

Selective optimization of the Rev-binding element of HIV-1

Lori Giver, David Bartel¹, Maria Zapp², Anna Pawul, Michael Green² and Andrew D. Ellington*
Department of Chemistry, Indiana University, Bloomington, IN 47405, ¹Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02118 and ²Program in Molecular Medicine, University of Massachusetts Medical Center, Worcester, MA 01605, USA

Received May 20, 1993; Revised and Accepted September 17, 1993

ABSTRACT

RNA molecules that can bind to the Rev protein of HIV-1 have been isolated from random sequence nucleic acid pools based on a minimal *Rev-binding element (RBE)* found within the *Rev Responsive Element (RRE)*. While the selected sequences are related to the wild-type element, they also contain substitutions that allow them to bind Rev up to 10-fold better *in vitro*. A hypothesized homopurine pairing at G48:G71 is generally replaced by A48:A71; the occasional selection of C48:A71 suggests that R71 may be in a *syn* conformation. These data support the structural model for the *RBE* originally proposed by Bartel *et al.* (1). Additional interactions with the Rev protein are promoted by the sequence CUC ... UYGAG, found in one class of high-affinity aptamers, but absent from the wild-type element. Within each class of aptamers different residues and substructures covary with one another to generate optimal Rev-binding surfaces. The interdependencies of different nucleotide substitutions suggest structural models for both the wild-type *RBE* and the selected high-affinity aptamers.

INTRODUCTION

The cytoplasmic accumulation of viral mRNAs that encode HIV structural proteins requires interactions between the Rev protein and a *cis*-acting *Rev Responsive Element (RRE)* (2–6). Rev binds specifically to the *RRE in vitro* (7–9). Sequences important for Rev-binding were first localized to a 66 nt domain of the *RRE* by mutational analyses and by nuclease and chemical protection studies (10–15). Sequences that can specifically bind to Rev have been mapped to a 'stem-internal loop-stem' *Rev Binding Element (RBE)* (1, 16–18; Figure 1). Nevertheless, structural models of the *RBE* differ substantially; for example, Bartel *et al.* (1) propose that residues G48 and G71 form a homopurine base-pair, while Holland *et al.* (19) suggest that G48 is unpaired. Minimal RNA oligonucleotides that contain the *RBE* bind Rev (16, 18, 20–23), and RNAs as short as 30 nts in length that can compete effectively with the 234 nt wild-type *RRE* for binding to Rev (1).

Such minimal *RBEs* could potentially be used as nucleic acid pharmaceuticals (NAPS) to block the interaction between the Rev

protein and its viral mRNA targets (24–28). However, inefficient uptake of free nucleic acids by cells coupled with practical limitations on the chemical synthesis of oligonucleotides necessitates that therapeutic sequences bind their targets with high affinity, but remain as short as possible.

Despite extensive mapping of the Rev-binding domains of the *RRE*, sparse data exist to guide the construction of minimal *RBEs* that could efficiently disrupt Rev:*RRE* complexes. Most mutational analyses of Rev-binding domain of the *RRE* have been carried out in the context of long (66 nt) viral mRNA sequences, and have focused on identifying sequences and structures necessary for normative, rather than improved, binding. While previous *in vitro* selection experiments (1) identified minimal *RBE* variants that could bind Rev slightly (1.5-fold) better than the wild-type element binds to Rev, only sequences that were closely related to wild-type were examined. Therefore, in order to identify short, high-affinity elements, we selected Rev aptamers (binding sequences; 29) from *RBE* pools containing limited (10–18 nt) random sequence tracts.

MATERIALS AND METHODS

Construction of random sequence pools

Sequences at the defined 5' and 3' ends of the 76.6 DNA pool are, respectively, 5'-GGTAATACGATCACTATAGGGAACTC-GATGAAGCGAGCT-3' and 5'-TACTGACTTCGGATCCCCTGC-(3'), while sequences at the defined 5' and 3' ends of the 79.9 pool are, respectively, 5'-GGTAATACGACTCACTATAGGG-AACTCGATGAAGCGAATT-3' and 5'-GCCTATCTATCGGAT-CCACG-3'. The underlined positions indicate bases covered by the PCR primers used for amplification; sequences in italics correspond to those found in the transcribed RNA pool.

DNA was synthesized on a Biosearch 8750 DNA Synthesizer (Milligen, Burlington, MA.) according to procedures supplied by the manufacturer. Random sequence tracts were introduced by delivery of an equimolar mix of ACG, and T phosphoramidites to the synthesis column. During synthesis of the 79.9 pool four separate 1 μ mol synthesis columns were used to incorporate 6, 7, 8, or 9 bases of an equimolar random sequence mix into the first random tract (near the 3' end of the oligonucleotide). Immediately prior to the synthesis of the second random sequence

* To whom correspondence should be addressed

tract (near the 5', end of the oligonucleotide) the synthesis columns were removed from the machine, opened, and the resins mixed. Each column was then refilled with the mixed resins and the synthesis continued so that again, 6,7,8, or 9 bases of an equimolar random sequence mix was incorporated into the second (5') random tract.

Following synthesis, the 76.6 DNA pool (0.9 μg ; 2×10^{13} sequences) was amplified in a large scale (3 ml.) PCR reaction [50 mM KCl, 10 mM Tris-Cl, pH 8.3, 1.5 mM MgCl_2 , 0.05% NP-40, 0.2 mM dNTPs, 0.5 μM 37.17, 0.5 μM 20.87, 60 U Taq Polymerase; 5 cycles of 94°C., 45"; 40°C., 75"; 72°C. 120"]. The 79.9 pool was amplified via a protocol similar to that used for the 76.6 pool, except that 7 cycles of amplification were carried out and gelatin was used in place of NP-40. Both DNA pools (2 μg) were transcribed *in vitro* using T7 RNA polymerase [40 mM Tris-Cl, pH 7.9, 26 mM MgCl_2 , 0.01% Triton X-100, 2.5 mM spermidine trihydrochloride, 2.5 mM NTPs, 5 mM DTT, 20 U RNasin (Promega, Madison, WI.), 100 U T7 RNA Polymerase (NEBL, Beverly, MA.), 40 nM $\alpha^{32}\text{P}$ UTP (3000 Ci/mmol; NEN, Boston, MA.) in 80 μl total volume] and isolated on an 8% denaturing polyacrylamide gel (30).

