tightly (see Figure 3). However, a minimum length was required as a substrate with a two base run bound comparable to random pool sequences. Including a single dideoxy G residue at the 3' primer terminus did not significantly affect the affinity of RT for the substrate (Figure 6).

The recovered sequences closely resembled the 3' end of the HIV ppt RNA which has a 6 base G run. Results suggest that HIV-RT uses the ppt to prime second strand synthesis not only because it is resistant to cleavage by RNase H, but also due to its unique structure which allows RT to bind tightly and in the proper orientation for nucleotide addition (20). With this in mind, it may actually be the structure rather than sequence of the primer-templates selected here that allowed them to bind so well. Sequences rich in G + C and containing G-tracts (stretched of G-C base pairs) are known to favor B-form to A-form transitions in solution (21–24) and crystal structures (25–27). Although HIV-RT can perform nucleotide catalysis on RNA–DNA (normally A'-form), DNA–DNA (normally B-form) and even RNA–RNA hybrids (normally A-form) in the case of tRNA extension, crystallographic analysis of several polymerases including RT indicates that the hybrid region immediately adjacent to the polymerase active site is in an

A'-form even for DNA–DNA substrates (28–31). The A'-form at the active site and that of RNA–DNA hybrids are not identical (32). It is also known that RT binds RNA–DNA hybrids much tighter than DNA–DNA (33–35). The propensity of G-C rich DNA hybrids to transition to the A-form suggests that it may require less distortion for these hybrids to conform to an A'-form that is catalytically competent. This could in turn lead to tighter binding.

Another possible explanation for the tight binding of the selected material is that the run of G residues could introduce a bend that helps the substrate fit into RT's active site, as is proposed to occur with the ppt (20). Crystal structures show that when bound to RT, an $\sim 45^\circ$ bend is observed in the primer-template (31). Although A-tracts are more typically associated with DNA bending, runs of G residues have also been shown to induce curvature to a lesser extent (36,37). This could help RT bind more tightly if the induced curvature favored a better fit to the active site. The curvature induced by A-tracts seems not to induce tight binding to RT as no tight binding primer-templates with long A runs were selected in these experiments. It is interesting in this regard that A- and G-tracts tend to induce curvature in the opposite directions with the former compressing the DNA minor groove and the latter the major (37).

