

Selection of RNA aptamers that bind HIV-1 LTR DNA duplexes: strand invaders

Chatchawan Srisawat¹ and David R. Engelke^{2,*}

¹Department of Biochemistry, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkoknoi, Bangkok 10700, Thailand and ²Department of Biological Chemistry, University of Michigan, 3200 MSRB III, 1150 W. Medical Center Drive, Ann Arbor, MI 48109-0606, USA

Received March 14, 2010; Revised July 16, 2010; Accepted July 22, 2010

ABSTRACT

RNA that can specifically bind to double-stranded DNA is of interest because it might be used as a means to regulate transcription of the target genes. To explore possible interactions between RNA and duplex DNA, we selected for RNA aptamers that can bind to the long terminal repeats (LTRs) of human immunodeficiency virus type 1 DNA. The selected aptamers were classified into four major groups based on the consensus sequences, which were found to locate in the non-stem regions of the predicted RNA secondary structures, consistent with roles in target binding. Analysis of the aptamer consensus sequences suggested that the conserved segments could form duplexes via Watson–Crick base-pairing with preferred sequences in one strand of the DNA, assuming the aptamer invaded the duplex. The aptamer binding sites on the LTR were experimentally determined to be located preferentially at these sites near the termini of double-stranded target DNA, despite selection schemes that were designed to minimize preferences for termini. The results presented here show that aptamer RNAs can be selected *in vitro* that strand-invade at preferred DNA duplex sequences to form stable complexes.

INTRODUCTION

Oligonucleotides that can bind specifically to the double-stranded (ds) DNA are of major interest due to their possible use as molecular tools for gene manipulations. Previous studies have shown that single-stranded (ss) DNA, RNA or synthetic nucleic acids [e.g. peptide nucleic acid (PNA)] can specifically recognize and bind

to the dsDNA targets through various interactions such as triple helix formation or strand invasion (1–7). As a result of binding, these dsDNA-directed oligonucleotides can prevent other trans-acting factors such as transcription factors from interacting with the dsDNA targets, leading to suppression/activation of transcription or inhibition of DNA replication (8–11). In addition, they can be modified to contain DNA-reactive agents or nucleic-acid motifs capable of recruiting other proteins, which then lead to modulation of the processes involving the target genes, e.g. expression, repair, recombination or mutagenesis (12–16). These properties make them very appealing as artificial gene modulators with therapeutic potential (17–20).

Among the nucleic acids, RNA is particularly interesting for such roles because it can be expressed from recombinant DNA constructs inside cells, making it suitable for intracellular use such as in gene therapy. RNA has previously been shown to specifically recognize and interact with dsDNA. One of the binding interactions is through a triple helix formation, in which the ssRNA inserts itself into the major groove of DNA and forms a triple-helical complex (21,22). This interaction has also been shown to affect the transcription of the target DNA (23–25). However, there are some limitations for this interaction. First, triple helix formation between dsDNA and ssRNA requires a long stretch of polypurine/polypyrimidine tract on the duplex DNA as a binding site for the RNA third strand. This would exclude a number of genes as potential targets because they might not contain such sequences. Moreover, some of the reported RNA–DNA triplex formation needs acidic conditions for a stable binding (21,26), restricting the use of ssRNA-directed triplexes *in vivo*.

To find RNA that can recognize and bind to duplex DNA targets at approximately physiological pH and ionic strength in a sequence-specific manner, *in vitro* aptamer selection, or SELEX (Systematic Evolution of Ligands by EXponential enrichment), techniques

*To whom correspondence should be addressed. Tel: +1 734 763 0641; Fax: +1 734 763 7799; Email: engelke@umich.edu

have been used. This approach can isolate high-affinity oligonucleotides capable of binding specifically to a particular ligand from a large pool ($>10^{15}$ sequences) of random-sequence ss oligonucleotides (27–30). In this study, the 3' long terminal repeat (LTR) of human immunodeficiency virus type 1 (HIV-1) was used as a target because this region is important for virus gene expression (31) and the LTR-binding RNA ligands (aptamers) thus generated might be useful for the suppression of HIV proliferation. Using this approach, it was expected that novel ssRNA–dsDNA interactions might be discovered and studied, leading to a better understanding that would allow a rational design of artificial RNA-based gene regulators.

MATERIALS AND METHODS

Materials

DNA oligonucleotides were synthesized by the University of Michigan DNA core facility. Streptavidin–agarose beads were from GIBCO BRL. Taq DNA polymerase, T7 RNA polymerase and Sequenase were prepared in house (32–34). Superscript II was from Invitrogen.

Generation of the random sequence RNA library

The RNA library was generated as previously described (35,36). Briefly, the library template was synthesized from 0.2 μ mol of ss deoxyoligonucleotides containing 40-nt randomized central sequence flanked by defined primer-binding sites with the sequence of 5'-AGTAATACGACTCACTATAGGGAGTCGACCGACCAGAA (N40) TATGTGCGTCTACATCTAGACTCAT-3'. The ssDNA was converted into dsDNA by a primer extension reaction using Taq DNA polymerase with 0.2 μ mol of the 3' N40 primer (5'-ATGAGTCTAGATGTAGACGCACATA-3'). The dsDNAs, which contains a T7 RNA polymerase promoter at the 5' end, were used as templates of an *in vitro* transcription reaction to generate an RNA library; the resulting RNA sequences were 5'-GGGAGUCGACCGACCAGAA [N40] UAUGUGCGUCUACAU CUAGACUCAU-3' (84 nt) with a calculated library complexity of $\sim 7 \times 10^{16}$ different RNA sequences. The *in vitro* transcription reaction was then extracted twice with phenol/chloroform, precipitated with ethanol, and resuspended in 20 ml of 50 mM Tris pH 8.0, 200 mM NaCl. The RNAs were separated from leftover nucleotides and primers using size exclusion chromatography on Sephadex G-100.

Preparation of dsDNA targets

The DNA fragments of the 3' LTR were generated from a plasmid containing HIV-1 genome, pBenn7 (37,38), kindly provided by Dr Gary J. Nabel. The LTR-325 and LTR-408 bp (Figure 1) corresponding to the positions 9261–9585 and 9219–9626 of the 3' LTR (accession number AF324493), respectively, were generated by PCR amplification with the following primers; 5' LTR-325 bp 5'-GGAGAGAACACCAGCTTG-3', 3' LTR-325 bp 5'-GTTCCCTAGTTAGCCAGAG-3', 5' LTR-408 bp

