DNA Aptamers Selected Against the HIV-1 trans-Activation-responsive RNA Element Form RNA-DNA Kissing Complexes*

(Received for publication, December 1, 1998, and in revised form, February 8, 1999)

Claudine Boiziaz, Eric Dausse, Ludmila Yurchenko‡, and Jean-Jacques Toulmé§
From INSERM U 386, IFR Pathologies Infectieuses, Université Victor Segalen, 146 rue Léo Saignat, P-33076 Bordeaux Cedex, France

In vitro selection was performed in a DNA library, made of oligonucleotides with a 30-nucleotide random sequence, to identify ligands of the human immunodeficiency virus type-1 trans-activation-responsive (TAR) RNA element. Aptamers, extracted after 15 rounds of selection-amplification, either from a classical library of sequences or from virtual combinatorial libraries, displayed an imperfect stem-loop structure and presented a consensus motif 5′ACTCCCAT in the apical loop. The six central bases of the consensus were complementary to the TAR apical region, giving rise to the formation of RNA-DNA kissing complexes, without disrupting the secondary structure of TAR. The RNA-DNA kissing complex was a poor substrate for Escherichia coli RNase H II, likely due to steric and conformational constraints of the DNA/RNA heteroduplex. 2′-O-Methyl derivatives of a selected aptamer were binders of lower efficiency than the parent aptamer in contrast to regular sense/antisense hybrids, indicating that the RNA/DNA loop-loop region adopted a non-canonical heteroduplex structure. These results, which allowed the identification of a new type of complex, DNA-RNA kissing demon-stage, the interest of in vitro selection for identifying non-antisense oligonucleotide ligands of RNA structures that are of potential value for artificially modulating gene expression.

In the antisense strategy, a DNA oligonucleotide is designed to hybridize to an RNA sequence, in order to inhibit specifically the reading of the encoded genetic information (1). Although RNA is a single chain nucleic acid, it adopts secondary and tertiary structures, which can prevent the hybridization of the antisense sequence. This is one of the likely explanations of the poor inhibition efficiency, if any, induced by some antisense oligonucleotides. The aptamer strategy has been successfully used for the selection of ligands against a large range of targets, such as proteins and small molecules (nucleotides, amino acids, dyes) (see Ref. 2 for a review). This methodology offers an alternative way for designing nucleic acid ligands against an RNA structure. Indeed, we previously demonstrated that in vitro selection of DNA ligands (aptamers) against DNA secondary structures led to the identification of sequences able to recognize the RNA targets through base pair formation and additional unidentified interactions (3, 4). This might be of high potential interest, as numerous RNA structures display a regulatory function through interaction either with proteins (such as the iron-responsive element interacting with the iron-responsive element-binding protein (5), the HIV trans-activation-responsive (TAR)3 element binding to the viral protein Tat (6), or the HIV Rev-responsive element promoting the export of retroviral RNA from the nucleus due to the binding with the viral protein Rev (6)) or with nucleic acids (like the dimerization-initiating sequence of HIV (7)). The binding of an oligonucleotide to such structures could prevent the interaction of the RNA with the regulatory partner, hence controlling the expression of the target gene.

We used the aptamer strategy to identify DNA ligands of the TAR RNA element; this RNA structure is a 59-nucleotide-long stem loop present at the 5′ end of HIV-1 RNA. TAR RNA mediates the trans-activation of transcription through the binding (i) of the viral protein Tat to a three nucleotide bulge in the upper part of the stem (6), and (ii) of cellular proteins to the upper region of the hairpin (8). The resulting RNA-protein complex increases the processivity of the RNA polymerase, allowing the high yield synthesis of the full-length retroviral genome (9). We hypothesized that oligonucleotides able to bind to TAR with a high affinity might compete with the TAR-binding proteins, thus preventing the transcription process.

A DNA library comprising more than 1012 different sequences was screened on the basis of oligonucleotide ability to bind to TAR. We identified DNA sequences (aptamers) displaying a binding constant of about 108 M−1. In contrast to antisense oligonucleotides, the selected aptamers did not invade the TAR stem-loop structure but rather fitted to the existing RNA structure. The selected aptamers could fold into imperfect stem-loop structures, displaying a consensus sequence in the aptamer loop, complementary to TAR loop. Therefore, TARP-aptamer interactions give rise to so-called “kissing complexes,” previously demonstrated to be involved in different natural RNA-RNA complexes (see Ref. 10 for a review). The binding properties of the DNA aptamers were analyzed as follows: we demonstrated that bases outside the loop play a key role in the binding to TAR, even though they do not directly interact with the RNA target. The upper part of the aptamer stem likely pre-organizes the loop to minimize the distortion required for loop-loop complex formation.

MATERIALS AND METHODS

Libraries of Candidates

Oligonucleotides—Four DNA libraries were investigated each one having a random sequence of 30 nucleotides (nt), flanked by 18 nt-long

---

*This work was supported in part by Genset, the “Agence Nationale de Recherches sur le SIDA,” and by the Biotechnology program of the “European Union”. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be

†Recipient of an INSERM PECO fellowship.

§To whom correspondence should be addressed. Tel.: (33) 5 57 57 10 17; Fax: (33) 5 57 57 10 15; E-mail: jean-jacques.toulmé@bordeaux.inserm.fr.