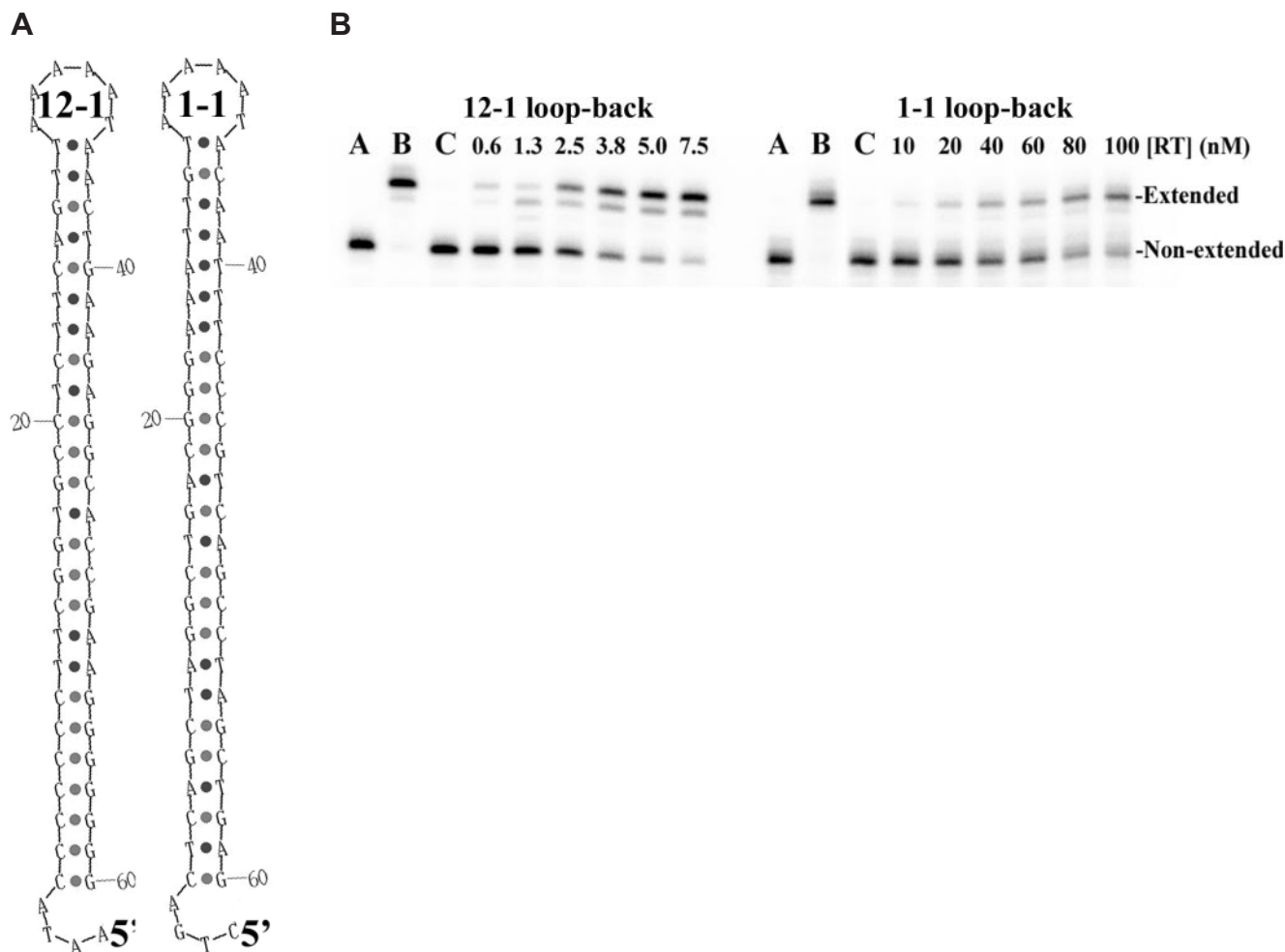


Figure 5. (A and B). K_d determination for 12-1 and 1-1 loop-back substrates. (A) Schematic diagrams of 12-1 and 1-1 loop-back substrates are shown. The 60 nt substrates are composed of only one strand of DNA and were derived from the original 12-1 and 1-1 substrates (see Figure 3) using only the nucleotides in the random region and four additional A residues in the loop. (B) An autoradiogram from a typical experiment used to determine the K_d value for substrates 1-1 and 12-1 loop-back. The amount of enzyme used in each reaction is listed above the individual lanes and was different for the two substrates. See Figure 3B for a description of the assay and markings. For 1-1 and 12-1 loop-back the K_d values were 26 ± 12 and 2.2 ± 0.4 nM (average of three experiments \pm SD).

An RNA version of the tight binding 12-1 loop-back substrate did not bind RT tightly as judged from the competition assay (data not shown). Two possible reasons for this are that the template region in this substrate is also RNA and that RT simply binds RNA primers relatively poorly. RT recognizes the ppt primer in the context of a DNA template and this may be pivotal for tight binding. Indeed, although our discussion has focused on the G run in the primer strand, the complementary deoxycytosine run in the template could be as important with respect to the substrate conforming for tight binding. As for binding to RNA primers, RT binds to the 5' end of most RNAs recessed on a longer DNA (primer configuration) and degrades rather than extending them (38,39). The ppt represents a special case where RT is directed to the 3' end for extension. The RNA version of 12-1 loop-back did not have the complete ppt sequence nor was the template region DNA. The substrate could be extended by RT but extension was inefficient compared with the DNA substrate (data not shown). Therefore, it was not clear that RT even bound preferentially to the 3' terminus on this substrate. Also, DNA versions of the ppt are actually more efficient primers than

the ppt (40). This suggests that even an optimized RNA primer does not function as well as DNA primers, although it is not clear whether this results from lower affinity for RT or inefficient nucleotide addition.

Reverse transcriptase from Moloney murine leukemia virus (MuLV) showed a preference for binding 12-1 over 1-1 while TY-3 RT did not. The latter finding showed that the G-rich primer-templates do not bind all RTs tightly but show some specificity. It is interesting that the ppt of MuLV closely resembles HIV with a 6 nt G run near the 3' terminus, while the TY-3 ppt has no such run. Again this suggests that the run of Gs may induce a structure that allows RT to bind tightly, but not all RTs have evolved to bind strongly to this motif. It would be appealing to see if a SELEX with TY-3 pulled out a sequence closely resembling its ppt or if MuLV-RT selected the same sequences as HIV-RT.