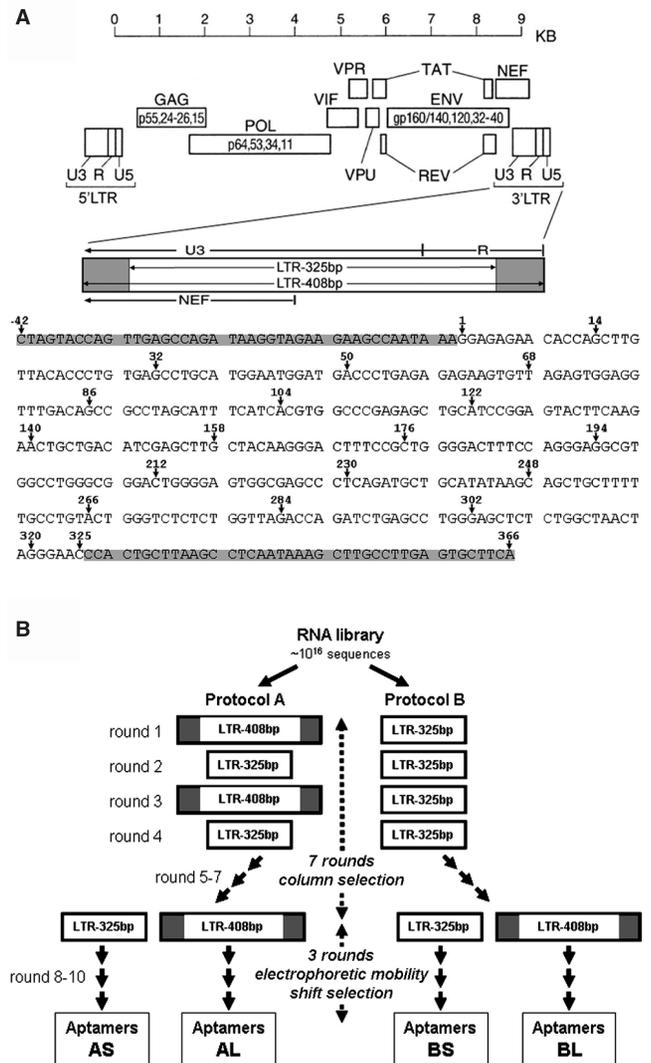


Figure 1. The dsDNA targets and the selection schemes used for the aptamer selection. (A) The intended DNA target is a 325-bp dsDNA segment of the HIV-1 3' LTR, which spans the U3 and R regions (LTR-325 bp), corresponding to the positions +1 to +325. The other target is a 408-bp DNA fragment (LTR-408 bp), which has about 40 bp of the LTR sequences (shaded boxes) extending from both ends of LTR-325 bp and corresponds to the positions –42 to +366. (B) Diagram of the selection schemes used to obtain the RNA specific to dsDNA.

5'-CTAGTACCAGTTGAGCCAGATAA-3' 3' LTR-408 bp 5'-TGAAGCACTCAAGGCAAGCTT-3'. The biotinylated fragments were generated using the primers containing biotin at the 5' ends through non-cleavable linkers (BioTEG). The PCR products were gel-isolated in 8% polyacrylamide gel and electroeluted. The purified LTRs were extracted with phenol/chloroform, precipitated in ethanol and resuspended in the binding buffer (50 mM HEPES pH 7.4, 10 mM MgCl₂ and 100 mM NaCl).

In vitro selection of duplex DNA-binding RNA ligands

The RNA ligands or aptamers capable of binding to the LTRs were selected from the RNA library for a total of 10 rounds (Figure 1) of binding, amplification of bound

RNAs and re-selection. During the first seven rounds, the biotinylated LTRs were immobilized on streptavidin-agarose beads and used as the targets for selection. RNA (0.5–40 nmol or 1–25 μ M) was incubated with the biotinylated LTR-bound streptavidin beads containing 0.025–0.4 nmol or 0.05–0.2 μ M of LTRs in the binding buffer (50 mM HEPES pH 7.4, 10 mM MgCl₂ and 100 mM NaCl). The binding reaction was done in a column at room temperature for 1 h. The beads were then washed and the RNAs were eluted from the beads with elution buffer (8 M urea, 5 mM EDTA). The eluted RNAs were amplified by reverse transcription polymerase chain reaction (RT-PCR) and *in vitro* transcription reactions to generate RNAs for the next round of selection as previously described (35,36). Subtraction steps with agarose beads were included, starting from Round 2, to remove RNAs having affinity toward the bead matrices instead of the DNA targets. In the last three rounds of selection (Rounds 8–10), the LTR-binding aptamers were enriched using electrophoretic mobility shift. Briefly, the radiolabeled RNA (5–10 pmol) was incubated with 5 μ g (20 pmol) of unmodified LTRs in the binding buffer at room temperature for 1 h. Non-specific DNA fragments (50 μ g) prepared from salmon sperm DNA cut with AluI were also added to the binding reaction as competitor DNAs to remove RNAs that might have non-specific affinity toward DNAs. The samples were then loaded into a non-denaturing 6% polyacrylamide in the binding buffer, and run at 60 V for 4 h. The shifted band corresponding to the DNA : RNA complexes were excised. The RNAs were then eluted with 8 M urea and undergone RT-PCR and *in vitro* transcription reactions. The PCR products from the final round were cloned into pGEM-T vector (Promega). Individual aptamers were characterized for the affinity to the LTRs by electrophoretic mobility shift assays and sequenced by dideoxy chain termination methods. Their sequences were analyzed and grouped according to the consensus sequences. The secondary structures were predicted using a secondary-structure prediction software, RNAstructure version 4.4 (39,40) and they were drawn using an RNA secondary structure visualization program, RnaViz 1.0 (41).

Electrophoretic mobility shift assay

The radiolabeled aptamer (5000 cpm or 0.1 pmol) was incubated with 1 μ g (5 pmol) of LTR-325 bp, 1.25 μ g (5 pmol) of LTR-408 bp or 5 μ g of non-specific DNA fragments prepared by combining various plasmids (e.g. pUC9, pBR322) and digested with MspI. The reaction was done in the binding buffer (5 μ l total volume) and incubated for 1 h at room temperature. The sample was then loaded into a non-denaturing 6% polyacrylamide in the binding buffer and run at 60 V for 4 h. The gels were dried and radioactive bands were visualized using a PhosphorImager (Molecular Dynamics). The assay was also used to test the binding of aptamers to various LTR fragments, i.e. the 5' fragments of 250 and 292 bp, and the 3' fragments of 116 and 75 bp, generated from PvuII-cleaved LTR-325 and LTR-408 bp, respectively.

To study the affinity of the aptamer toward the LTR targets, the radiolabeled aptamer (20 nM) was incubated with different concentrations of either LTR-325 or LTR-408 bp (0–10 μ M) and the electrophoretic mobility shift assays were done as described above. The apparent dissociation constant was then determined as described by Soukup *et al.* (22).

Linker-scanning analysis of the aptamer binding sites

When Dnase I and S1 nuclease footprinting of the RNA-DNA interactions failed to give definitive results, we used substitution mutants in the LTR to define the binding sites. Various LTR-325 bp fragments containing 18 bp of the NdeI-XhoI-SalI linker sequences sequentially replacing the wild-type sequence from positions +1 to +325 were generated by PCR amplification. Most of the fragments were amplified from the plasmids containing the linker-scanning mutants of HIV-1 LTR (42). Some linker-scanning LTR fragments, i.e. those containing the linker sequences at +16 to +33, +302 to +320 and +321 to +325, were generated by PCR-directed mutagenesis. The PCR fragments were purified using polyacrylamide gel electrophoresis and the electrophoretic mobility shift assays of the aptamers and the linker-scanning LTR fragments were done as described above. It is noted that the linker-scanning mutants used here were derived from the HIV-1 strain HBX2 (42), the LTR sequence of which is different from that of the HIV-1 strain used in this selection at the following positions: +42A→G, +54T→G, +109G→A and +236G→C (Figure 1A). We found no significant difference in the aptamer binding to the 'wild type' LTR fragments derived from these two strains.