Selection of Rev-binding aptamers

Every sequence in the random sequence population was represented $>10^2$ times in the first round of selection. An equivalent amount of selected RNA (0.65 μg ; 13×10^{13} sequences) was used in subsequent rounds. For each cycle of selection, 6.5 μg *E. coli* tRNA (Boehringer, Indianapolis, IN.) was mixed with 25–30 ng of active Rev protein (31) in a total volume of 10 μl of 50 mM KCl, 50 mM Tris-Cl, pH 8.0 (1 \times Binding Buffer, BB). The mixture was allowed to equilibrate for 10' at ambient temperature and then either 76.6 or 79.9 pool RNAs (0.65 μg ; conformers equilibrated at 65°C. for 3' and then ambient temperature for 5') were added in 25 μl of 1 \times BB. Again, the mixture was allowed to equilibrate for 10' and, finally, 8 μg of wild-type *RBE* (transcribed from *ScaI*-cut plasmid pD2 (1), gel isolated, and thermally equilibrated) was added in 15 μl of 1 \times BB. The final 50 μl mixture (0.5 μM RNA pool: 5 μM tRNA: 5 μM wild-type *RBE*: 40–50 nM Rev) was then equilibrated for an additional 60' and vacuum filtered (5–10 inches Hg) on HAWP nitrocellulose filters (Millipore, Bedford, MA.). The filters were washed twice with 500 μl of 1 \times BB and the RNAs eluted twice (100°C, 5') with 200 μl of 7 M urea, 100 mM sodium citrate, pH 5.0, 3 mM EDTA. The combined eluates were phenol and chloroform extracted, ethanol precipitated, and the selected RNAs amplified by reverse transcription [40 mM KCl, 50 mM Tris.Cl, pH 8.0, 6 mM MgCl_2 , 0.8 mM dNTPs, 1–2 μM reverse transcription primer [20.87 for 76.6, or 20.86 for 79.9], 5 U Avian Myeloblastosis Virus reverse transcriptase (Seikagaku, St Petersburg, FL) in a total volume of 20 μl . RNA was mixed with primer and buffer, denatured at 75°C. for 3', and allowed to anneal at 25°C. for 5'; the remaining components were then added and reverse transcription proceeded at 42°C. for 30'] followed by PCR. *In vitro* transcription of the PCR products was used to produce RNA for further cycles of selection.

The pool assays described in Figure 2 were similar to selections except that the total volume of the mixture was 300 μl and the concentrations of components were 0.13 μM RNA pool : 1.3 μM tRNA: 0.13 μM wild-type *RBE*: 7 nM Rev. This assay represents a direct competition between pool sequences and domain II of the *RRE* for binding to Rev. After precipitation,

samples were electrophoresed on a 10% denaturing gel and quantitated using a Molecular Dynamics (Sunnyvale, CA) Phosphorimager. Binding ratios were determined via the following formula: [$\#$ counts filtered pool/ $\#$ counts unfiltered pool]/[$\#$ counts filtered wild-type *RBE*/ $\#$ counts unfiltered wild-type *RBE*]. Appropriate numbers of background counts were subtracted from each value prior to determining the binding ratio. In all pools and clones, increases in aptamer binding are mirrored by decreases in *RRE* binding, indicating that both RNAs bind to a common site on Rev. Under the assay conditions we have used, the binding ratio should be an accurate and internally controlled measure of equilibrium binding constants for *RBE* variants relative to the wild-type element.

Sequencing selected clones

PCR DNA (1 μg) was isolated following round 4 of the 76.6 selection and was cleaved with *Bam*HI and *Sac*I and ligated into appropriately restricted MpLG8 (300 ng). MpLG8 was constructed by introducing synthetic DNA corresponding to the sequence GAAAACCTCGATGAAGCGAGCTCGACAACGG-ATCCCTGCAAGCTT between the *Eco*RI and *Hind*III sites of M13Mp18 (32). PCR DNA (1 μg) from round 4 of the 79.9 selection was cleaved with *Bam*HI and *Sac*I and ligated into appropriately restricted MpLG1 (300 ng). MpLG1 was constructed by introducing synthetic DNA corresponding to the sequence GAATTGGAACCTCGATGAAGCGAATTCTGAAA-GGATCC between the *Eco*RI and *Bam*HI sites of M13Mp18. Single-stranded DNAs were prepared from single plaques according to standard procedures (30). DNA was sequenced according to the procedures supplied with Sequenase 2.0 (USB, Cleveland, OH.); any ambiguities that appeared were resolved by re-sequencing the clones using deoxyinosine in place of deoxyguanosine.

Individual aptamers were transcribed from PCR DNA made from ssDNA clones. Assays for aptamer activity (Tables I and II) were carried out using conditions similar to those used in the selections, except that the total volume of the mixture was 100 μl and the concentrations of components were 0.5 μM aptamer : 5 μM tRNA : 0.5 μM Rev-binding domain of the *RRE*: 20 nM Rev; that is aptamer: wild-type *RBE* ratio was 1:1 rather than 1:10, and the [aptamer + wild-type *RBE*] : Rev ratio was 50:1 rather than 110:1. In addition, aptamer RNAs and wild-type *RBE* were mixed and thermally equilibrated prior to adding them to the [Rev + tRNA] mixture. Competitions were allowed to proceed for 60' following the addition of all components.

RESULTS

Random sequence pool construction

Assuming that most high-affinity *RBE*s would have secondary structures akin to the wild-type, we chose a short stem-internal loop-stem Rev-binding element as a starting point for our selections (Figure 1a). The structural context of this *RBE* is somewhat different than that of the wild-type element: Rev-binding sequences are completely enclosed within 30 nts, rather than a several kilobase long stretch of mRNA. In addition, the 'left' stem (Figure 1) is closed by four rather than two base-pairs, and is no longer part of an extended three helix junction.

Two different random sequence RNA pools were constructed to search for optimal high-affinity minimal *RBE*s. One pool, termed '76.6' (Figure 1b), contained a completely randomized 'internal loop' region, but left stem sequences necessary for

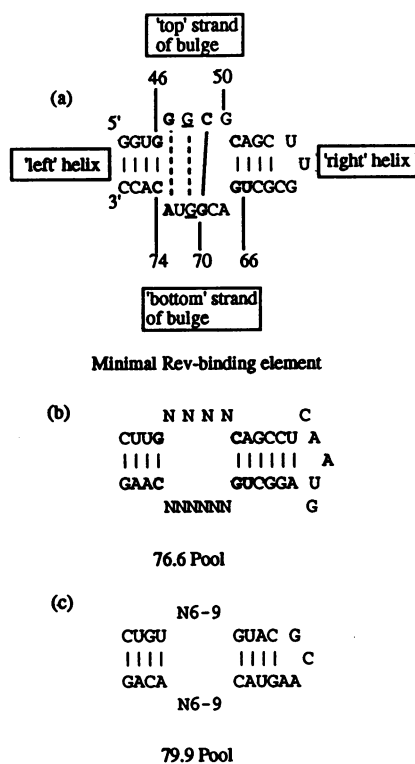


Figure 1. Construction of random sequence *RBE* pools. (a) Sequence of a wild-type minimal *RBE* (1). Bases are numbered according to Malim *et al.* (5, 13). Sequences in boldface have been shown to be important for Rev recognition; underlined residues may form a homopurine base-pair. The orientation of this element as described in the text is defined here. (b) Sequence of the 76.6 pool. (c) Sequence of the 79.9 pool.