As was noted in the introduction, single-stranded DNA molecules that bind HIV-RT with high affinity have been selected using SELEX techniques (10). The selected ligands bound with affinities similar to the tight binding primer-templates found here. Among the selected ligands were several

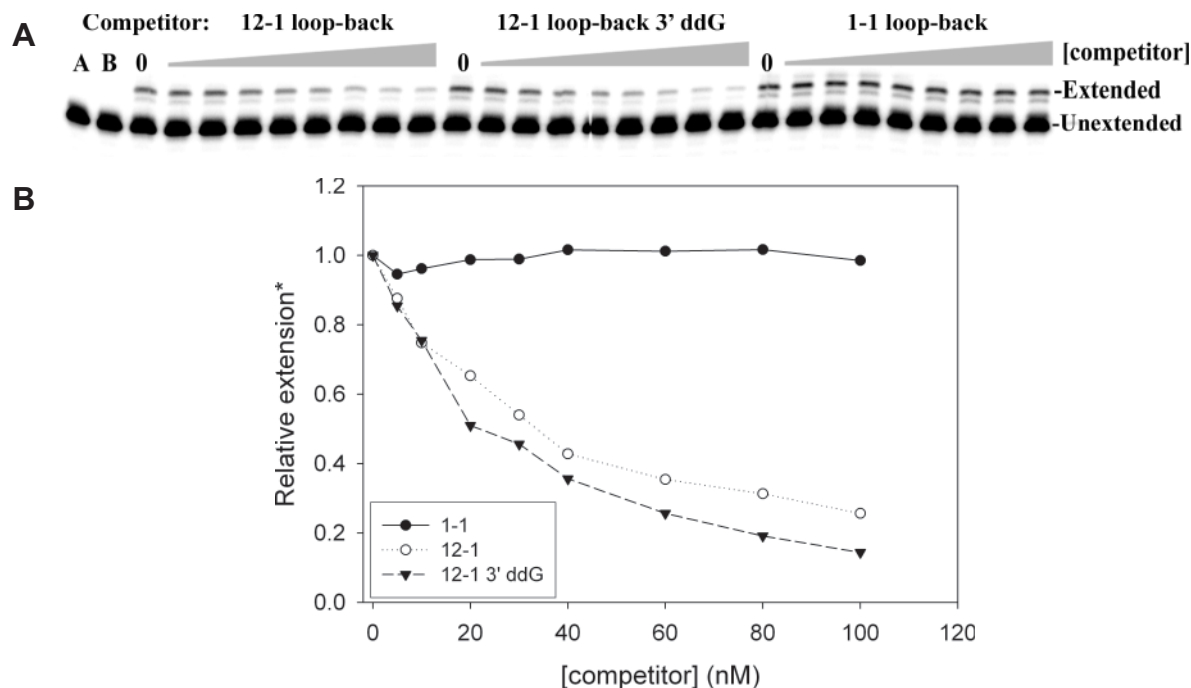


Figure 6. (A and B). Competition assay with 12-1 loop-back competed against itself or 1-1 or 12-1 3' ddG loop-backs. Shown is an autoradiogram of a competition assay using 5' P-32 end-labeled 12-1 loop-back substrate (10 nM). The substrate was incubated with HIV-RT (2 nM) and increasing amounts of cold competitor substrate (from left to right, 0, 5, 10, 20, 30, 40, 60, 80 and 100 nM) for 1 h. Extension was then initiated by the addition of dNTPs and heparin trap (see Figure 3) and incubation was continued for 2 min. Lane A, no enzyme added; lane B, trap control with 2 nM HIV-RT (see Figure 3). (B) A graph of relative extension versus concentration of competitor is shown for the experiment in (A). Values are relative to the reaction with no competitor added (labeled '0' in each set). The amount of extended substrate in this reaction was quantified using a phosphoimager and set equal to 1. This experiment was repeated with similar results.

predicted to fold and form 3' recessed termini that contained a run of G residues. Since the invariant region of these molecules contained a run of 4 Gs at the 3' terminus, this is not surprising. The authors concluded that the folded structure of the ligands was most important to binding RT with high affinity rather than particular sequences. In light of what we show here it would be interesting to see if the G run played a role in the tight binding for some of the ligands. It was clearly not the only important parameter as several tight binding ligands without this motif were also selected.

Overall, these results demonstrate that the sequence of a primer-template can play a role in determining how tightly it binds to RT. With HIV-RT the difference can be an order of magnitude or more. Whether this is a general phenomenon for all polymerases or is unique to some reverse transcriptases is not known. The resemblance of the sequences recovered here to the HIV ppt suggests that tight binding may result from evolutionary pressure to recognize a particular structure induced by the sequence. If this were the case, other polymerases would probably not show such a striking range of affinities.