---

1 The abbreviations used are: TAR, trans-activation-responsive; HIV, human immunodeficiency virus; nt, nucleotides; PCR, polymerase chain reaction; EMSA, electrophoretic mobility shift assay; DEPC, diethyl pyrocarbonate; FI, FI1, FI11, and FIV, Family I, II, III, and IV, respectively.
conserved sequences used for PCR amplification. Families I, II, and III were obtained by in situ chemical (FI) or enzymatic (FII, FIII) ligation of so-called “half-candidates” from two sub-libraries, 5‘-GCAGTCTCTGGACATCCGCC(5N)15 and 5‘-P-NGCC(5N)15-GTCTCTGGACATCCGCC-CCGC (5N representing any of the four natural nucleic acid bases). Family IV was 5‘-GCAGTCTCTGGACATCCGCN15-GTCTCTGGATCGAC-AGCGC. Oligonucleotide primers were P1, 5‘-GCAGTCTCTGGACATCCGCC, and P2, 5‘-CTCGCCTGGATCGACAGC.

Chemical Ligation (FI)—The mixture of half-candidates (100 pmol of each) was heated for 1 min at 80 °C and then cooled down in 5 min from 80 to 40 °C. 2 pmol of TAR in a final volume of 10 μl of 50 mM sodium cacodylate buffer, pH 7.0, containing 50 μM NaCl, 10 mM MgCl2 were then added to the candidates. A further linear decrease of temperature from 40 to 20 °C, followed by a decrease from 20 to 4 °C at -1.5 °C/h, allowed the candidates to equilibrate in the presence of TAR. 10 μl of 400 mM imidazole, pH 7.0, 200 mM NiCl2, 100 mM BrCN were then added at 4 °C, and the ligation reaction was allowed to proceed for 24 h at 4 °C.

Enzymatic Ligation (FII and FIII)—The same TAR/candidate mixing and annealing conditions as above were used, except the buffer (50 mM Tris acetate, pH 7.5, and 10 mM dithioerythritol), in a final volume of 20 μl of 400 mM imidazole, pH 7.0, 200 mM NiCl2, 100 mM BrCN were then added at 4 °C, and the ligation reaction was allowed to proceed for 24 h at 4 °C.

In Vitro Selection

15 rounds of selection/amplification were performed with each family.

Selection—10–75 pmol of single-stranded PCR-amplified candidates were incubated for 2 h at 4 °C in the presence of 1 pmol of 3′ end-biotinylated TAR, in D buffer (10 mM Tris-HCl, pH 7.5, 10 mM MgCl2, 50 mM NaCl, and 1 mM dithioerythritol), in a final volume of 20 μl. TAR candidates were captured by magnetic streptavidin-coated beads (Promega), and the bound candidates were then eluted with Proteinase K.

Amplification—Double-stranded amplification was performed in 50 μl with 25% of the TAR-bound candidates as template and 50 pmol of each primer, using Taq polymerase (0.5 units from Promega). Single-stranded candidates were then produced from the previous reaction by 15 rounds of selection/amplification with each primer P1 and P2.

Cloning—After the 15th round of selection, the candidates were further amplified with primers P1lon (5′-AATTCCTTCGACATCGTCGACCATCCGCC) and P2lon (5′-GGCGCTCTAGACTGCGATCGACAGC) and cloned in pBluescript. Inserts were sequenced by the Sanger method, either with T7 sequencing kit (Amersham Pharmacia Biotech) or by automatic sequencing with dye terminators (Perkin-Elmer).

Identification of Positive Clones, Electrophoretic Mobility Shift Assay

Transfected bacteria were lysed by 5 min heating at 96 °C and then mixed with 50 pmol of each primer P1 and 5′-phosphorylated primer P2. After PCR amplification, single-stranded candidates were obtained by incubation for 2 h at 37 °C with 3 units of a exonuclease III (Life Technologies, Inc.) in a 67 mM glycine, NaOH buffer, pH 9.4, containing 2.5 mM MgCl2, in order to selectively remove the phosphorylated strand. Candidates were phenol-extracted, precipitated, and evaluated for TAR binding by electrophoretic mobility shift assay (EMSA); 50 nM 32P-5′-end-labeled TAR were incubated with candidates in D buffer for 2 h at 4 °C. The mixture was then loaded on a 10% polyacrylamide gel containing 50 mM Tris acetate, pH 7.5, and 10 mM magnesium acetate, run at 4 °C for 18 h at 10 V/cm. For the subsequent analysis of chemically synthesized aptamers, the same procedure was used, except that the oligonucleotide of interest (5 nM) was labeled.

Footprints

In all experiments, 50 nM 5′-end-radiolabeled oligonucleotide (either TAR or the candidate) was preincubated for 20 min at 4 °C with 200 nM of the partner, in 8 μl (final volume) of the selection buffer. Radiolabeled TAR was then partially digested with either 2 ng of RNase A (Boehringer Mannheim) or 0.5 units of RNase T1 (Boehringer Mannheim) for 20 min at 4 °C. The 5′-end-radiolabeled DNA aptamer was digested by 200 units of S1 nuclease (Boehringer Mannheim) for 20 min at 4 °C. Diethyl pyrocarbonate (DEPC, 10%) reaction and potassium permanganate (KMnO4, 2 mM) modification were performed at 4 °C for 90 and 4 min, respectively. The DNA aptamer was then ethanol-purified and cleaved by 1 μM piperidine, for 30 min at 90 °C. After DEPC modification, radiolabeled TAR was ethanol-purified, resuspended in 10 μl of Tris-HCl, 1 mM MgCl2, and 50 pmol of each primer P1 and P2 were added to the candidates. A further linear decrease of temperature from 40 to 20 °C, followed by a decrease from 20 to 4 °C at -1.5 °C/h, allowed the candidates to equilibrate in the presence of TAR. 10 μl of 400 mM imidazole, pH 7.0, 200 mM NiCl2, 100 mM BrCN were then added at 4 °C, and the ligation reaction was allowed to proceed for 24 h at 4 °C.