recognition by Rev intact. It was expected that aptamers selected from this pool would be similar to the wild-type *RBE*. The other pool, termed '79.9' (Figure 1c), contained two random sequence tracts corresponding to both the 'internal loop' and its flanking stems. Both random sequence tracts were systematically varied from 6 to 9 nts in length. Since all specific sequence contacts necessary for Rev-binding were removed from the 79.9 pool, it was expected that selected ligands might differ significantly from the wild-type element, and, in addition, that those aptamers that did resemble the wild-type 'stem-internal loop stem' would be found in both orientations relative to the fixed sequence flanking helices.

The random sequence tracts introduced into these pools were short enough to permit exhaustive searches for optimal Rev-binding elements. The 76.6 pool contained $4^{10} \approx 1 \times 10^6$ possible variants, while the 79.9 pool contained $[4^{12} + 2 \times 4^{13} + 3 \times 4^{14} + 4 \times 4^{15} + 3 \times 4^{16} + 2 \times 4^{17} + 4^{18}] \approx 1 \times 10^{11}$ possible variants. To ensure that all possible variants could compete for binding to Rev, about 1.5×10^{13} RNA molecules were used in each round of selection.

Selection of Rev-binding aptamers

In all selections, the amount of active Rev protein was limiting relative to the amount of pool RNAs in order to promote competition between individual sequences. The Rev protein was initially mixed with a 10-fold excess of tRNA to saturate non-specific binding sites, then pool RNAs were added. The Rev:pool RNA complexes that formed were then challenged with a 10-fold

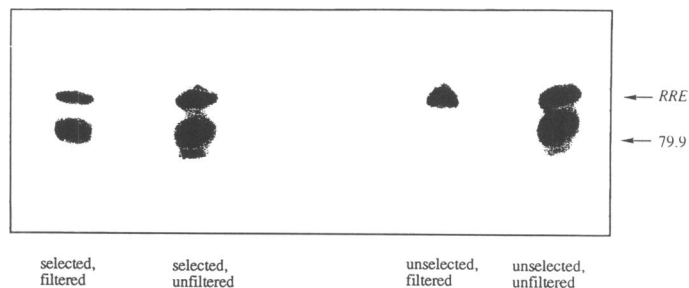


Figure 2. Multiple cycles of selection and amplification enrich the 79.9 RNA pool in Rev-binding aptamers. The two bands in each lane represent respectively, the 94 nt wild-type Rev-binding domain of the *RRE* and the 73 to 79 nt 79.9 pool. These molecules were allowed to compete for binding to a limiting amount of Rev protein. This mixture was then passed through a nitrocellular filter; Rev protein and any RNAs that remained bound to Rev were retained. 79.9 and wild-type *RBE*s were subsequently eluted from the filter and separated from one another by gel electrophoresis. The number of counts in each band was determined using a Molecular Dynamics Phosphorimager. In order to compare the ratio of bound to unbound RNA for each species both before and after selection, one-twentieth of the sample was precipitated prior to filtration and loaded directly on the gel for comparison with the filter sample. The activity of a given pool or aptamer is defined in relation to the wild-type element by the formula: $[\# \text{ counts filtered pool} / \# \text{ counts unfiltered pool}] / [\# \text{ counts filtered wild-type RBE} / \# \text{ counts unfiltered wild-type RBE}]$. This number represents the relative affinities of pool and wild-type RNAs for Rev.

excess of a 94 nt-long version of the Rev-binding domain of the *RRE*. Pool sequences that remained bound to Rev after the challenge were isolated by filtration. Selected RNAs were amplified via reverse transcription, PCR amplification, and *in vitro* T7 transcription, and allowed to again compete for binding to Rev. Multiple cycles of selection and amplification were carried out to enrich the pools in variants that had off-rates that were significantly longer (or equilibrium dissociation constants that were significantly smaller) than those of the wild-type *RBE*.

Following the third cycle of selection, the pools were tested for their ability to compete with the wild-type *RBE* for binding to Rev. Figure 2 demonstrates that the binding ability of the 79.9 pool has improved significantly even after this limited selection. The selected RNAs bind at least 31-fold better than unselected, and are now at least 2.8-fold better than the wild-type *RBE*. Similar results were obtained for the 76.6 pool. Selections were continued for one additional cycle and individual aptamers were cloned and analyzed.

Aptamers selected from the 76.6 pool identify optimal motifs for sequence- and structure-specific recognition of the *RBE* by Rev

As expected, aptamers selected from the 76.6 pool resemble the wild-type *RBE*. Residues analogous to G47, C49, G70, and A73 are present in all sequenced aptamers (Table I). In addition, a homopurine covariation observed by Bartel *et al.* (1) between positions 48 and 71 is found in 30136 of the selected 76.6 aptamers (Figure 3). An A48:A71 combination is found in 20 clones, while the wild-type G48:G71 combination is found in 10 clones. These results are consistent with those of Bartel *et al.* (1), who showed that *RBE*s containing the A:A combination bind Rev more tightly than those that contain the wild-type G:G combination. Interestingly, C48:A71 is the only other base combination that we find between positions 48 and 71 (6/36, at least three independent isolates). This restricted set of covariations is most simply explained by hypothesizing the existence of a set

Table I. Sequences and activities of clones derived from the 76.6 pool

Clone	Sequence		Activity ^a
	GCTCTTG NNNN CAGCCT CAATG AGGCTG NNNNNN CAAG		
	47 50	68 73	
Wild-type ^b	GGCG	ACGGTA	
G48:G71			
9 (2)	GGCG	GTGGTA	1.9±0.21 (3)
18 (2)	GGCG	GAGGTA	2.0 (2)
123	GGCG	GTGGTA	
5	GGCA	GAGGTA	
8	GGCA	GTGGTA A	
19	GGCA	GTGGTA	1.2±0.08 (3)
21	GGCA	GAGGTA	
40	GGCT	CTGGTA	
A48:A71			
6 (4)	GACA	CAGATA	3.1±0.26 (3)
116	GACA	CAGATA	2.1 (1)
120	GACA	CAGATA	
124	GACA	CAGATA	
129	GACA	A	
204 (2)	GACA	CAGATA	
34	GACA	GAGATA	
203	GACA	AAGATA	
111	GACA	GAGAAA	
121	GACA	CAGAAA	
126	GACA	CAGAAA	2.7±0.29 (3)
127	GACA	GCGAAA	
3	GACT	CTGATA	
13	GACT	CTGAAA	
27	GACG	ATGATA	
38	GACT	CTGATA	
C48:A71			
15 (4)	GCCG	ATGATA	1.7 (1)
23	GCCA	GAGATA	
113	GCCA	GTGATA	

^aThe activity of a variant is its binding ability relative to domain II of the wild-type *RRE* in a competitive assay in which Rev was limiting relative to RNA (Materials and Methods). Thus, these numbers should represent relative K_d 's. Numbers in parentheses indicate the number of times a variant was assayed; standard deviations are given where three assays were performed on the same RNA.