ACKNOWLEDGEMENTS

The authors thank Dr Stuart Le Grice from HIV Drug Resistance Program, NCI-Frederick for the RNase H minus E478>Q HIV-RT and TY-3 RT, and Dr Michael Parniak, University of Pittsburgh for HIV-RT drug resistant mutants K65R and D67N/K70R/T215F/K219Q. This work was supported by National Institute of General Medicine grant

number GM051140. Funding to pay the Open Access publication charges for this article was provided by the grant GM051140.

Conflict of interest statement. None declared.

REFERENCES

- Ellington, A.D. and Szostak, J.W. (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature*, **346**, 818–822.
- Joyce, G.F. (1989) Amplification, mutation and selection of catalytic RNA. *Gene*, **82**, 83–87.
- Tuerk, C. and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, **249**, 505–510.
- Joshi, P. and Prasad, V.R. (2002) Potent inhibition of human immunodeficiency virus type 1 replication by template analog reverse transcriptase inhibitors derived by SELEX (systematic evolution of ligands by exponential enrichment). *J. Virol.*, **76**, 6545–6557.
- Joshi, P.J., North, T.W. and Prasad, V.R. (2005) Aptamers directed to HIV-1 reverse transcriptase display greater efficacy over small hairpin RNAs targeted to viral RNA in blocking HIV-1 replication. *Mol. Ther.*, **11**, 677–686.
- Joshi, P.J., Fisher, T.S. and Prasad, V.R. (2003) Anti-HIV inhibitors based on nucleic acids: emergence of aptamers as potent antivirals. *Curr. Drug Targets Infect. Disord.*, **3**, 383–400.
- Tuerk, C., MacDougall, S. and Gold, L. (1992) RNA pseudoknots that inhibit human immunodeficiency virus type 1 reverse transcriptase. *Proc. Natl. Acad. Sci. USA*, **89**, 6988–6992.
- Jaeger, J., Restle, T. and Steitz, T.A. (1998) The structure of HIV-1 reverse transcriptase complexed with an RNA pseudoknot inhibitor. *EMBO J.*, **17**, 4535–4542.
- Chen, H. and Gold, L. (1994) Selection of high-affinity RNA ligands to reverse transcriptase: inhibition of cDNA synthesis and RNase H activity. *Biochemistry*, **33**, 8746–8756.