Results

Combinatorial Libraries—We screened for their binding to the TAR RNA element a 30-nt randomized oligodeoxynucleotide library which theoretically contained 43010 different sequences. However, as the size of our experiment allowed us to handle about 10 pmol, only one out of 170,000 sequences was present in the solution at the first selection step, thus substantially reducing the diversity of the library. To circumvent this limitation, we considered the possibility to use a template-assisted combinatorial strategy; two sub-libraries were synthesized, each one with a 15-nt random region linked, either at its 5′ end (sub-library 1) or at its 3′ end (sub-library 2), to fixed sequences used for PCR amplification with primers P1 and P2 in further steps of the selection. Moreover, the sequences in the sub-library 2 were 5′-phosphorylated. Therefore each sub-library contained about 1010 different sequences that were present on the average 6-105 times each at the scale of the experiment.

Upon simultaneous mixing with the TAR stem-loop RNA, oligomers from sub-libraries 1 and 2 may bind independently to the target RNA. We hypothesized that a bound candidate of the sub-library 1 might have its 3′ end in an appropriate position for being ligated to the 5′ end of a bound candidate of the sub-library 2, thus generating a 30-nucleotide central region. Only such ligated candidates can be amplified from P1 and P2. The ligation reaction was achieved either chemically (Family I, FI) or enzymatically, under two different sets of conditions (Families II and III, FII and FIII; see “Materials and Methods”). For comparison, a fourth family (FIV) corresponded to a library of “standard” combinatorial sequences comprising a 30-nucleotide random stretch between the fixed regions corresponding to P1 and P2.

Sequences exhibiting affinity for 3′-biotinylated TAR RNA were captured by magnetic streptavidin beads. The selection was stopped after the 15th round, and candidates from the four families were cloned and sequenced.

Selected Candidates, Three Classes of Sequences Emerged—Direct PCR amplification was performed from the bacterial colonies, and the crude oligonucleotide solutions were incubated with 32P-5′-end-labeled TAR and tested by EMSA. 21 clones, representing 100% of TAR at 50 μM and were sequenced, as well as 55 other clones that did not induce any shift at this concentration. All together, 76 clones (12 in FI, 18 in FII, 20 in FIII, and 26 in FIV) were sequenced.

Primary structure analysis led to four classes of sequences (Fig. 1) as follows. (i) Class A was composed of sequences containing the octamer 5′-ACTCCCAT (sequences II-25, II-29, II-36, III-43 and IV-40 showed one point mutation in the consensus). Three families (FII, III, and IV) were represented in...
DNA Aptamers for the HIV-1 TAR RNA Element

**Class A**

- 11-14 AAGCAATATATACGAACTACACACTCCATCTATGGTC
- 11-15 AAGCAATATATACGAACTACACACTCCATCTATGGTC
- 11-17 ACTATGACGTGATAGCTACACACTCCATCTATGGTC
- 11-25 AGGCAATATATACGAACTACACACTCCATCTATGGTC 3 X
- 11-29 ATATAGGCTACCTCCATCTTACGAGT
- 11-36 CAGGATCTACACTCCATCTTATGATGGT
- 11-25 CAGGGAATGATACACTCCCATCTATGGTC ++
- 11-33 ACATGAATGATACACTCCCATCTATGGTC ++
- 11-43 ACAAGGAGATACACTCCCATCTATGGTC ++
- 11-45 ACAAGGAGATACACTCCCATCTATGGTC ++

**Class B**

- 11-06 TACAGCAATACAATCCTCCTCCTCCTTGG
- 11-23 ACAAGATATGCTCCCTCCCTCCCTCCCTCCCT
- 11-26 GACACGACATCCCTCTCCCTCCCTCCCT
- 11-27 AGGAGAATGCTCCCTCTCCCTCCCTCCCT
- 11-35 ACAAGATATGCTCCCTCCCTCCCTCCCT
- 11-36 TGCCAAATGCTCCCTCTCCCTCCCTCCCT
- 11-05 ACCAAACATCTGCTCACTCCCTCCCTCCCT
- 11-07 TACAAAGCAATACCTCCTCCTCCTCCTCCT
- 11-08 ACAAGATATGCTCCCTCCCTCCCTCCCT
- 11-10 CAAGACACTGCTCCCTCTCCTCCTCCT
- 11-11 ACAAGATATGCTCCCTCCCTCCCTCCCT
- 11-18 CAACAAACTGCTCCCTCTCCTCCTCCT
- 11-21 ACAATGACATCTTCTATG
- 11-22 AGTACAGGATACACTCCCATCTATGGTC
- 11-26 AGATACAGGATACACTCCCATCTATGGTC
- 11-27 TGCACATATGCTCCCTCCCTCCCTCCCT