^bNumbering based on Malim *et al.*, 1989, 1990.

of approximately isosteric base pairs between positions 48 and 71 (Figure 3).

Residues analogous to G50, A68, C69, and U73 have not been previously identified as important for sequence- or structure-specific recognition. Similarly, we do not find a requirement for any particular sequence or pairing at these positions, but do find a strong sequence covariations between positions 50, 68, and 69 and the central homopurine combination at 48:71. For example, those sequences that contain the wild-type G48:G71 do not have wild-type A68 and C69, but rather G68 and U/A69 (10/10), while aptamers that contain A48:A71 contain A50 (16/20) rather than G50, and C68A69 (10/20). These differences with the wild-type sequence were seen in several independent isolates, and may have functional consequences, since these aptamers all bind Rev better than the wild-type *RBE* within domain II of the *RRE*. Increases in affinity cannot be attributed to the flanking sequences common to all 76.6 aptamers; an RNA that contained the wild-type *RBE* in the context of 76.6 flanking sequences bound Rev no better than does domain II of the *RRE*.

Most authors (1, 16, 19, 22, 23, 33) automatically pair G50 and C69 and assume that they are part of an extended helical stack in which A68 is bulged, though Bartel *et al.* (1) found that a Watson-Crick interaction at this position was not important for Rev-binding. We now find that in the context of the wild-type element, Watson-Crick pairing of position 50 with either position 68 or position 69 actually yields a sub-optimal binding sequence. The bases that are selected at these positions can only rarely (3/36) form canonical Watson-Crick pairings. The paucity

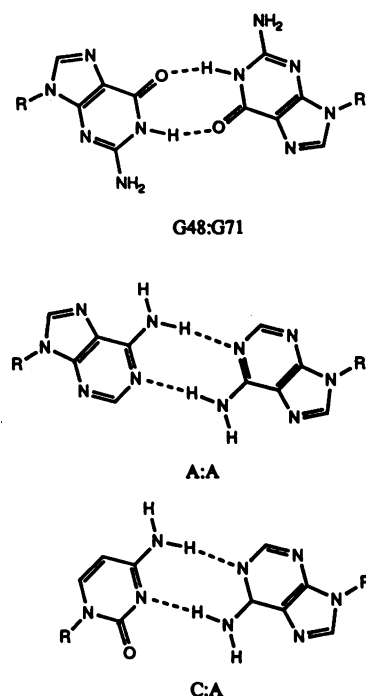


Figure 3. Structural models of base-pairs isosteric to G48:G71. These structures represent the hypothesized homopurine interaction in the wild-type and selected *RBE*s. The G48:G71 pairing shown (top) is proposed for the wild-type *RBE*, while the A:A pairing (middle) is proposed for a number of *RBE* variants that exhibit increased affinity for Rev. It can be seen that the hydrogen bonding patterns, location of ring structures, and placement of the ribose relative to the bases are virtually the same in these two pairings. The proposed C:A pairing (bottom) was discovered in aptamers selected from the 76.6 pool. While cytosine lacks the imidazole ring of adenosine, it nonetheless can form the same hydrogen bonds and places the ribose in much the same position relative to the base. As shown, the strands containing these nucleotide pairings are antiparallel to one another, with one base in *syn* and the other in the *anti* conformation. While other models for interacting bases can also be drawn (e.g., structures VII, V, and XXVI in ref. 34, p. 120), their relative geometries are much less similar.

of potential Watson-Crick pairings is especially striking given the sequences that are selected at positions 50 and 68: if the A (22/36), and U (4/36) residues selected at position 50 were randomly apportioned with the U (15/36) and A (20/36) residues selected at position 68 (rather than covarying away from a Watson-Crick base-pair, as indicated), then a Watson-Crick pairing would be expected to occur roughly 50% of the time, rather than less than 10%. When sequence is allowed to vary freely, selected *RBE*s actively avoid forming a Watson-Crick base-pair between positions 50 and 68. In addition, no (0/36) potential Watson-Crick base-pairs can be formed between positions 50 and 69.

Aptamers selected from the 79.9 pool reveal new high-affinity Rev-binding motifs

In the 79.9 pool all salient structural features of the *RBE* (including its length) were allowed to vary. While there was no apparent size selection of 79.9 aptamers, a specific sequence motif (CUC ... GAG) is present in most of the selected RNAs (21/31; Table II). There are also two examples of (CGC ... GUG). These motifs are unknown in any *RBE* that has so far been sequenced or constructed. The hypothesized structures of these aptamers (Figure 4) illustrate how this motif may replace C49:G70 (and