RESULTS

In vitro selection of the aptamers specific to the dsDNAs containing the HIV-1 3' LTR sequence

The aim of this study is to evaluate the possibility of generating RNA aptamers capable of specific recognition of dsDNA. In this study, the 3' LTR of HIV-1 was chosen as the target. It is 325 -bp long spanning the regions of U3 and R regions (Figure 1A), thus called LTR-325 bp. The *in vitro* selection, or SELEX (28), technique was used to obtain RNA that can bind specifically to LTR-325 bp. The binding was done in the binding buffer containing 50 mM HEPES pH 7.4, 10 mM MgCl₂ and 100 mM NaCl. In our preliminary selections, the aptamers had a tendency to bind to the ends of dsDNA, probably due to transient strand separation near the DNA ends and DNA : RNA duplex formation. To avoid this problem and to promote the binding of RNA toward the internal region of dsDNA, several selection schemes were used (Figure 1B). The first strategy (protocol A) used alternating dsDNA targets, i.e. LTR-325 and LTR-408 bp. LTR-408 bp has extra sequences of ~40 bp extending from both 5' and 3' ends of the LTR-325 bp. These two fragment targets were alternated during the first seven rounds of column selection in order to avoid enrichment of the undesired aptamers with preferential binding to the dsDNA ends of LTR-325 bp. The subsequent three rounds of

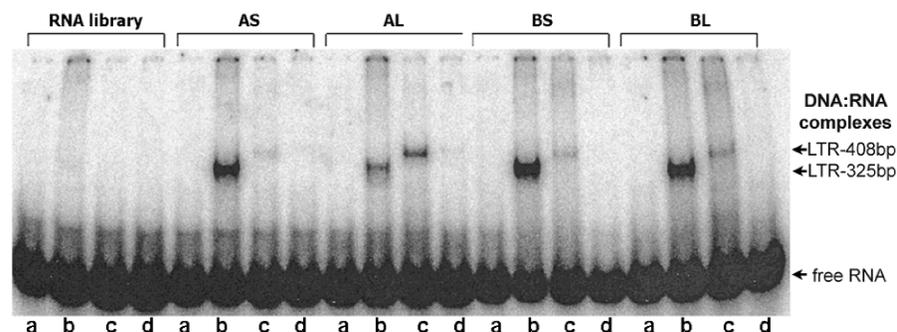


Figure 2. Electrophoretic mobility shift assays of the round 10 aptamers pools AS, AL, BS and BL to the double-stranded LTR targets compared with that of the RNA library. Lanes a–d correspond to RNA only, RNA + LTR-325 bp, RNA + LTR-408 bp and RNA + non-specific dsDNA fragments, respectively. The arrows indicates the position of free RNA, and the complexes between RNA and either LTR-325 bp or LTR-408 bp.

electrophoretic mobility shift selections used either the short LTR-325 bp or long LTR-408 bp DNA fragments. The resulting aptamer pools were designated as aptamers AS and AL, respectively. In contrast, the second strategy (protocol B) used only LTR-325 bp as the target throughout the column selection. Then LTR-325 bp was used again as the target during the electrophoretic mobility shift selections to generate the aptamer BS or the target was shifted to LTR-408 bp to generate the aptamer BL in order to enrich the aptamers that favor the binding to the internal region, not to the ends, of the target DNAs.

After a total of 10 rounds of selection, the binding of the aptamers to the target dsDNA fragments was assessed using electrophoretic mobility shift assays (Figure 2). The starting pool of RNA from the library showed no detectable affinity to any dsDNA targets, i.e. LTR-325 and LTR-408 bp, or non-specific dsDNA fragments derived from AluI-cut salmon sperm DNA. In contrast, the aptamer pools showed specific binding to the LTR fragments used in the selection, but not to the non-specific dsDNA. Interestingly, the aptamers bound the two LTR fragments differently, depending on the selection schemes. For example, the aptamers AL, which were generated from the scheme that strongly favored the binding to both LTR-325 and LTR-408 bp, bound comparatively well to both fragments, whereas the BS aptamers, which was selected against LTR-325 bp for the entire selection, showed stronger binding to LTR-325 bp, compared with LTR-408 bp.

Cloning and characterizing the dsDNA-binding aptamers

Individual aptamers from different selection schemes were cloned and their sequences and binding affinities to the LTR fragments were determined. Based on the consensus sequences, the aptamers could be classified into four major groups, i.e. Groups 1–4, and an unclassified group, in which the individual aptamers do not have any shared sequence (Figure 3A–D). In the predicted secondary structures of most aptamer groups, the consensus sequences were found to be in the non-base-pairing regions, suggesting that the ‘ss’ regions have probable roles in specific recognition of targets. The number of aptamer clones screened, frequencies and binding characteristics of the aptamers from various groups are summarized in Table 1.

It is interesting to note that the aptamers within each group share not only the consensus sequences, but also the binding characteristics. Most of the unclassified aptamers showed weak or no binding to the LTR targets in the mobility shift assays. The aptamers from Groups 1 and 2 aptamers bound strongly to LTR-325 bp but weakly, if not at all, to LTR-408 bp, whereas the Group-3 aptamers bound almost exclusively to LTR-408 bp. From their binding characteristics, the Groups 1–3 aptamers seem to bind to the ends of the target dsDNA. In contrast, the Group-4 aptamers bound comparatively well to both LTR-325 and LTR-408 bp. This gave us an initial impression that their binding site probably resided within the internal region of the dsDNA target, which is the desired property for the study (however, this proved not to be the case based on the results of the subsequent experiments below).

We, therefore, focused on the properties of the Group 4 members. To assess the binding strengths of the aptamers toward the LTR fragments, the apparent dissociation constants (K_d) of the aptamers were determined by electrophoretic mobility shift assays. The apparent K_d values of a representative aptamer from Group 4 (clone AL66) to LTR-325 and LTR-408 bp were 1.2 and 0.3 μ M, respectively (Figure 3E). The binding of Group-1 and -2 aptamers to LTR-325 bp in this assay (data not shown) displayed strengths in the same approximate range (0.1–2 μ M).