**Class C**

- 11-02 AGGGATATGCTCATTCTAATCATATCCCT
- 11-03 ACCAAACATCTGCTCCCTCCCTCCCT
- 11-07 ACAATGACATCTTCTTCTCCTCCTCCT
- 11-08 ACAAGATATGCTCCCTCCCTCCCTCCCT
- 11-10 ACAAGATATGCTCCCTCCCTCCCTCCCT
- 11-15 ACAATGACATCTTCTTCTCCTCCTCCT
- 11-16 AGGAGAATGCTCCCTCTCCTCCTCCTCCT
- 11-19 CAAGACACTGCTCCCTCTCCTCCTCCT
- 11-20 AGGAGAATGCTCCCTCTCCTCCTCCT
- 11-21 CAGGATCTACACTCCCATCTTATGATGGT
- 11-22 GCCAGACATCTTCTTCTCCTCCTCCT
- 11-23 GCCAGACATCTTCTTCTCCTCCTCCT
- 11-24 AGATACAGGATACACTCCCATCTATGGTC
- 11-25 AGGAGAATGCTCCCTCTCCTCCTCCT
- 11-26 AGGAGAATGCTCCCTCTCCTCCTCCT
- 11-27 AGATACAGGATACACTCCCATCTATGGTC
- 11-28 TGCACATATGCTCCCTCCCTCCCTCCCT

**Class D**

- 11-11 GGCGAGGGTGAGGATCTGTAAGTTAT
- 11-09 GGCGAGGGTGAGGATCTGTAAGTTAT 12 X

**Fig. 1. Selected sequences.** The 30-nucleotide randomized region only is represented. Sequences are named after the family of selection (I–IV), ranked in classes A, B, C, or D, and classified according to primary structure criteria (see text). In A and B classes, the octameric consensus and the pyrimidine-rich region, respectively, are in the primary structure criteria (see text). In A and B classes, the octameric consensus and the pyrimidine-rich region, respectively, are in the primary structure criteria (see text). In A and B classes, the octameric consensus and the pyrimidine-rich region, respectively, are in the primary structure criteria (see text). In A and B classes, the octameric consensus and the pyrimidine-rich region, respectively, are in the primary structure criteria (see text). In A and B classes, the octameric consensus and the pyrimidine-rich region, respectively, are in the primary structure criteria (see text).

- **Class A**
  - 11-04 ACCTGACCTCAAACACTCCCTCCT
  - 11-05 TTGAGATATATATACGAACTACAC
  - 11-06 AGGAGAATGCTCCCTCTCCTCCTCCT

- **Class D**
  - 11-04 ACCTGACCTCAAACACTCCCTCCT
  - 11-05 TTGAGATATATATACGAACTACAC

**D. Sekkai, C. Boiziaux, and J. J. Toulmé, unpublished observations.**
The secondary structures of IV-04, III-2539, III-3339, and IV-4039 were predicted by computer folding (12) and confirmed by enzymatic and chemical mapping (see Fig. 4). Truncated IV-0439 is indicated by solid black arrows. The sequences corresponding to the primers are in italic, the octameric consensus in bold face, and the mutated positions are shaded. +A and −T indicate an insertion or a deletion, respectively. The $K_f$ values (in nM) determined by EMSA, in D buffer at 4 °C, of intact aptamers (in parentheses) and of mutated or point-deleted aptamers are given. PG, propylene glycol linker.

**Fig. 2.** Secondary structure and $K_f$ values of some class A′ candidates. The secondary structures of IV-04, III-2539, III-3339, and IV-4039 were predicted by computer folding (12) and confirmed by enzymatic and chemical mapping (see Fig. 4). Truncated IV-0439 is indicated by solid black arrows. The sequences corresponding to the primers are in italic, the octameric consensus in bold face, and the mutated positions are shaded. +A and −T indicate an insertion or a deletion, respectively. The $K_f$ values (in nM) determined by EMSA, in D buffer at 4 °C, of intact aptamers (in parentheses) and of mutated or point-deleted aptamers are given. PG, propylene glycol linker.

The bulge area, known to be crucial for Tat binding (13) were mutated or permuted; no modification of the affinity was detected with “TAR UCU,” “TAR del,” “TAR UA,” or “TAR CG” compared with the wild type RNA, except for the CG permutation, which induced a modest 2.5-fold destabilization (Fig. 5). Finally, a TAR with a truncated bottom stem (“mini-TAR”) showed a slightly increased affinity (2-fold) compared with the full-length TAR (Fig. 5). These results indicate that the interaction between TAR and the class A aptamers is driven by the TAR loop and the top part of the aptamer.

The Aptamer Stem Organizes the Aptamer to Stabilize the Complex—To determine the role of the stem, the aptamer IV-04 was progressively truncated from the bottom of the stem region; the affinities of the resulting IV-04 derivatives were evaluated by EMSA (Fig. 6). As reported in Table I, no dramatic loss of affinity was detected until a 12-mer is generated. The five nucleotides at the 5′ end were unessential to complex formation as was the bottom of the stem and the internal loop (IV-0461 and IV-0449). The deletion up to nt 23 and beyond nt 48 induced a 6-fold decrease of the affinity for IV-0449 compared with the parent aptamer ($K_f = 120$ nM, Table I). Then the progressive deletion of one base pair at a time (nt 23/48, 24/47, 25/46) was without major further consequence (aptamers IV-0424 to IV-0440; $K_f = 150$ nM, Fig. 6). The deletion of the bulged G49 in IV-0449 did not weaken the association of the mutated aptamer with TAR (not shown). In contrast, IV-0442 had a $K_f$ value of 1 μM. It is only 12 nt long and contains the consensus sequence but certainly cannot fold in a stable stem-loop structure under the conditions used for binding assays (Table I). This indicated that the active form of the aptamer was a structured stem-loop presenting the consensus in the loop.