Table II. Sequences and activities of clones derived from the 79.9 pool

Clone	Sequence ^a					Activity ^b
	ATTCTGT	N6-9	GTAC GCAA GTAC	N6-9	ACAG	
Wild-type		46 51 (GGGGGC)		66 74 (TGACGGTAC)		
Class I						
1		<u>ACTCC</u>		<u>GTCGAGAA</u>		- 4.3 (1)
2		<u>TTAGGACTC</u>		<u>TGAGATACT</u>		2.4 (2)
13		<u>CCTGGCTC</u>		<u>TAGAGAC</u>		3.6 (1)
14		<u>GACTC</u>		<u>TGGAGAA</u>		5.4±0.05 (3)
23		<u>GGACTC</u>		<u>TTGAGATAC</u>		3.1 (1)
50		<u>GTCTC</u>		<u>TGAGAAACG</u>		0.4 (1)
63		<u>GACTCC</u>	T	<u>GTTGAGCA</u>		7.7, 9.5, 9.6
74		<u>AGACTC</u>		<u>TCGAGATAT</u>		4.2 (1)
75		<u>TGGACTCC</u>		<u>GTCGAGAT</u>		1.9 (1)
83		<u>GGACTC</u>		<u>TGAGAAAC</u>	C	3.6 (1)
88		<u>GACTCTTT</u>		<u>AGAGTAT</u>		1.4 (2)
15		<u>GACGC</u>		<u>TTTAT</u>		4.1 (2)
17		<u>GACGGCTC</u>		<u>GCCTGTAT</u>		3.9 (1)
Class II						
3		<u>TGTGCTCC</u>		<u>GTTGAGGT</u>		1.9 (2)
4		<u>AGGTTGAG</u>	- -	<u>GGCTCTCT</u>		2.6 (1)
20		<u>ACTCTC</u>		<u>GATCGAGAC</u>		1.0 (1)
51		<u>GAGCTC</u>		<u>TCGAGGT</u>		2.0 (1)
53		<u>TTGAG</u>		<u>CGCTC</u>		3.2 (2)
58		<u>ACTCCT</u>		<u>GTTGAGAC</u>		0.3 (1)
72		<u>ATGAGA</u>	-	<u>CGACTCT</u>		1.4 (1)
73		<u>GCTCGT</u>		<u>GCTTGGGA</u>		6.4, 6.5, 7.4, 9.7
86		<u>GTAGAG</u>		<u>CGCTCC</u>		2.3 (2)
87		<u>TTTGGAG</u>		<u>CGCGCTCG</u>		7.4 (2)
92		<u>GTAGAG</u>	-	<u>AGCGCTCC</u>		0.7 (1)
Class III						
22		<u>GTAGGTT</u>		<u>ACGGGTT</u>		0.2 (1)
59		<u>GCTTC</u>		<u>TGATAT</u>		4.6 (2)
61		<u>GACATC</u>		<u>CTTGA</u>		2.4 (1)
76		<u>GACTTCG</u>	T	<u>CGACTGAC</u>		3.9 (2)
79		<u>GACATTT</u>		<u>GTTTAT</u>		
85		<u>GACATC</u>		<u>CTTGA</u>		
18	G	<u>AGCATCTT</u>		<u>GAGAGAGCA</u>		0.5 (1)

^aThe (CUC ... GAG) motif is underlined in Class I and II sequences, and the C and G that may correspond to C49:G70 are underlined in Class III sequences.
^bActivities are defined as in Table I and Materials and Methods. Numbers in parentheses indicate the number of times a variant was assayed; standard deviations are given where three assays were performed on the same RNA. Standard deviations are greater for aptamers that bound much better than wild-type, since it was difficult to quantitate the small amount of wild-type RBE that remained bound to Rev during the competition. In these cases, individual values obtained from several independent experiments are listed.

possibly C51:G67) of the wild-type structure. Assuming that the CUC and GAG sequences form a short helical stack, these results also suggest that the normally invariant C49:G70 may exist as a Watson-Crick base-pair.

Almost all aptamers that contain the (CUC ... GAG) motif also contain a novel structural feature not found in the wild-type RBE a pyrimidine bulge in the 'right' helix (apparent exceptions are aptamers 79.9-17 and 79.9-88). This bulge is flanked on one side by the paired (CUC ... GAG) motif, and on the other by 4 to 7 base-paired helical stems. The length of the bulge varies from one base (always a U) to three; the second base from the 3' end of the bulge is always a U. The overall structure of the motif is thus CUC:(UY)GAG.

A second sequence motif found in the selected 79.9 Rev aptamer population is a sequence-specific recognition element that resembles G46G47:A73C74 in the wild-type RBE (16/31). This motif is frequently (12/16) formed by the deletion of a constant region U adjacent to the 'top' N(6-9) random sequence tract (the 3'-most U in the CUGU sequence of Figure 1). Residues that resemble G46, A73, and C74 are recruited from the constant sequence stem while G47 is selected from the random sequence tract and takes the place of the deleted U residue.

Many of the selected 79.9 aptamers (Class I, composed of 79.9-1, -2, -13, -14, -15, -17, -23, -50, -63, -74, -75, -83, and -88) contain each of these defining motifs (Figure 4a) and an unpaired W (A or U) at position 72. Exceptions within this class

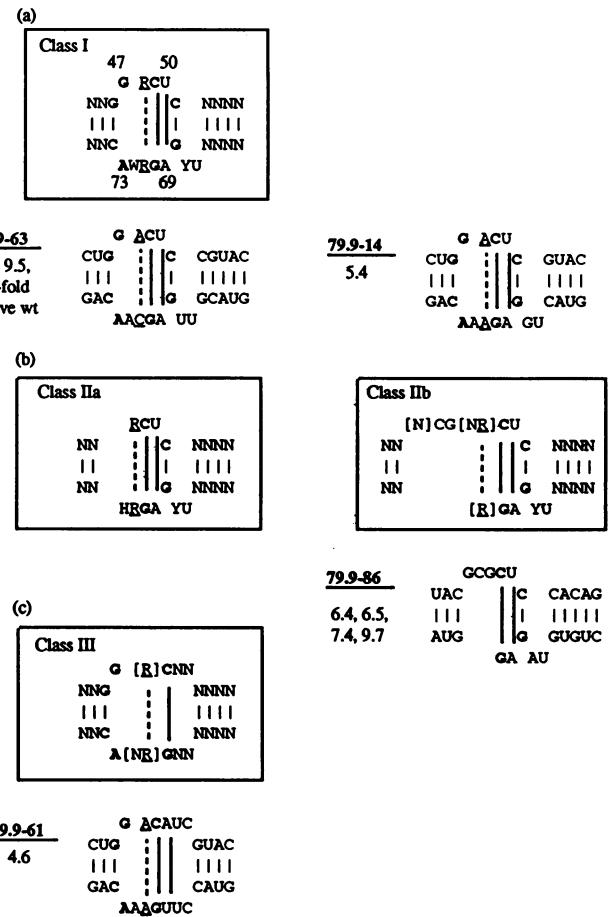


Figure 4. Schemes of Rev-binding aptamers selected from the 79.9 pool. (a) Class I comprises sequences that contain a CUC:(UY)GAG motif, a GG:CA motif, and the central homopurine base-pair. Bold and underlined bases are as in Figure 1. W indicates A or U, H indicates any base but G; Y indicates a pyrimidine; R indicates a purine; N indicates any base. Lines between N's indicate a base-pair. Numbers represent possible similarities to the wild-type RBE; the proposed homopurine pair is connected by dotted lines and is adjacent to an invariant C:G pairing. (b) Class II comprises sequences that correspond to the right-half of the Class I RBE. Class IIa has an unpaired W adjacent to the central homopurine pairing. Class IIb consists of elements that place the stem-loop portion of the minimal RBE on the left rather than right and contain a stretch of unpaired residues 5' to CUC. (c) Class III comprises sequences that do not contain a (CUC ... GAG) motif, but do contain recognition elements normally found in the left-hand side of the Class I RBE. The structures and affinities of the four best aptamers from Table II are also shown, adjacent to their Class.

include: 79.9-1 contains U rather than G47; 79.9-13 contains U:A rather than G46:C74; 79.9-63 contains A:C rather than R48:R71; and 79.9-74 contains A:U rather than G46:C74.