10. Schneider,D.J., Feigon,J., Hostomsky,Z. and Gold,L. (1995) High-affinity ssDNA inhibitors of the reverse transcriptase of type 1 human immunodeficiency virus. *Biochemistry*, **34**, 9599–9610.
11. Brody,E.N. and Gold,L. (2000) Aptamers as therapeutic and diagnostic agents. *J. Biotechnol.*, **74**, 5–13.
12. Gold,L. (1995) The SELEX process: a surprising source of therapeutic and diagnostic compounds. *Harvey Lect.*, **91**, 47–57.
13. Sambrook,J. and Russell,D.W. (2001) *Molecular Cloning: A Laboratory Manual, 3rd edn*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
14. Peliska,J.A. and Benkovic,S.J. (1992) Mechanism of DNA strand transfer reactions catalyzed by HIV-1 reverse transcriptase. *Science*, **258**, 1112–1118.
15. Hsieh,J.C., Zinnen,S. and Modrich,P. (1993) Kinetic mechanism of the DNA-dependent DNA polymerase activity of human immunodeficiency virus reverse transcriptase. *J. Biol. Chem.*, **268**, 24607–24613.
16. Wöhrl,B.M., Tantillo,C., Arnold,E. and Le Grice,S.F. (1995) An expanded model of replicating human immunodeficiency virus reverse transcriptase. *Biochemistry*, **34**, 5343–5356.
17. DeStefano,J.J., Cristofaro,J.V., Derebail,S., Bohlayer,W.P. and Fitzgerald-Heath,M.J. (2001) Physical mapping of HIV reverse transcriptase to the 5' end of RNA primers. *J. Biol. Chem.*, **276**, 32515–32521.
18. Arion,D., Kaushik,N., McCormick,S., Borkow,G. and Parniak,M.A. (1998) Phenotypic mechanism of HIV-1 resistance to 3'-azido-3'-deoxythymidine (AZT): increased polymerization processivity and enhanced sensitivity to pyrophosphate of the mutant viral reverse transcriptase. *Biochemistry*, **37**, 15908–15917.
19. Arion,D., Borkow,G., Gu,Z., Wainberg,M.A. and Parniak,M.A. (1996) The K65R mutation confers increased DNA polymerase processivity to HIV-1 reverse transcriptase. *J. Biol. Chem.*, **271**, 19860–19864.
20. 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.
21. Ivanov,V.I., Minchenkova,L.E., Minyat,E.E., Frank-Kamenetskii,M.D. and Schyolkina,A.K. (1974) The B to A transition of DNA in solution. *J. Mol. Biol.*, **87**, 817–833.
22. Minchenkova,L.E., Schyolkina,A.K., Chernov,B.K. and Ivanov,V.I. (1986) CC/GG contacts facilitate the B to A transition of DNA in solution. *J. Biomol. Struct. Dyn.*, **4**, 463–476.
23. Zavriev,S.K., Minchenkova,L.E., Frank-Kamenetskii,M.D. and Ivanov,V.I. (1978) On the flexibility of the boundaries between the A-form and B-form sections in DNA molecule. *Nucleic Acids Res.*, **5**, 2657–2663.
24. Pilet,J., Rottman,F. and Brahms,J. (1973) Structural parameters of single and double stranded helical polyribonucleotides. *Biochem. Biophys. Res. Commun.*, **52**, 517–523.
25. Haran,T.E., Shakked,Z., Wang,A.H. and Rich,A. (1987) The crystal structure of d(CCCCGGGG): a new A-form variant with an extended backbone conformation. *J. Biomol. Struct. Dyn.*, **5**, 199–217.
26. Wang,A.H., Fujii,S., van Boom,J.H. and Rich,A. (1982) Molecular structure of the octamer d(G-G-C-C-G-G-C-C): modified A-DNA. *Proc. Natl Acad. Sci. USA*, **79**, 3968–3972.
27. Mooers,B.H., Schroth,G.P., Baxter,W.W. and Ho,P.S. (1995) Alternating and non-alternating dG–dC hexanucleotides crystallize as canonical A-DNA. *J. Mol. Biol.*, **249**, 772–784.
28. Kunkel,T.A. and Wilson,S.H. (1998) DNA polymerases on the move. *Nature Struct. Biol.*, **5**, 95–99.
29. Brautigam,C.A. and Steitz,T.A. (1998) Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes. *Curr. Opin. Struct. Biol.*, **8**, 54–63.
30. 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 [see comments]. *Science*, **282**, 1669–1675.
31. 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 Å resolution shows bent DNA. *Proc. Natl Acad. Sci. USA*, **90**, 6320–6324.
32. Wöhrl,B.M., Krebs,R., Goody,R.S. and Restle,T. (1999) Refined model for primer/template binding by HIV-1 reverse transcriptase: pre-steady-state kinetic analyses of primer/template binding and nucleotide incorporation events distinguish between different binding modes depending on the nature of the nucleic acid substrate. *J. Mol. Biol.*, **292**, 333–344.
33. Cristofaro,J.V., Rausch,J.W., Le Grice,S.F. and DeStefano,J.J. (2002) Mutations in the ribonuclease H active site of HIV-RT reveal a role for this site in stabilizing enzyme-primer-template binding. *Biochemistry*, **41**, 10968–10975.
34. DeStefano,J.J., Bambara,R.A. and Fay,P.J. (1993) Parameters that influence the binding of human immunodeficiency virus reverse transcriptase to nucleic acid structures. *Biochemistry*, **32**, 6908–6915.
35. Yu,H. and Goodman,M.F. (1992) Comparison of HIV-1 and avian myeloblastosis virus reverse transcriptase fidelity on RNA and DNA templates. *J. Biol. Chem.*, **267**, 10888–10896.
36. Biburger,M., Niederweis,M. and Hillen,W. (1994) Oligo[d(C)(G)] runs exhibit a helical repeat of 11.1 bp in solution and cause slight DNA curvature when properly phased. *Nucleic Acids Res.*, **22**, 1562–1566.
37. Merling,A., Sagaydakova,N. and Haran,T.E. (2003) A-tract polarity dominates the curvature in flanking sequences. *Biochemistry*, **42**, 4978–4984.
38. DeStefano,J.J., Mallaber,L.M., Fay,P.J. and Bambara,R.A. (1993) Determinants of the RNase H cleavage specificity of human immunodeficiency virus reverse transcriptase. *Nucleic Acids Res.*, **21**, 4330–4338.
39. DeStefano,J.J. (1995) The orientation of binding of human immunodeficiency virus reverse transcriptase on nucleic acid hybrids. *Nucleic Acids Res.*, **23**, 3901–3908.
40. Fuentes,G.M., Rodriguez-Rodriguez,L., Fay,P.J. and Bambara,R.A. (1995) Use of an oligoribonucleotide containing the polypurine tract sequence as a primer by HIV reverse transcriptase. *J. Biol. Chem.*, **270**, 28169–28176.