Loop-Loop Interactions in TAR-IV-04 Complex Distort the RNA/DNA Duplex Structure—As 2′-O-methyl oligonucleotides have a higher affinity for RNA than DNA (14), we hypothesized that 2′-O-methyl (2′-O-Me) modifications at the positions leading to the potential formation of Watson-Crick pairs in loop-loop complexes could stabilize IV-04-TAR complexes. But the modified oligonucleotide IV-0439-WC2′-O-Me displayed a 3-fold lower affinity for TAR than the unmodified parent aptamer (Table I). About the same affinity was obtained for the fully modified molecule ($K_f = 160$ nM). This indicated that the loop-loop duplex did not behave as a canonical double-stranded heteroduplex.

We investigated the activity of the E. coli RNase H on a radioactively labeled TAR bound to IV-04 at a saturating concentration. As shown on Fig. 7a, two faint bands corresponding to cleavage at G33 and G34 were observed, demonstrating that the TAR loop was involved in a DNA/RNA hybrid. When TAR was mixed with truncated derivatives of IV-04 (IV-0429,26,20,12), the cleavage efficiency increased, but the cleavage pattern remained unchanged (Fig. 7a). The efficiency of cleavage of TAR by RNase H (at an aptamer concentration 50-fold over the $K_f$, i.e. sufficient to ensure a complete hybridization) was inversely related to the affinity.

Role of the Bulged T at the Apter IV-04—The computer-predicted structure of IV-04, confirmed by the footprints performed with chemicals and S1 nuclease, suggested the presence of a bulged T (nt 41) at the top of the stem (Fig. 2). The deletion of this T (IV-0449(T−41)) or the introduction of an A in the opposite strand (IV-0449(+A40)), both modifications, which lead to a perfect double-stranded stem next to the loop of the aptamer, induced a large destabilization of the complex ($K_f > 1 \mu M$, Fig. 2). However, the substitution in IV-0439 of T41 by A, C, G or by a propylene glycol linker induced minor changes in affinity, leading to $K_f$ values of 60, 150, 20, and 250 nM, respectively, compared with 50 nM for the parent aptamer (Fig. 2). This supports a structural role, rather than a sequence effect of this bulged residue. The pattern of RNase H cleavage of TAR induced by IV-0449(+A40) was totally unexpected and differed from that obtained with other IV-04 derivatives (Fig. 7a). Multiple cleavage sites were detected with a decreasing sensitivity in the order $U_{31}, C_{30} > C_{29}, G_{28}, A_{27}, G_{26} > U_{25}, U_{24}$ (Fig. 7b). In this region, only $U_{31}$ was predicted to be paired in the TAR RNA/aptamer DNA hybrid (Fig. 3); no complementarity between the TAR and the other nucleotides of this sequence could be identified. This pattern resulted from the insertion of a single nucleotide compared with the parent IV-0439. In contrast to IV-0439(+A40), the IV-0449(T−41) derivative (in which the T41 was deleted) led to the same cleavage sites as IV-0439, although with a much higher efficiency (not shown). In this latter case, the cleavage yield was also inversely related to the affinity.

In class A aptamers, a conserved T was identified at position +2 or +3 relative to the 3′ end of the consensus sequence (Fig. 1). We investigated the properties of three other class A′ candidates.
aptamers, in which this T was deleted. For III-33 and III-25, no loss of affinity was obtained compared with the parent aptamer, in contrast to IV-40 for which $K_d$ increased from 120 nM for the wild type sequence to more than 1 mM (Fig. 2).

According to secondary structure predictions, this T was located either in a bulge or in the stem depending on the aptamer. So, the presence of this conserved T might not be significant with respect to the binding property of class A aptamers to TAR.

**DISCUSSION**

We have selected aptamers against the HIV-1 TAR RNA element from DNA libraries containing candidates randomized at 30 positions. In order to take full advantage of the molecular diversity of such a population ($>10^{18}$ different sequences) that surpasses by about $10^5$ times the number of individuals that can be screened at a time in our *in vitro* selection experiment, we used a template-assisted combinatorial strategy for the first round of selection (15). This approach relied on the spontaneous self-assembly on the surface of the target molecule, of short oligomers belonging to two different sub-libraries. The amplification by polymerase chain reaction (PCR) used to produce the second generation of candidates that will be screened at the next selection step requires a physical link between an oligomer from sub-library 1 and another one from sub-library 2. This link was provided by a ligation step prior to the PCR. Therefore, as any sub-library 1 sequence can potentially be combined to any sub-library 2 sequence, we can in principle explore the full diversity of this virtual library, i.e. $4^{30}$ entities. The general interest of such libraries over standard selection methods is discussed in detail in Ref. 15. The ligation was ensured either chemically (FI) or enzymatically under two different conditions (FII and FIII), and the results of a 15-round selection with these three families were compared with that obtained with a conventional library with 30 random positions (FIV). The analysis of 76 clones belonging to the four families led to the identification of 20 sequences defining a single class of aptamers (class A) exhibiting affinity for TAR (the high background level is partly ascribed to the fact that the selection procedure was not stringent enough and that no coun-

---

**Fig. 4. Footprints of TAR-IV-04 complexes.**

- **a**, radiolabeled TAR (TAR*, 50 nM) was incubated in the absence (−) or in the presence (+) of 200 nM IV-04 and digested by RNase A (A) or RNase T1 (T1). The sequence of TAR is given to the left. **b**, radiolabeled IV-04* (IV-04*, 50 nM) was incubated in the absence (−) or in the presence (+) of 200 nM TAR and digested by S1 nuclease (S1). The sequence of IV-04 is given to the left. **c**, summary of major cleavage sites in TAR-IV-04 complex induced by S1 nuclease (∆), RNase T1 or A ([darrow]), DEPC (○), and K MnO 4 (○). Open and solid symbols refer to protected and sensitized sites in the complex, respectively.