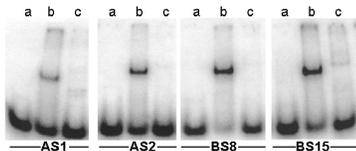
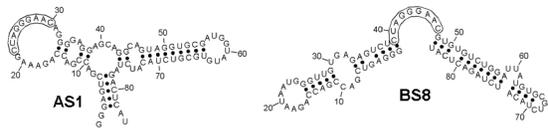
Determination of the aptamer binding sites on the LTR targets

To approximate the binding sites of the various aptamer classes, LTR-325 and LTR-408 bp were cleaved with a blunt-end restriction enzyme, PvuII, to generate the 5' fragments of 250 and 292 bp and the 3' fragments of 116 and 75 bp, respectively (Figure 4). These fragments were used as the targets of the representative aptamers from each group in electrophoretic mobility shift assays. Both Group-1 and -2 aptamers were previously shown to bind strongly to the LTR-325 bp. In this experiment, the binding sites of the aptamers from Groups 1 and 2 were located to the 75-bp 3' fragment and the 250-bp 5' fragment of LTR-325 bp, respectively. In contrast, the Group-4 aptamers, which bind to both LTRs, surprisingly

A Group 1

```

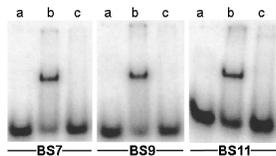
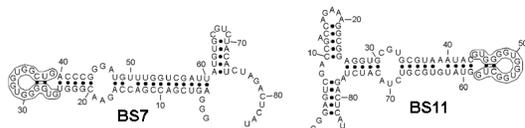
AS1          ACUAGGGAAACAGGGGAGGAGCAGGCGAGUAGGUGCGAUGG
AS2          GCAGGUCAGAUAAACAGCUAGGGAAACUGAGGCUAGG
AS6          ACAGAUUGUCCGGUAGCAGCUUAGGAAACGGUGCU
AS12         AUGGCGUGUGAUUUGGGCUAGGGAAUUAUGGGUGCGUGG
AL68         UUGGGAGCCUUGCGUGGUCUAGUACUAGGGAACGAGUC
BS3          GGGAUUUUGGGGUAACAUCUUAACUAGGGAAACAAAU
BS4          UUGGGAGCCUUGCGGUGGUCUAGUACUAGGGAACGAGUC
BS8          UAAUGGGUGUGAGAGUCUAGGGAAACGUGUGUCUGGAU
BS15         GGGUAGGAUAAAUAACUAGGGAAACAGUGGGGUGAGGUG
BS14         CGGAGGUGGCGUUGAUUACUAGGGAAACAGUGGUCUU
BL2          UUGGGAGCCUUGCGGUGGUCUAGUACUAGGGAACGAGUC
BL6          UUGGGAGCCUUGCGGUGGUCUAGUACUAGGGAAACGAGUC
Consensus sequence CUAGGGAAAC
    
```



B Group 2

```

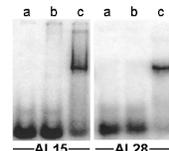
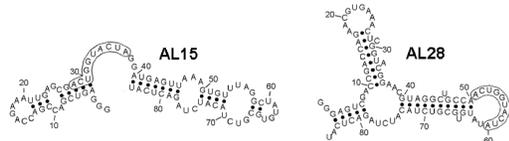
BS7          CGGGUGGGGUGGUGGUCGACCCGGGAUGUUGGUCGAU
BS9          GGUUGGGGUGGUGGUGAACCUGGUAAGGUUAGACCGG
BS10         UGCUGGGGUGGUGGUCGACCGUUGUUGUAACCCCG
BS11 AGGCGGAGGUGCGUGGUAUACUGGGGUGGUGGUCG
Consensus sequence GUGGGUGGUGGUCG
    
```



C Group 3

```

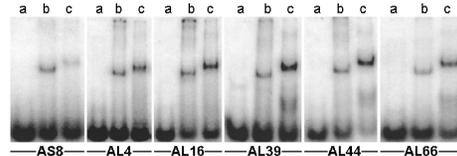
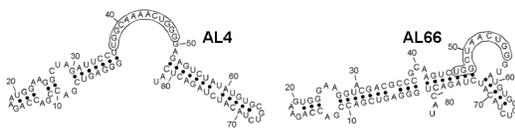
AL15         AUUGAGGCACUGGUACUAGGAGAGUUAAGUGUUUAGGC
AL28         CGUGAAACUCGUGACGGAAACGUGAGCCACUGGUACUA
Consensus sequence ACUGGUACUA
    
```



D Group 4

```

AS8          GUGACUUAACGAGUCGGAUCGUAUCCUGGCU-AACUGGAAUA
AS15         CUGAAGGCAGAGCUGGGGUCGAGUCCUACUGGGGAUA
AL4          UGGAAAGCUGAUAUCCUUGGCAAAACUGGGGAGAGUCUA
AL16         GGGUGAAGACUUCGCGAAACUGGGGAAUUGGGUUAGU
AL20         AAAUUGGUUCUCUGGGGUCUACUGGGGAAUUGGUGGUC
AL23         AUGGUAUUGGAAAGACUCUAGGGGGGAGUCCUACUGG
AL27         GGGGGGUGGACUUGGCUAAC-GGUACGAAUAGAGUCUUAAG
AL39         GUTGGUUUUGGCUUUGCGAAACUUGGGGUGAGGCUACGUG
AL40         GUAUGGUAUTUCGUAUAACUAAUUGCGAAACUGGGUAC
AL41         AUGAGAGGUGUAGUCCGAAACUUGGGGAAUGANNAGGGUG
AL44         AUAUGUUGUUCGGAUCCUGGCAAAACUGGGGAGAGUCUA
AL57         UUGGUAUGACUUCUGGCAAAACAGGUAUCAGAGUCUAAGG
AL60         GGCUCUUGCGAGACUCUUGGGGGGCUAGAGUAGGGA
AL64         GUUAUUCUGAUGCGUGACUCUUGGUGAACAUGGGAGA
AL66         GUGGGAAGGUACAGCGCCGCAAGUCUUGGCU-AACUGGG
Consensus sequence UGGCNNAACUGGG
    
```



E Dissociation constant determination

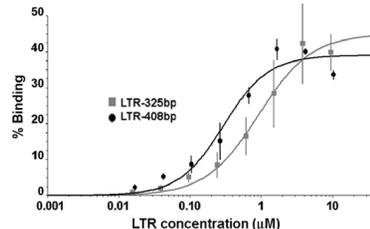
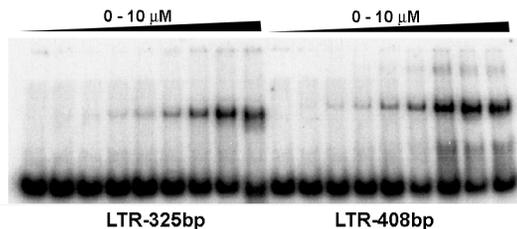


Figure 3. The sequences, predicted secondary structures (39,40), and binding characteristics of the aptamers. The aptamers are classified into four major groups, i.e. group 1–4, based on the consensus sequences [(A)–(D), respectively]. Only the randomized sequence of each aptamer is shown with the consensus sequence underlined. The consensus sequences (shown in boxes) are frequently located within the non-base-pairing regions of the predicted secondary structures. Each aptamer group also displays different binding affinity to the different double-stranded LTR targets as shown in electrophoretic mobility shift assays (Lanes a–c correspond to RNA only, RNA + LTR-325 and RNA + LTR-408 bp, respectively). (E) The apparent dissociation constants (K_d) of a representative aptamer from group 4 (clone AL66) were determined by electrophoretic mobility shift assays; the plot shows % binding as a mean \pm S.E. The K_d to LTR-325 and LTR-408 bp are 1.2 and 0.3 μ M, respectively.

do not seem to bind at the same binding site on LTR-325 and LTR-408 bp. The representative aptamer bound most strongly to the 75-bp 3' fragment of LTR-325 and to the 292-bp 5' fragment of LTR-408 bp. These two fragments are not overlapping, suggesting that the Group-4 aptamers have at least two separate binding sites for the

different LTR targets. The Group-4 aptamer binding sites could be deduced from this experiment; the first one is from positions –42 to –1 and the other is from +251 to +325 (Figure 1), respectively.