The remaining 79.9 aptamers represent half-molecules of the Class I RBE. Class II includes the right half of Class II starting from (but not necessarily including) the hypothesized homopurine base-pair (Class II: 79.9-3, 4, -20, -51, -53, -58, -72, -73, -86, -87, and -92) (Figure 4b). It is interesting to note that while Class I is composed almost exclusively of A:A combinations, Class II contains 4 G:G combinations. Class II aptamers frequently contain an unpaired H (any residue but G) residue at a position corresponding to U72 (Class IIa: aptamers 3, 20, 58, and 73), or from 24 unpaired residues with a preference for CG along the 'upper' strand (Class IIb: aptamers 4, 53, 72, 86, 87, and

92). Perhaps because the sequence of the 'right' flanking stem prevented residues from being readily recruited into a GG:AC motif (as occurred in the 'left' flanking stem of Class I aptamers), Class IIb is comprised of all and only those 79.9 aptamers that were selected in 'reverse' orientation; that is, these sequences contained a 5' (GAG... CUC) 3' motif assumed to be equivalent to 5' (CUC ... GAG) 3' but reversed relative to the flanking constant sequence stems. It should also be noted that a less stable alternative structure can be drawn for high-affinity Class IIb aptamers such as 79.9-#86 and 79.9-#92. This alternative structure would have only 3, rather than 4 base pairs in the left-hand stem but would maximize similarity with the wild-type structure by including bases similar to G48, G71, U72, and A73 in the internal loop.

Class III includes the left half of Class I, sometimes (4/6) including the homopurine base-pair (Class III: 22, 59, 61, 76, 79, and 85) (Figure 4c). Class III aptamers contain only a few recognizable structural features, such as a C.G base-pair that may correspond to C49:G70, and frequently contain a number of bases that appear to be unpaired. This class is defined at least in part by negation, in that it does not contain the CUC:(UY)GAG motif of Classes I and II.

Aptamers from each class were assayed for their ability to bind relative to the wild-type *RBE*. Even though aptamers such as 79.9-20 or 79.9-58 represent sequences corresponding to only the right-half of the Class I *RBE*, they nevertheless have other structural features that allow them to bind three-fold better than wild-type. Conversely, although aptamers such as 79.9-61 or 79.9-79 lack critical sequences that define the right-half of the Class I *RBE*, they make improved interactions throughout the remainder of the element that allow them to bind four-fold better than wild-type. When different selected structural features are combined in a single molecule, aptamer that bind significantly better than wild-type are generated: for example, aptamer 79.9-14 contains GG:AC, the improved A:A combination, the (CUC ... GAG) motif, and a two base pyrimidine bulge and binds 5.4-fold better than wild-type.

DISCUSSION

We have selected Rev-binding aptamers from random sequence pools based on the wild-type Rev Binding Element. Sequence analysis of these aptamer and functional comparisons with the natural element suggest that residues in the *RBE* can be parsed into two distinct sets: those that are important for high-affinity interactions with Rev in any sequence context (context-independent), and those that modulate Rev-binding in a context-dependent manner. This distinction is important both for a better understanding of the results of site-directed mutagenesis experiments, and for the design of minimal, high-affinity *RBEs*.

Residues that are either invariant among selected aptamer sequences, or that are strongly associated with high-affinity elements can be said to be context-independent. For example, we find that high-affinity *RBEs* must have a C at position 49 and a G at position 70. All aptamers from the 76.6 selection contained C49 on the 'top' strand and G70 on the 'bottom' strand of the minimal *RBE* (Figure 1), and the five most active clones selected from the 79.9 pool also have residues that are analogous to C49:G70, usually as part of the CUC:(UY)GAG motif (4/5).

Similarly, the G46G47:A73C74 motif appears to promote context-independent high-affinity interactions with Rev. Again, all aptamers from the 76.6 selection have residues corresponding

to G47 and A73, and three of the five most active clones from the 79.9 selection have also recreated the GG:AC motif. Further, in the 79.9 Class I and III aptamers, this motif is frequently (12/31) generated by deleting a constant residue (U) adjacent to the random sequence tract, and replacing it with a G; this substitution reproduces the GG:AC motif. The relative rarity of deletions in the synthesized DNA pool (typically much less than 1 % per position in studies that utilize synthetic oligonucleotides to generate site-directed mutants (37) suggests that there has been a strong selection for this motif, and, hence, that the gain in binding energy due to the formation of this motif must be substantial. Only those aptamers that fall into 79.9-Class IIa uniformly lack sequences that may correspond to G46 and G47, and these *RBEs* are the least improved in the 79.9 selection.

The selection of G46, G47, C49, G70, A73, and C74 from completely random sequence tracts indicates that they are important for high-affinity Rev binding regardless of the sequence context in which they are found; apparently no other combination of sequences at these positions will yield RNA ligands that can bind better. These invariant residues most likely form base-specific contacts with Rev, or are directly involved in positioning base-specific contacts.

Our results also support the existence of a context-independent interaction that is structure-specific, as opposed to sequence-specific. The highest-affinity aptamers from both the 76.6 and 79.9 selections contain a strong sequence covariation between positions analogous to 48 and 71 in the wild-type element. This covariation is most readily wanted for by assuming that these positions interact; given that the helical stems of the *RBE* juxtapose positions 48 and 71 on anti-parallel strands, a base-pair between these positions can be modeled as a symmetrical homopurine interaction (1; Figure 3). Since the hypothesized A:A and G:G pairings have very different numbers and arrangements of hydrogen bond donors and acceptors along their Watson-Crick faces, it is possible that Rev specifically recognizes either the phosphate backbone or the shape of the RNA at positions 48 and 71.