---

**Fig. 5. Affinity of mutated TAR elements for the aptamer IV-04.** The $K_d$ values were evaluated by EMSA in D buffer at 4 °C.

---

**Fig. 6. Affinity of IV-04 and IV-04* for TAR.** Radiolabeled IV-04 (a) and IV-04* (b) were incubated with an increasing concentration of TAR (in nM) and analyzed by band-shift assay. The complexes are indicated by an arrow.
The ligation step did not provide a clear benefit to the selection process. Indeed, winners (class A aptamers) were obtained from Families III and IV. At least four reasons might account for this outcome as follows. First, the template-assisted ligation of the candidates occurred only once at the first step of the process. Therefore, any further round will weaken the early potential enrichment of the library. Second, for our selected sequences, a limited number of residues actually contact the target: at most six bases engage direct interactions (hydrogen bonds) with the TAR RNA. Third, ligation occurred in the octameric consensus sequence, located in the loop-loop duplex, which can impede the enzyme or chemical reagent action. Finally, only combinations for which the 5'-phosphate group of one candidate is in contact with the 3'-OH end of another candidate can be ligated, thus restricting the selected sequences to a sub-class.

In all families, the selected sequences showing binding ability for TAR displayed two important common features as follows: (i) a stem-loop structure and (ii) an octameric sequence in the loop, partly complementary to the TAR loop. We demonstrated that the TAR-aptamer complex involves loop-loop interactions, showing that kissing complexes constitute a valid recognition mode for RNA stem-loops by DNA sequences. Such interactions between two RNA stem-loops were previously extensively studied. It was demonstrated that kissing RNA complexes play a key role in the regulation of plasmid replication (16) and in the dimerization of the retroviral genome (7). But this is to our knowledge the first time that such a complex is described between DNA and RNA hairpins.

The structure of the double-stranded region resulting from loop-loop interactions is unknown. Recent NMR studies have shown that kissing RNA hairpins form a quasi-continuous structure of three coaxially stacked helices (17, 18); the loop-loop helix is distorted compared with the A-form RNA and is bent toward the major groove. This reduces the distance between the two strands allowing a single phosphodiester bond to bridge the loops across the major groove. One such complex was formed between the HIV-1 TAR and TAR*, an RNA hairpin

### Table I

Affinity of IV-04 derivatives for TAR

| Oligonucleotides | \( K_d \) values were evaluated by EMSA in D buffer at 4 °C with 5 nM radiolabeled aptamers. 2'-O-Methyl nucleotides are underlined. |
|------------------|----------------------------------------------------------------------------------------------------------------------------------|
| IV-04            |  |  |
| GCACT ACT        | 15 | ACT |
| CT-CGTCG CCAGCA-GGCATG-TAC  | C |
| GA CGACG GTGTGAC AT  | C |
| C CTA G T AC | ACT |
| IV-04<sub>2</sub> | 6 | ACT |
| CT-CGTCG CCAGCA-GGCATG-TAC  | C |
| GA CGACG GTGTGAC AT  | C |
| C CTA G T AC | ACT |
| IV-04<sub>39</sub> | 17 | ACT |
| CCAGCA-GGCATG-TAC  | C |
| GTGTGAC AT  | C |
| IV-04<sub>39</sub> | 23 | ACT |
| GGCATG-TAC  | C |
| GTGTGAC AT  | C |
| IV-04<sub>20</sub> | 26 | ACT |
| CATG-TAC  | C |
| GTAC AT  | C |
| IV-04<sub>12</sub> | 30 | ACT |
| TA C  | C |
| AT C  | C |
| IV-04<sub>39</sub> 2'O-Me | 17 | ACT |
| CCAGCA-GGCATG-TAC  | C |
| GTGTGAC AT  | C |
| IV-04<sub>39</sub> WC 2'O-Me | 17 | ACT |
| CCAGCA-GGCATG-TAC  | C |
| GTGTGAC AT  | C |

**DNA Aptamers for the HIV-1 TAR RNA Element**
with a fully complementary loop (17). Our TAR-DNA aptamer kissing complexes can potentially involve six base pairs. This would require the disruption of the TAR C^{29–G^{36}} base pair. Alternatively, only 5 base pairs can be formed, leaving an intact TAR stem (Fig. 3). TAR-DNA aptamer complexes will likely adopt a structure different from that of TAR-TAR\(^\text{a}\), as in the former case, loop-loop interactions will generate an RNA/DNA duplex flanked by double-stranded DNA (the aptamer stem) on one side and double-stranded RNA (the TAR stem) on the other side. RNA and DNA helices are A- and B-forms, respectively, whereas RNA/DNA hybrids adopt a different conformation for the two strands (19). This leads to substantially different groove widths and thus to different interstrand distances. Whereas the gap across the major groove is about 10 Å for a 6-base loop in the A-form geometry, it is about twice as large in the B-form geometry (20). Stronger distortions and/or longer “connectors” will be required for the RNA/DNA than for the RNA-RNA kissing complex. Indeed, one or two bases (A\(^{32}\) of IV-04, C\(^{50}\) of TAR, and either C\(^{33}\) of IV-04 or C\(^{29}\) of TAR, see Fig. 3) are necessary to connect the stacked helices, in contrast to what has been previously described for the TAR-TAR\(^\text{a}\) and the ColE1 complexes (17, 18).