To refine the above findings, a detailed mapping of the aptamer binding sites on the LTR targets was performed.

Table 1. The frequencies and binding characteristics of the aptamers derived from different selection schemes in electrophoretic mobility shift assays

Selection schemes	No. of aptamers screened	Group 1 bind to LTR-325 bp (%)	Group 2 bind to LTR-325 bp (%)	Group 3 bind to LTR-408 bp (%)	Group 4 bind to both LTRs (%)	Unclassified (no consensus sequences)			
						bind to LTR-325 bp (%)	bind to LTR-408 bp (%)	bind to both LTRs (%)	No or weak binding (%)
AS	15	27	–	–	13	20	–	–	40
AL	54	2	–	4	22	4	2	7	59
BS	15	33	27	–	–	13	–	–	27
BL	15	13	–	–	–	13	–	–	74

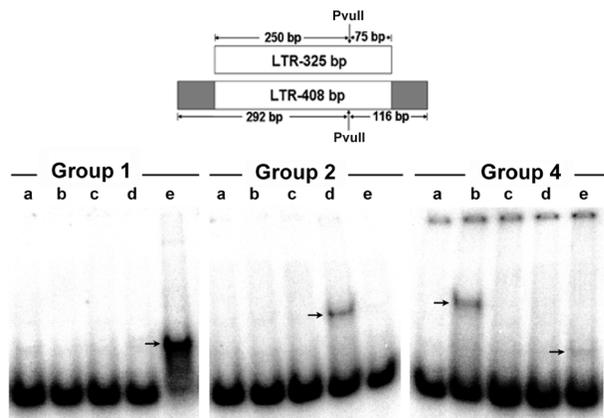


Figure 4. Determination of the aptamer binding sites using PvuII-cleaved LTR fragments. The double-stranded LTR-325 and LTR-408 bp were cleaved with a restriction enzyme, PvuII, into smaller fragments, which were used as targets in electrophoretic mobility shift assays; Lane a corresponds to RNA only, Lanes b and c to RNA + either the 5' or 3' fragment of LTR-408 bp (292- and 116-bp fragments, respectively), Lanes d and e correspond to RNA + either the 5' or 3' fragment of LTR-325 bp (250- and 75-bp fragments, respectively). The arrows indicate the complexes between the aptamer and the cleaved DNA fragments.

Initial attempts to assess the binding sites through 'foot-printing' of DNA cleavage accessibility did not give definitive results, so identification of the binding sites was performed through characterization of the effects of 'linker scanning' mutations. PCR fragments of the LTR-325 bp mutants, which have the NdeI-XhoI-SalI linker sequences (18 bp) sequentially replacing the wild-type sequence from positions +1 to +325 were generated from previously characterized clones (22). The mutants are named 'n/m' where n and m are the 5'- and 3'-most positions, respectively, where the wild-type sequence is replaced with the linker sequence. These mutants were used as the targets for the aptamers in electrophoretic mobility shifts assays. The binding sites of the aptamers could then be deduced from the loss of aptamer binding to the mutants containing the putative binding sites. As shown in Figure 5, the representative aptamer from Group 1 lost its binding to the LTR mutants 302/319 and 320/325, suggesting that the binding site of Group-1 aptamers are located at the 3' end of LTR-325 bp,

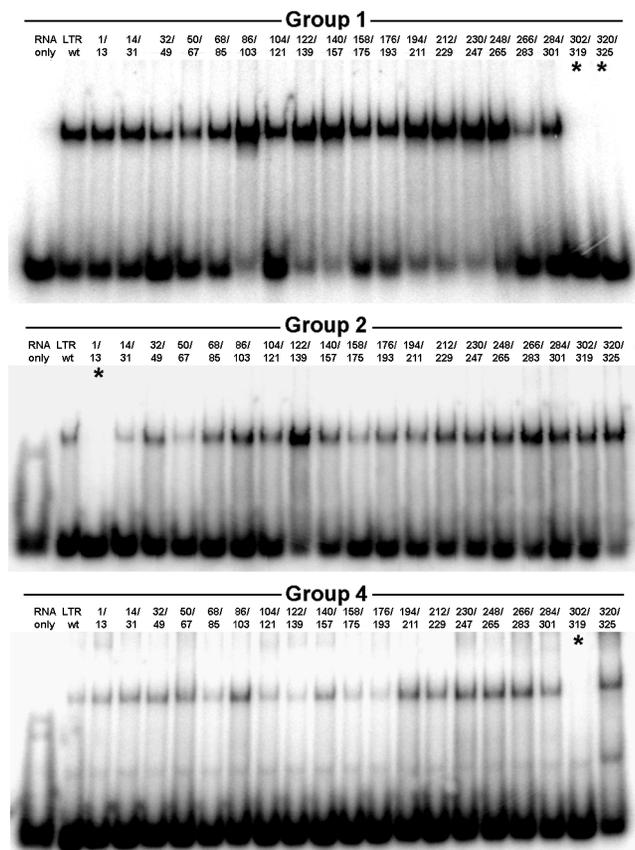


Figure 5. Determination of the aptamer binding sites on LTR-325 bp using linker-scanning LTR mutants. A set of the double-stranded LTR-325 bp mutants that have the NdeI-XhoI-SalI linker sequences (18 bp) sequentially replacing the wild-type sequence from positions +1 to +325 was used as the targets for the aptamers in electrophoretic mobility shifts assays; the numbers in each lane indicate the 5'-most and 3'-most positions, where the wild-type sequence is replaced with the linker sequence. Asterisks indicate the linker-scanning LTR mutants that show very little or no binding with the aptamers.

corresponding to the positions +302 to +325. In contrast, that of the Group-2 aptamer was located at the 5' end of LTR-325 bp, corresponding to the positions +1 to +13. For the Group-4 aptamers, which appeared to have two separate binding sites on LTR-325 and LTR-408 bp, its binding site on LTR-325 bp were

located near the 3' end, corresponding to the positions +302 to +319. The results from this experiment are in agreement with those of the experiment using the PvuII-cleaved LTR fragments for the binding site determination.