For the proposed homopurine pairing to occur, one of the bases must be in a *syn* conformation. When both positions 48 and 71 are purines, it is impossible to judge which base has 'flipped'. We find that several (6/36; three independent isolates) 76.6 Rev aptamers contain C48:A71. These residues can be modeled as a base pair that forms the same hydrogen bonds as, and is roughly isosteric with, the proposed homopurine interaction (Figure 3). Given this model and the fact that pyrimidine residues rarely assume a *syn* conformation (34), the preferential selection of only one of two possible C:A pairings from a pool that closely resembles the wild-type *RBE* suggests that G (or A) 71 may be *syn* in the wild-type element. The proposed C:A base-pairing is not only structurally plausible, but also appears to be functional, since the most frequently arising C:A clone (aptamer 76.6-15) has 1.7 times the activity of the wild-type *RBE*. Our model extends the work of Iwai *et al.* (22), who show that deaza-G and inosine substitutions at position 48 are active in Rev-binding. However, while this model is consistent with the data, other interpretations are also possible. Different base combinations at positions 48 and 71 may stack differently in the overall structure and thus have very different local geometries.

Homopurine pairings analogous to those shown in Figure 3 may be found in other natural RNAs. Phylogenetic analysis of an interior bulge in 16S rRNA reveals sequence covariations that can be most easily explained by G:G and A:A pairings (35).

Similarly, Parker and Siliciano (36) have recently demonstrated that guanosine residues that comprise the GU ... AG motif in eukaryotic introns may interact to form a G:G base-pair. Interestingly, the substitution of the guanosines at the splice junctions produces a phenotypic pattern that is similar to the substitution of G48 and G71 of *RBEs*: single substitutions destroy splicing or Rev-binding activity, while C:A double substitutions can restore these (very different) functions.

Interactions between Rev and the context-independent structure-specific recognition element at 48:71 appear to be modulated by base substitutions at other residues. For example, position 50 does not on its own appear to be critical for sequence-specific recognition, since it can be either G (wild-type) or A in aptamers selected from the 76.6 pool. Similarly, position 68 can be either A (wild-type) or G or C. However, positions 50, 68, and 69 covary with the identity of the hypothesized non-Watson-Crick pairings at positions 48 and 71 and appear to modestly improve the Rev-binding ability of different aptamers relative to a wild-type control. In structural terms, results from both the 76.6 and 79.9 selections can be most easily rationalized by assuming that different homopurine pairings may stack slightly differently with the invariant C49:G70 pairing. Further sequence changes are then required to compensate for these alterations in the homopurine:Watson-Crick stack, and to 'fine-tune' the presentation of contacts to the Rev protein.

While alternative hypotheses can be proposed to account for the covariations at positions 50, 68, and 69, they are not tenable. Direct contacts between Rev and each alternative sequences are unlikely, since G50:A68:G69 (wild-type *RBE*), G50:G68:U/A69 (aptamers with a G48:G71), and A50:C68:A69 (aptamers with an A48:A71) present non-overlapping networks of hydrogen bond donors and acceptors. It could also be supposed that these covariations represent negative interactions (37) that protect essential secondary structural elements in the *RRE*. For example, the dearth of C residues at position 50 and of G residues at position 69 might represent a negative interaction that avoids the shifting of bases involved in the C49:G70 pairing. The problem with this hypothesis is that, while it excludes some interactions, it does not explain why many other seemingly innocuous combinations (e.g., U50 or C69) occur only rarely or never.

These data from our *in vitro* genetic experiments complement and extend the results of site-specific mutational analyses. For example, Iwai *et al.* (22) and Kjems *et al.* (18) have constructed minimal *RBEs* that delete A68. This deletion gives a moderate (three-fold) reduction in activity, a result that is in agreement with our finding that the identity of position 68 is not critical for Rev-binding. In addition, though, we have also demonstrated that position 68 can modulate the activity of *RBE* variants, and that double to pentuple substitutions of the wild-type element that retain position 68 actually have improved Rev-binding abilities.

Our selections have identified functional sequence substitutions that would have been difficult to discover by site-directed mutagenesis alone. For example, identification of the proposed homopurine pairing between residues 48 and 71 and the attendant sequence context effects involving residues 50, 68, and 69 would have required from 16 (4²) to more than 1000 (4⁵) site-directed mutations of the wild-type *RBE*. Similarly, it is difficult to conceive of a pathway of single mutations that could have led to the discovery of the high affinity CUC:(UY)GAG motif revealed by the 79.9 selection. In fact this motif may make additional contacts with the Rev protein that have no counterpart in the wild-type element. This hypothesis is consistent with the

fact that *RBEs* that contain this motif can bind from up to tenfold (1.4 kcal/mole of binding energy) better than wild-type, irrespective of whether additional binding elements appear to be present (Figure 4, compare 79.9-63 from Class I and 79.9-86 from Class IIa).

It can be argued that our aptamers represent the best possible minimal Rev-binding sequences. The random sequence pools used in our selections initially spanned all possible stem-internal loop-stem structures that were similar in size to the wild-type *RBE*, suggesting that we have exhaustively searched the sequence space occupied by small (<40 base) *RBE* variants. For example, although only a limited number of clones from the selected 76.6 pool were analyzed, the commonality of sequence motifs that was revealed strongly suggests that we have found the optimum set of base substitutions for the 'internal loop' sequence of the wild-type *RBE*; no better set of single (10 possible) to dectuple (59,049 possible) 'internal loop' substitutions exists. Aptamers selected from the 79.9 pool also revealed strong preferences, in this case for a novel CUC:(UY)GAG motif, it is found in four of the five tightest-binding aptamers, and is present in the 79.9 Classes I, IIa, and IIb. While these data suggest that this motif is an optimal binding sequence, it is also possible that it may merely represent the most efficient route to generating high-affinity ligands for Rev. That is, if a higher affinity ligand had to 'fix' many more than the seven bases found in this motif it would necessarily be present in the initial population at a lower frequency and, hence, might not show up until more cycles of selection had been performed. However, given that many of the 79.9 aptamers we selected fixed not only these seven bases but also the homopurine pairing (2/16 combinations), and the GG:A-C motif (four bases), it is unlikely that higher affinity RNA ligands based on the internal loop structure will be found. The CUC:(UY)GAG motif did not arise during the more constrained 76.6 selection because the random sequence tracts were of shorter uniform length. Fixed sequence recognition elements (G46:C74 and C51:G67) may also have biased the 76.6 selections towards selecting Rev-binding aptamers that were similar to wild-type.