The top of the stem of the aptamers seems to play a crucial role in the binding process. It is striking that the base pairs next to the loop are weak, A-T or G-T pairs (Fig. 2). Moreover, bulged bases or internal loops are frequently observed in the upper part of the stem. This suggests that a weak double-stranded structure in the vicinity of the binding loop was selected. Indeed, the removal of the bulged T\(^{41}\) in the aptamer IV-04 or the insertion of an A in the opposite strand, both modifications restoring perfect helicity in this region and consequently stabilizing the stem, led to a 25–40-fold increase in \(K_d\). The nature of the nucleic acid base was not important, as a non-nucleotide linker partly restored the affinity, indicating a structural role for this bulged residue. These imperfect structures are reminiscent of the kissing complexes formed by natural antisense RNAs involved in bacterial plasmid replication. In most of the cases where a single stem-loop is involved, the antisense RNA does not fold in a perfect hairpin structure. As demonstrated by Hjalt and Wagner (21), the deletion of the bulge and of the internal loop of the upper stem in CopA antisense RNA increased the \(K_d\) of the kissing intermediate up to 14-fold, thus showing the structural role of these abnormalities in the helicity for kissing complexes. Therefore, secondary and tertiary structures in these regions are crucial for the proper presentation of the interacting loops in RNA-RNA as well as RNA-DNA loop-loop complexes.

The \(E.\ coli\) ribonuclease H displayed a very low activity on the TAR RNA/IV-04 DNA hybrid (Fig. 7). Three explanations can be proposed as follows. (i) For the length of the heteroduplex, it was reported that heteroduplex as short as 4 base pairs are substrates for this enzyme (22). Even though 6 bases of the IV-04 loop are complementary to the upper part of the TAR RNA element (including the last G residue next to the TAR loop), we do not know the actual number of paired bases. (ii) The formation of the kissing complex requires that the connecting residues between the 3′ ends of the loop-loop “helix” and either the TAR or the aptamer stems cross the grooves of the loop-loop region (18, 23). These linkers might interfere with the RNase H binding and/or activity. Indeed, pseudo-half-knot antisense DNA-RNA hairpin complexes were reported to be cleaved with a reduced efficiency compared with a linear double-stranded RNA duplex, indicating that loop regions crossing either the major or the minor groove of the loop-loop helix interfered with RNase H (24). However, in both cases (“duplex length” and “limited access”), the IV-04 derivatives share the same loop region, i.e. form the same duplex region and have the same connector length. Therefore, we strongly favor the third parameter as the key determinant of low RNase H activity on RNA-aptamer complexes: (iii) the bent structure of the kissing complex. RNase H requires a nucleic acid region upstream (with respect to the RNA strand) from the cleaved region for binding, likely through electrostatic interactions (Fig. 8a) (25). This means that, in IV-04-TAR complexes, the enzyme should interact with the aptamer stem in order, for the catalytic site, to be appropriately positioned on the loop-loop heteroduplex (Fig. 8b). Therefore any distortion at the heteroduplex-aptamer stem junction will displace the catalytic site away from the substrate region. It is known that RNA-RNA hairpin complexes adopt a bent conformation in order to allow loop-loop interactions (18, 23). A similar structure for aptamer-TAR complexes will preclude a good contact between the catalytic site and the loop-loop region once the enzyme is bound to the aptamer stem. The high efficiency of cleavage obtained with the short IV-04_{12} derivative is in reasonable agreement with this hypothesis; in contrast to the bent rigid structure of TAR-IV-04 complex, the 3′ part of IV-04_{12} derivative will provide a flexible binding site for the enzyme compatible with catalytic site facing the TAR RNA hybridized loop (Fig. 8c). When increasing the stem length, the complex is more stacked and constrained, hence more stable, but resulting in a worse substrate for RNase H. Different curvatures induced by different stems (for instance IV-04_{12}(+A\(^{30}\)) and IV-04_{12}(-T\(^{41}\)) will generate various structures for the complexes that are sensed by RNase H, eventually leading to cuts outside the paired region (Fig. 8d). Cleavages outside the RNA/DNA duplex were previously reported for
Fig. 8. Model for RNase H binding and cleavage of aptamer IV-04-TAR RNA complexes. The interaction of the E. coli RNase H basic protrusion with the double strand is indicated by +++ and the catalytic site by an arrow. The binding and digestion schemes of RNase H on a perfect RNA/DNA double strand (a), on the IV-04-TAR kissing complex (b), on the IV-04_{s=TAR} (c), and on the mutated IV-04_{s=A(TAR)} (d) are shown. The RNA and DNA strands are, respectively, double and single lines.

regular (26, 27) and chemically modified duplexes (28, 29).

It is striking that a full-length (30-mer) antisense sequence corresponding to the random region of the library was not selected. This confirmed that antisense oligodeoxynucleotides are not good ligands for structured RNA regions. Indeed, it was demonstrated that anti-TAR DNA 17-mers, including a sequence complementary to the top part of the TAR element and which adopted a hairpin structure, had a poor affinity (30); this oligomer was therefore not able to give rise to kissing complex formation.