Analysis of the aptamer/target binding interactions

Because the consensus sequences of most aptamers are located within the non-base-pairing regions of the predicted secondary structures (Figure 3A–D), it was hypothesized that the aptamers bind to the LTR targets through the consensus sequences. Therefore, the possible interactions between the LTR sequences at the putative binding sites (from the above experiments) and the consensus sequences were speculated. For the Group-1 aptamers, which bind strongly to LTR-325 bp, their consensus sequence (5'-CUAGGGAAC-3') could potentially form a perfect duplex with the bottom strand of LTR-325 bp at positions +317 to +325 using canonical base-pairing (Figure 6). This is consistent with the binding site determined using the linker-scanning experiment, which is in the region of positions +302 to +325. A similar finding is observed with the Group-3 aptamers,

which bind only to LTR-408 bp. All 10 nucleotides of the Group-3 consensus sequence (5'-AUCAUGGUC A-3') could potentially form perfect canonical base-pairing with the top strand at the 5' end of LTR-408 bp, spanning the positions -41 to -32. For the Group-4 aptamers, which have two separate binding sites on LTR-325 and LTR-408 bp, their consensus sequence (5'-UGGCNNAACUGGG-3') is predicted to form complementary base-pairing with both LTR targets in spite of some degree of mismatches and the use of non-canonical base-pairing. One binding site (Site 1) is at the 5' end of LTR-408 bp at positions -36 to -25 (top strand), while the other site (Site 2) is at the 3' end of LTR-325 bp at positions +310 to +321 (bottom strand); this is in agreement with the results from the previous experiments (Figures 4 and 5), which defined the Group-4 aptamer binding sites on LTR-408 and LTR-325 bp at positions -42 to -1 and +302 to +319, respectively. However, for the Group-2 aptamers, which were shown to bind to the 5' end of LTR-325 bp at positions +1 to +13 (Figures 4 and 5), the base complementarity between the consensus sequence (5'-GUGGGGUGGUGGCUG-3') and the putative target sequence is not as obvious as those of the

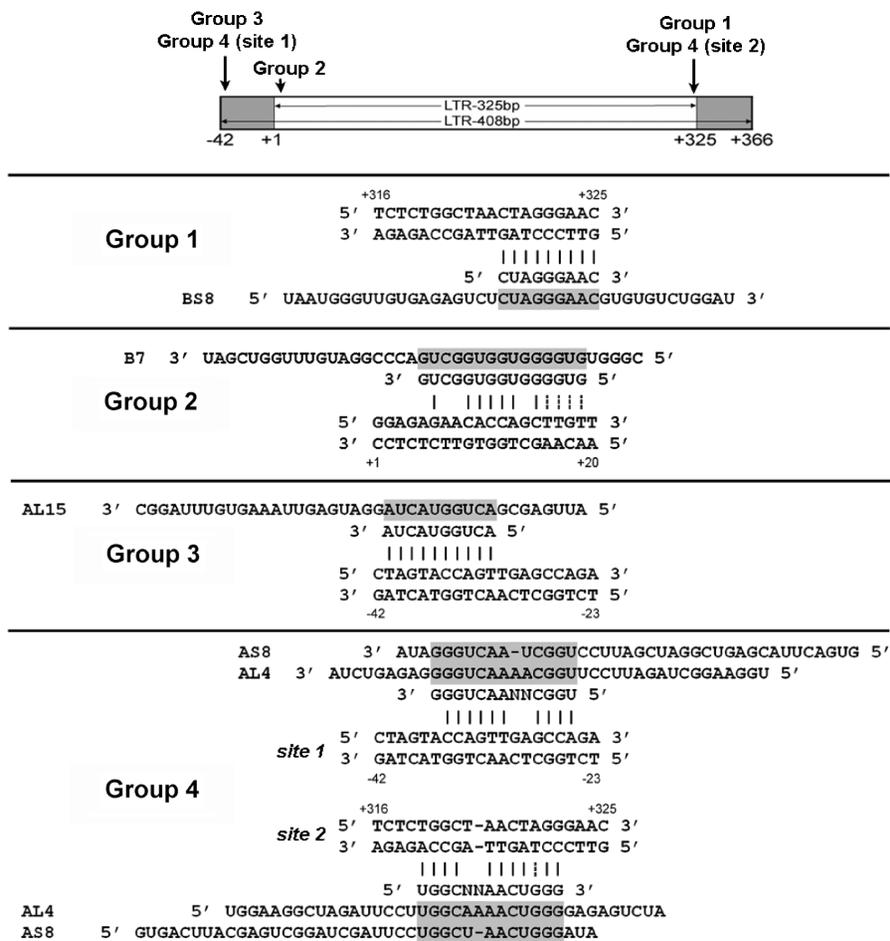


Figure 6. The putative binding sites of various aptamer groups on LTR-325 and LTR-408 bp. The representative aptamer and consensus sequence from each group are aligned with the target sequence(s) on the LTRs. Solid and dashed lines indicate the canonical and non-canonical base-pairings, respectively.

other aptamer groups (Figure 6). Their consensus sequence (15-nt long) is quite G rich (%G ~67), and only 11 nucleotides (out of 15) are predicted to form base-pairing with the putative target sequence, i.e. the top strand at the 5' end of LTR-325 bp, from positions +6 to +19; the binding thereof was confirmed by electrophoretic mobility shift assays (data not shown). Non-canonical base-pairing predictions (4 out of 11) are also more frequently observed in this group.

DISCUSSION

RNA that can specifically bind to dsDNA can be very useful as a means to regulate transcription of the target DNA. However, some of previously described RNA–dsDNA interactions still have some limitations (21,26). Therefore, in this study, we aimed to explore novel interactions between RNA and dsDNA at approximately neutral pH and modest ionic strength. The 3' LTR of HIV-1 was chosen as the target in this study. This target DNA does not contain a long stretch of polypurine/polypyrimidine tract; the longest stretch is only 10 bp, which is unlikely to favor a stable triplex formation (21,22). Thus, novel interactions other than triplex could be expected.