The results of our *in vitro* selections should prove relevant *in vivo*. Our minimal *RBEs* have been selected to successfully compete with a large (94 base) subsegment of the *RRE* for Rev binding *in vitro*. This subsegment (domain U), contains all sequences necessary for *in vivo* Rev responsiveness. Substitutions that impair or restore *RBE* activity *in vitro* have been introduced into a full-length *RRE* and shown to have analogous effects on Rev response *in vivo* (1, 19). *RRE* 'decoys' have also proven to be highly effective in inhibiting the replication of HIV-1 (38). Similarly, TAR mimics that interact with the IRV regulatory protein Tat have been shown to block viral replication when expressed *in vivo* (25). Indeed, HIV may prove to be uniquely susceptible to treatment by optimized aptamers decoys since *in vivo* constraints on viral structures may have prevented sequences similar to our optimized *RBEs* from arising during evolution. For example, changing C68A69, the wild-type sequence, to G68U/A69, the sequence found in aptamer containing the G:G base-pairing, changes the first two positions of a threonine codon in the envelop protein. Alternately, higher affinity *RBEs* may have been avoided because they interfere with proper *in vivo* regulation of HIV gene expression. Finally, a population of improved Rev-binding aptamers can be used to inhibit viral replication, as opposed to a single RNA species. In contrast to other pharmaceuticals, this approach will make it extremely difficult for HIV-1 to accumulate mutations that can

simultaneously evade a multitude of different RRE decays. Such an effect has already been observed by Lisiewicz *et al.* (28) with multiple antisense RNAs.

ACKNOWLEDGEMENTS

D.B. was supported by a grant from Hoechst and M.G. and M.Z. were supported by the National Institutes of Health.

REFERENCES

1. Bartel, D.P., Zapp, M.L., Green, M.R. and Szostak, J.W. (1991) *Cell*, **67**, 529–536.
2. Rosen, C.A., Terwilliger, E., Dayton, A., Sodroski, J.G. and Haseltine, W.A. (1985) *Proc. Natl. Acad. Sci. USA*, **85**, 2071–2075.
3. Emerman, M., Vazeux, R. and Peden, K. (1989) *Cell*, **57**, 1155–1165.
4. Felber, B.K., Hadzopoulou-Cladaras, M., Kaadaras, C., Copeland, T. and Pavlalkis, G.N. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 1495–1499.
5. Malim, M.H., Hauber, J., Le, S.-Y., Maizel, J.V. and Cullen, B.R. (1989) *Nature*, **338**, 254–257.
6. Hammarskjold, M.-L., Helmer, J., Hammarskjold, B., Sangwan, I., Albert, L. and Rekosh, D. (1989) *J. Virol.*, **63**, 1959–1966.
7. Zapp, M.L. and Green, M.R. (1989) *Nature*, **342**, 714–716.
8. Daly, T.J., Cook, K.S., Gray, G.S., Malone, T.E. and Rusche, J.R. (1989) *Nature*, **342**, 816–819.
9. Cochrane, A.W., Chen, C.-H. and Rosen, C.A. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 1198–1202.
10. Dayton, E.T., Powell, D.K. and Dayton, A.I. (1989) *Science*, **246**, 1625–1629.
11. Heaphy, S., Dingwall, C., Emberg, I., Gait, M.J., Green, S.M., Karn, J., Lowe, A.D., Singh, M. and Skinner, M.A. (1990) *Cell*, **60**, 685–693.
12. Holland, S.M., Ahmad, N., Maitra, R.K., Wingfield, P. and Venkatesan, S. (1990) *J. Virol.*, **64**, 5966–5975.
13. Malim, M.H., Tiley, L.S., McCarn, D.F., Rusche, J.R., Hauber, J. and Cullen, B.R. (1990) *Cell*, **60**, 675–683.
14. Olsen, H.S., Nelbock, P., Cochrane, A.W. and Rosen, C.A. (1990) *Science*, **247**, 845–848.
15. Kjems, J., Brown, M., Chang, D.D., and Sharp, P.A. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 683–687.
16. Heaphy, S., Finch, J.T., Gait, M.J., Karn, J. and Singh, M. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 7366–7370.
17. Dayton, E.T., Konings, D.A.M., Powell, D.M., Shapiro, B.A., Butini, L., Maizel, J.V. and Dayton, A.I. (1992) *J. Virol.*, **66**, 1139–1151.
18. Kjems, J., Calnan, B.J., Frankel, A.D. and Sharp, P.A. (1992) *EMBO J.*, **11**, 1119–1129.
19. Holland, S.K., Chavez, M., Gerstberger, S. and Venkatesan, S. (1992) *J. Virol.*, **66**, 3699–3706.
20. Cook, K.S., Fisk, G.J., Hauber, J., Usman, N., Daly, T.J. and Rusche, J.R. (1991) *Nucleic Acids Res.*, **19**, 1577–1583.
21. Huang, X., Hope, T.J., Bond, B.L., McDonald, D., Grahl, K. and Parslow, T.G. (1991) *J. Virol.*, **65**, 2131–2134.
22. Iwai, S., Pritchard, C., Mann, D.A., Karn, J. and Gait, M.J. (1992) *Nucleic Acids Res.*, **20**, 6465–6472.
23. Tiley, L.S., Malim, M.H., Tewary, H.K., Stockley, P.G. and Cullen, B.R. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 758–762.
24. Baltimore, D. (1988) *Nature*, **335**, 395–396.
25. Sullenger, B.A., Gallardo, H.F., Ungers, G.E. and Gilboa, E. (1990) *Cell*, **63**, 601–608.
26. Tuerk, C. and Gold, L. (1990) *Science*, **249**, 505–510.
27. Riordan, M.L. and Martin, J.C. (1991) *Nature*, **350**, 442–443.
28. Lisiewicz, J., Sun, D., Klotman, K., Agrawal, S., Zamecnik, P. and Gallo, R. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 11209–11213.
29. Ellington, A.D. and Szostak, J.W. (1990) *Nature*, **346**, 818–822.
30. Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (1987) *Current Protocols in Molecular Biology*. Greene Publishing Associates, New York.
31. Zapp, M.L., Hope, T.J., Parslow, T.G. and Green, M.R. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 7734–7738.
32. Messing, J. (1983) *Methods Enzymol.*, **101**, 20–78.
33. Karn, J., Dingwall, C., Finch, J.T., Heaphy, S. and Gait, M.J. (1991) *Biochimie*, **73**, 9–16.
34. Saenger, W. *Principles of Nucleic Acid Structure* (1984) Springer-Verlag, New York.
35. Gutell, R.R. and Woese, C.R. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 663–667.
36. Parker, R. and Siliciano, P.G. (1993) *Nature*, **361**, 660–662.
37. Couture, S., Ellington, A.D., Gerber, A.S., Cherry, J.M., Doudna, J.A., Green, R., Hanna, M., Pace, U., Rajagopal, J. and Szostak, J.W. (1990) *J. Mol. Biol.*, **215**, 345–358.
38. Lee, T.C., Sullenger, B.A., Gallardo, H.F., Ungers, G.E. and Gilboa, E. (1992) *New Biol.*, **4**, 66–74.