We selected DNA molecules that might inhibit TAR-dependent activation of transcription. It is expected that IV-04 will not directly interfere with the binding of Tat, as the binding sites of these two molecules on TAR are different. In contrast, the cellular proteins TRP 185 (31), TRBP (32), and the Tat-cyclin T-CDK9 complex (33) bind to TAR at the level of the loop. This could allow DNA aptamers to prevent the transcription activation controlled by these proteins. The competition between Tat or TRBP and the aptamers is presently under investigation.

Acknowledgments—We are grateful to Sandra Grelier and Justine Michel for their technical assistance.

REFERENCES

1. Hélene, C., and Toulmé, J. J. (1990) Biochim. Biophys. Acta 1049, 99–125
2. Gold, L., Polisky, B., Uhlenbeck, O., and Yarus, M. (1995) Annu. Rev. Biochem. 64, 763–797
3. Mishra, R. K., Le Tinèvez, R., and Toulmé, J. J. (1996) Proc. Natl. Acad. Sci. U. S. A. 93, 10679–10684
4. Boiziau, C., Daussé, E., Mishra, R. M., Ducongé, F., and Toulmé, J. J. (1997) Antisense Nucleic Acid Drug Dev. 7, 369–380
5. Stripecker, R., Oliveira, C. C., McFarland, J. E. G., and Hentze, M. W. (1994) Mol. Cell. Biol. 14, 5896–5909
6. Gait, M. J., and Corn, J. (1993) Trends Biochem. Sci. 18, 255–259
7. Paullart, J. C., Marquet, R., Skrjintin, E., Ehresmann, C., and Ehresmann, B. (1996) Biochimie (Paris) 78, 639–653
8. Gatignol, A., Duarte, M., Davier, L., Chang, Y. N., and Jeang, K. T. (1996) Gene Express 5, 217–228
9. Graebel, M. A., Churcher, M. J., Lowe, A. D., Gait, M. J., and Corn, J. (1995) Proc. Natl. Acad. Sci. U. S. A. 90, 6184–6188
10. Wagner, E. G. H., and Simons, R. W. (1994) Annu. Rev. Microbiol. 48, 713–742
11. von Ahlen, U., and Noller, H. F. (1995) Science 267, 234–237
12. Jaeger, J. A., Turner, D. H., and Zuker, M. (1989) Proc. Natl. Acad. Sci. U. S. A. 86, 7706–7710
13. Churcher, M. J., Lamont, C., Hamy, F., Dingwall, C., Green, S. M., Lowe, A. D., Butler, J. G., Gait, M. J., and Corn, J. (1993) J. Mol. Biol. 230, 90–110
14. Iriberri, A. M., Sprott, B. S., Neuner, P., Sutlton, J., Ryder, U., and Lamond, A. I. (1990) Proc. Natl. Acad. Sci. U. S. A. 87, 7747–7751
15. Hug, I., and Lehn, J. M. (1997) Proc. Natl. Acad. Sci. U. S. A. 94, 2106–2110
16. Tomizawa, J. I. (1986) Cell 47, 89–97
17. Chang, K. Y., and Tinoco, I. (1994) Proc. Natl. Acad. Sci. U. S. A. 91, 8705–8709
18. Marino, J. P., Gregorius, R. S. J., Csonkovski, G., and Crothers, D. M. (1995) Science 268, 1448–1454
19. Federoff, O. Y., Salazar, M., and Reid, B. R. (1993) J. Mol. Biol. 233, 509–523
20. Haasnoot, C. A. G., Hilbers, C. W., Van der Marel, G. A., Van Boom, J. H., Singh, U. C., Pattabiraman, N., and Kollman, P. A. (1986) J. Biomol. Struct. Dyn. 3, 643–655
21. Hjalt, T. A. H., and Wagner, E. G. H. (1995) Nucleic Acids Res. 23, 580–587
22. Inoue, H., Hayase, Y., Imura, A., Iwai, S., Miura, K., and Hoshina, E. (1987) Nucleic Acids Res. 15, 6131–6147
23. Chang, K. Y., and Tinoco, I. (1995) J. Mol. Biol. 269, 52–66
24. Lima, W. F., Mohan, V., and Crooke, S. T. (1997) J. Biol. Chem. 272, 18191–18199
25. Morikawa, K., and Katayanagi, K. (1998) Crystal Structures of RNase H from Prokaryotes, Ribonucleases H (Crouch, R. J., and Toulmé, J. J., eds) pp. 181–193, Les Editions INSERM, Paris
26. Lima, W. F., and Crooke, S. T. (1997) J. Biol. Chem. 272, 27513–27516
27. Le Tinèvez, R., Mishra, R. K., and Toulmé, J. J. (1998) Nucleic Acids Res. 26, 2273–2278
28. Larrouty, B., Boiziau, C., Sprott, B., and Toulmé, J. J. (1995) Nucleic Acids Res. 23, 3434–3440
29. Toulmé, J. J., Le Tinèvez, R., and Brossalina, E. (1996) Biochimie (Paris) 78, 663–673
30. Ecker, D. J., Vickers, T. A., Bruice, T. W., Freier, S. M., Jenison, R. D., Manoharan, M., and Zounes, M. (1992) Science 257, 958–961
31. Wu, F., Garcia, J., Sigman, D., and Gaynor, R. (1991) Nucleic Acids Res. 2273–2278
32. Gatignol, A., Buckler-White, A., Berkhout, B., and Jeang, K. T. (1991) Science 251, 1597–1600
33. Wei, P., Garber, M. E., Fang, S. M., Fischer, W. H., and Jones, K. A. (1998) Cell 92, 451–462

DNA Aptamers for the HIV-1 TAR RNA Element 12737