SELEX was used to discover RNA aptamers capable of binding to the dsDNA fragments of the 3' LTR. Four major groups of the LTR-specific aptamers were isolated, i.e. Groups 1–4. Interestingly, these aptamers tend to bind to the termini of the duplex DNA, probably through the invasion of the DNA strands followed by base-pairing formation between the target and aptamer consensus sequences, resulting in RNA : DNA duplexes. This end invasion is also supported from the findings that the RNA aptamers, particularly from Groups 1 and 2, no longer bind to LTR-408 bp, although their target sequences are present within this longer LTR fragment, but not at the termini (Figure 3A–B). Based on this approach, we found that binding of aptamers to the ends of the target DNAs seems to be a major mode of interactions as the previous data from two independent aptamer selections using only LTR-325 bp as a target also produced similar results; the end-binding aptamers with the consensus sequences exactly the same as those of Groups 1 and 2 were isolated. Because of this, different schemes were devised to minimize the end-binding RNAs by using alternating short and long LTR fragments that have different sequences at the dsDNA ends as targets (Figure 1B). However, the RNA aptamers obtained still bind to the ends of both target DNAs. In addition to the Group-1 and -2 aptamers isolated before, the selection schemes led to the isolation of the Group-4 aptamers, which can bind to the termini of both LTRs. They seem to do this by adopting the use of non-canonical base-pairing to form duplexes with the target sequences with some degree of tolerance to base mismatches. This is in contrast to the aptamers that bind exclusively to only one target (either LTR-325 or LTR-408 bp) such as those from Groups 1 and 3; only canonical base-pairing is used and there is no mismatch

between the target and aptamer consensus sequences. Of the end-binding aptamers, those from Group 2 are quite interesting. Unlike the other end binders, complementarity between the aptamer consensus sequence of Group 2 and the target DNA sequence is not continuous and seems to have more non-canonical base-pairing and mismatches. Due to the G-rich nature of these Group-2 aptamers, it might be possible that the aptamers might adopt non-standard structures such as a quadruplex or G-quartet to mediate the binding with the target DNA. Nevertheless, the actual nature of the binding interaction is still not clear and might be worth further investigation.

There are several explanations as to why the RNA aptamers prefer to bind to the end of DNA instead of the sequence in the internal region. Because the aptamers recognize the dsDNA through base-pairing with the target sequences, separation of DNA strands is required and the base pairs at the DNA termini separate more readily than the base pairs in the internal region of DNA (43–45); this makes the DNA termini the preferred sites for strand invasion by the aptamers. In addition, if the binding occurs within the DNA target, two helix boundaries will form at the edges of RNA–DNA strand-invasion complexes and the energetic cost of such boundary formation is unacceptably high (46). Therefore, the aptamer binding to the sequences at the DNA ends is more favorable because no new helix boundary needs to be formed. Recently, there is a study of oligonucleotides with similar binding characteristics to the LTR-binding aptamers, i.e. short PNAs with mixed-base composition (46). These mixed-base PNAs invade the DNA strands of the targets at the termini to form PNA : DNA duplexes. However, the end invasion is efficient only at the very end of the DNA termini; no binding is observed if the PNA-binding site is placed just two nucleotides away from the end and no mismatch is tolerated. On the other hand, the aptamers, particularly from Group 4, can bind to the LTRs even though their binding sites are situated up to 4 bp away from the ends and they can tolerate some mismatches. Furthermore, the aptamers can bind to the dsDNA targets at physiological pH and ionic strength, while the mixed-base PNAs do so only at low salt conditions; higher salt (>20 mM) inhibits the binding. It might be possible that other motifs in the aptamer structures besides the consensus sequence might help stabilize the binding as well.

Although the dsDNA-binding aptamers isolated in this study can bind to the target dsDNA by strand invasion and formation of RNA : DNA duplexes, their binding strengths are somewhat moderate (apparent K_d in a low micromolar range). Thus, they might or might not be effective *in vivo* as therapeutic tools. Moreover, their potential aptamer targets (i.e. the sequences at the dsDNA ends) might not be common inside cells except in some instances, e.g. during DNA replication, where a number of DNA ends are generated, or during transcription, where the DNA strands are transiently separated in a transcription bubble favoring strand invasion; such conditions would favor the aptamer binding to dsDNA. In addition, the presence of topoisomerases or supercoiled structures of DNA *in vivo* could reduce the resulting DNA

strain generated during the strand separation and might theoretically favor the aptamer binding to the internal DNA sequences. These could make the aptamers more useful as gene-directed therapeutic agents or molecular tools for genetic analysis or gene manipulations.

ACKNOWLEDGEMENTS

We would like to thank Dr Gary J. Nabel for a kind gift of the plasmid containing HIV1 genome. We also thank Dr James C. Alwine for providing the plasmids containing the linker-scanning mutants of HIV-1 LTR.

FUNDING

National Institutes of Health grants AI054239 and GM082875 (to D.R.E.). Funding for open access charge: University of Michigan, Department of Biological chemistry account.

Conflict of interest statement. None declared.

REFERENCES

- Demidov, V.V. and Frank-Kamenetskii, M.D. (2001) Sequence-specific targeting of duplex DNA by peptide nucleic acids via triplex strand invasion. *Methods*, **23**, 108–122.
- Duca, M., Vekhoff, P., Oussedik, K., Halby, L. and Arimondo, P.B. (2008) The triple helix: 50 years later, the outcome. *Nucleic Acids Res.*, **36**, 5123–5138.
- Frank-Kamenetskii, M.D. and Mirkin, S.M. (1995) Triplex DNA structures. *Annu. Rev. Biochem.*, **64**, 65–95.
- Grunweller, A. and Hartmann, R.K. (2007) Locked nucleic acid oligonucleotides: the next generation of antisense agents? *BioDrugs*, **21**, 235–243.
- Koizumi, M. (2006) ENA oligonucleotides as therapeutics. *Curr. Opin. Mol. Ther.*, **8**, 144–149.
- Lohse, J., Dahl, O. and Nielsen, P.E. (1999) Double duplex invasion by peptide nucleic acid: a general principle for sequence-specific targeting of double-stranded DNA. *Proc. Natl Acad. Sci. USA*, **96**, 11804–11808.
- Nielsen, P.E. and Egholm, M. (2001) Strand displacement recognition of mixed adenine-cytosine sequences in double stranded DNA by thymine-guanine PNA (peptide nucleic acid). *Bioorg. Med. Chem.*, **9**, 2429–2434.
- Diviacco, S., Rapozzi, V., Xodo, L., Helene, C., Quadrioglio, F. and Giovannangeli, C. (2001) Site-directed inhibition of DNA replication by triple helix formation. *FASEB J.*, **15**, 2660–2668.
- Guntaka, R.V., Varma, B.R. and Weber, K.T. (2003) Triplex-forming oligonucleotides as modulators of gene expression. *Int. J. Biochem. Cell Biol.*, **35**, 22–31.
- Shen, C., Rattat, D., Buck, A., Mehrke, G., Polat, B., Ribbert, H., Schirrmester, H., Mahren, B., Matuschek, C. and Reske, S.N. (2003) Targeting bcl-2 by triplex-forming oligonucleotide—a promising carrier for gene-radiotherapy. *Cancer Biother. Radiopharm.*, **18**, 17–26.
- Wang, G. and Xu, X.S. (2004) Peptide nucleic acid (PNA) binding-mediated gene regulation. *Cell. Res.*, **14**, 111–116.
- Chin, J.Y. and Glazer, P.M. (2009) Repair of DNA lesions associated with triplex-forming oligonucleotides. *Mol. Carcinog.*, **48**, 389–399.
- Rogers, F.A., Vasquez, K.M., Egholm, M. and Glazer, P.M. (2002) Site-directed recombination via bifunctional PNA-DNA conjugates. *Proc. Natl Acad. Sci. USA*, **99**, 16695–16700.
- Simon, P., Cannata, F., Concordet, J.P. and Giovannangeli, C. (2008) Targeting DNA with triplex-forming oligonucleotides to modify gene sequence. *Biochimie*, **90**, 1109–1116.
- Suzuki, T. (2008) Targeted gene modification by oligonucleotides and small DNA fragments in eukaryotes. *Front. Biosci.*, **13**, 737–744.
- Torigoe, H., Katayama, T. and Tsukamoto, Y. (2007) Triplex formation-based artificial transcription factor to regulate target gene expression. *Nucleosides Nucleotides Nucleic Acids*, **26**, 1463–1466.
- Christensen, L.A., Finch, R.A., Booker, A.J. and Vasquez, K.M. (2006) Targeting oncogenes to improve breast cancer chemotherapy. *Cancer Res.*, **66**, 4089–4094.
- Jain, A., Wang, G. and Vasquez, K.M. (2008) DNA triple helices: biological consequences and therapeutic potential. *Biochimie*, **90**, 1117–1130.
- Ray, A. and Norden, B. (2000) Peptide nucleic acid (PNA): its medical and biotechnical applications and promise for the future. *FASEB J.*, **14**, 1041–1060.
- Uil, T.G., Haisma, H.J. and Rots, M.G. (2003) Therapeutic modulation of endogenous gene function by agents with designed DNA-sequence specificities. *Nucleic Acids Res.*, **31**, 6064–6078.
- McDonald, C.D. and Maher, L.J. 3rd (1995) Recognition of duplex DNA by RNA polynucleotides. *Nucleic Acids Res.*, **23**, 500–506.
- Soukup, G.A., Ellington, A.D. and Maher, L.J. 3rd (1996) Selection of RNAs that bind to duplex DNA at neutral pH. *J. Mol. Biol.*, **259**, 216–228.
- Belotserkovskii, B.P., De Silva, E., Tornaletti, S., Wang, G., Vasquez, K.M. and Hanawalt, P.C. (2007) A triplex-forming sequence from the human c-MYC promoter interferes with DNA transcription. *J. Biol. Chem.*, **282**, 32433–32441.
- Carbone, G.M., McGuffie, E.M., Collier, A. and Catapano, C.V. (2003) Selective inhibition of transcription of the Ets2 gene in prostate cancer cells by a triplex-forming oligonucleotide. *Nucleic Acids Res.*, **31**, 833–843.
- Karympalis, V., Kalopita, K., Zarros, A. and Carageorgiou, H. (2004) Regulation of gene expression via triple helical formations. *Biochemistry*, **69**, 855–860.
- Semerad, C.L. and Maher, L.J. 3rd (1994) Exclusion of RNA strands from a purine motif triple helix. *Nucleic Acids Res.*, **22**, 5321–5325.
- Ellington, A.D. and Szostak, J.W. (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature*, **346**, 818–822.
- Fitzwater, T. and Polisky, B. (1996) A SELEX primer. *Meth. Enzymol.*, **267**, 275–301.
- Robertson, D.L. and Joyce, G.F. (1990) Selection in vitro of an RNA enzyme that specifically cleaves single-stranded DNA. *Nature*, **344**, 467–468.
- Tuerk, C. and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, **249**, 505–510.
- Naghavi, M.H., Schwartz, S., Sonnerborg, A. and Vahlne, A. (1999) Long terminal repeat promoter/enhancer activity of different subtypes of HIV type 1. *AIDS Res. Hum. Retroviruses*, **15**, 1293–1303.
- Engelke, D.R., Krikos, A., Bruck, M.E. and Ginsburg, D. (1990) Purification of *Thermus aquaticus* DNA polymerase expressed in *Escherichia coli*. *Anal. Biochem.*, **191**, 396–400.
- Laing, L. (1993) Thermodynamic study of the folding and ligand interactions for the 23S rRNA GTPase center. *Doctoral dissertation*. Johns Hopkins University.
- Tabor, S. and Richardson, C.C. (1989) Selective inactivation of the exonuclease activity of bacteriophage T7 DNA polymerase by in vitro mutagenesis. *J. Biol. Chem.*, **264**, 6447–6458.
- Srisawat, C. and Engelke, D.R. (2001) Streptavidin aptamers: affinity tags for the study of RNAs and ribonucleoproteins. *RNA*, **7**, 632–641.
- Srisawat, C., Goldstein, I.J. and Engelke, D.R. (2001) Sephadex-binding RNA ligands: rapid affinity purification of RNA from complex RNA mixtures. *Nucleic Acids Res.*, **29**, E4.
- Benn, S., Rutledge, R., Folks, T., Gold, J., Baker, L., McCormick, J., Feorino, P., Piot, P., Quinn, T. and Martin, M. (1985) Genomic heterogeneity of AIDS retroviral isolates from North America and Zaire. *Science*, **230**, 949–951.
- Gendelman, H.E., Phelps, W., Feigenbaum, L., Ostrove, J.M., Adachi, A., Howley, P.M., Khoury, G., Ginsberg, H.S. and Martin, M.A. (1986) Trans-activation of the human

- immunodeficiency virus long terminal repeat sequence by DNA viruses. *Proc. Natl Acad. Sci. USA*, **83**, 9759–9763.
39. Mathews,D.H., Disney,M.D., Childs,J.L., Schroeder,S.J., Zuker,M. and Turner,D.H. (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl Acad. Sci. USA*, **101**, 7287–7292.
 40. Mathews,D.H., Sabina,J., Zuker,M. and Turner,D.H. (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.*, **288**, 911–940.
 41. De Rijk,P. and De Wachter,R. (1997) RnaViz, a program for the visualisation of RNA secondary structure. *Nucleic Acids Res.*, **25**, 4679–4684.
 42. Zeichner,S.L., Kim,J.Y. and Alwine,J.C. (1991) Linker-scanning mutational analysis of the transcriptional activity of the human immunodeficiency virus type 1 long terminal repeat. *J. Virol.*, **65**, 2436–2444.
 43. Pohorille,A., Ross,W.S. and Tinoco,I. Jr (1990) DNA dynamics in aqueous solution: opening the double helix. *Int J Supercomput Appl*, **4**, 81–96.
 44. Nonin,S., Leroy,J.L. and Gueron,M. (1995) Terminal base pairs of oligodeoxynucleotides: imino proton exchange and fraying. *Biochemistry*, **34**, 10652–10659.
 45. Andreatta,D., Sen,S., Perez Lustres,J.L., Kovalenko,S.A., Ernstring,N.P., Murphy,C.J., Coleman,R.S. and Berg,M.A. (2006) Ultrafast dynamics in DNA: “fraying” at the end of the helix. *J. Am. Chem. Soc.*, **128**, 6885–6892.
 46. Smolina,I.V., Demidov,V.V., Soldatenkov,V.A., Chasovskikh,S.G. and Frank-Kamenetskii,M.D. (2005) End invasion of peptide nucleic acids (PNAs) with mixed-base composition into linear DNA duplexes. *Nucleic Acids Res.*, **33**, e146.