JOURNAL OF VIROLOGY, Dec. 1995, p. 7559–7569 Vol. 69, No. 12
0022-538X/95/$04.00+0

High-Resolution Mapping of the Human T-Cell Leukemia Virus Type 1 Rex-Binding Element by In Vitro Selection

SCOTT BASKERVILLE,[1] MARIA ZAPP,[2] AND ANDREW D. ELLINGTON[1]*

Departments of Chemistry and Microbiology, Indiana University, Bloomington, Indiana 47405,[1] and Program in Molecular Medicine, Howard Hughes Medical Institute, University of Massachusetts Medical Center, Worcester, Massachusetts 01605[2]

Received 19 May 1995/Accepted 25 August 1995

Interactions between the Rex protein of HTLV-1 and the genomic Rex-binding element (XBE) mediate the cytoplasmic transport of viral mRNAs. However, it is uncertain which RNA sequences and structures contribute to Rex recognition. A portion of the viral genome that spanned the XBE was partially randomized, and functional Rex-binding variants were selected. Alignment of selected Rex-binding sequences revealed positions that were functionally conserved between different molecules. A model is presented in which a subset of the selected residues are in direct contact with Rex. Positions that covaried with one another were also found. These covariations support a secondary-structural model in which a central paired stem is symmetrically flanked by two bulge loops. On the basis of this model, site-directed mutations of the XBE were constructed and each half molecule was found to bind independently to Rex. The functional residues and secondary structures in the XBE half molecules bear a remarkable resemblance to the transactivation response region element of HIV-1. Since the transactivation response region element is known to interact specifically with arginine residues in the Tat protein, these results suggest that the XBE binds to the arginine-rich RNA-binding domain of Rex in a similar manner. This model is supported by the selection data.

Human T-cell leukemia virus type 1 (HTLV-1) is the primary etiological agent of adult T-cell leukemia (27). In the early portion of the retroviral life cycle, regulatory factors such as Rex and Tax are synthesized, while in the late phase of infection these factors govern the accumulation of structural proteins. In particular, the Rex protein of HTLV-1 specifically binds to a portion of the 3′ untranslated region deemed the Rex-responsive element (XRRE) and promotes the cytoplasmic appearance of incompletely spliced mRNAs that encode *env* and *gag* (1, 2, 8, 11, 12, 22; see Fig. 1A).

Understanding the mechanism of Rex action requires a detailed description of the molecular interactions between Rex and the XRRE. The regions of the XRRE that interact with Rex and contribute to its function have been roughly mapped by deletion and mutation analysis (1, 2, 4, 8, 22, 24) and by modification interference analyses (5; Fig. 1b). These reports all substantially agree on the general features of the secondary structure of the XRRE and suggest that a small subsegment of stem IID is the primary Rex-binding element (XBE; 10, 22, 25). However, it is unclear what the precise structure of the XBE is, and at least three slightly different folds have been advanced (5, 11, 22). Moreover, only a few of the residues that may contribute to Rex recognition have been probed by mutational studies.

Understanding XBE-Rex interactions may also provide insights into the regulatory mechanisms of other viruses. Rex is a member of a family of RNA-binding proteins that includes human immunodeficiency virus type 1 (HIV-1) Rev and Tat, the HDV antigen, and the N protein of bacteriophage lambda (15, 17). Unlike proteins that are related by extensive sequence and structural homologies, these diverse viral factors have been grouped together because of the presence of short, concentrated patches of arginine (arginine-rich motifs [ARMs]) in their RNA recognition domains. However, while the sequence compositions of ARMs are similar, their structures are likely quite different. For example, the ARM of HIV-1 Rev spans positions 34 to 50, contains 10 arginine residues, and has been shown to form an alpha helix that can specifically recognize the Rev-responsive element even when separated from the remainder of the protein (21). Although the ARM of HTLV-1 Rex also contains numerous arginines (7 of the first 17 residues), there is little or no sequence similarity to the Rev ARM, and given the fact that 4 of the first 17 residues are prolines, this portion of Rex probably does not assume an alpha-helical configuration.

Although the sequences and structures of the various viral ARMs are different, there may be some commonality in how they recognize RNA ligands. While other arginine-rich proteins, such as histones and protamines, form nonspecific charge-charge interactions with nucleic acids, each of the ARMs tends to be specific for a given RNA sequence and structure and can recognize its cognate viral genome in a complex background of cellular RNAs. As an example, nuclear magnetic resonance studies have suggested that at least one of the arginines in the Tat ARM contributes to sequence- and structure-specific recognition of the HIV-1 transactivation response region (TAR). The positively charged guanidino head group of arginamide forms a pseudo base triple with a TAR G · C base pair and is locked in place by a refolding of the phosphate backbone of the TAR bulge-loop (19). Interestingly, sequence variants of the Rev-binding element (RBE) of HIV-1 (7) and at least one structural representation of the wild-type XBE also strongly resemble a TAR variant that shows enhanced binding to a Tat peptide (26).

A more complete definition of the XBE not only should aid in understanding the mechanism of Rex action but also may provide more general insights into how ARMs bind to their RNA substrates. Unfortunately, a prohibitive number of XBE mutants would be needed to determine completely which sequences and structures are essential for Rex recognition. For

* Corresponding author. Phone: (812) 855-6071. Fax: (812) 855-8300.

example, examining alternative base-pairing schemes for the different secondary-structural models that have been proposed would require that upwards of 40 single and 128 double substitutions be constructed and assayed. In the absence of such an extensive set of phylogenetic or mutational data, in vitro genetic selections have proven to be a useful tool for high-resolution structural mapping of functional nucleic acids. By synthesizing random-sequence populations, the relative binding abilities of extremely large numbers of sequence substitutions can be assessed in parallel. As an example, the primary- and secondary-structural features of the wild-type RBE were, in large part, delineated by analyzing RNA aptamers (binding sequences) selected from a partially randomized library (3). Similarly, sequence covariations that supported the existence of numerous base-base interactions were identified by analyzing Rev-binding sequences isolated from a completely random library (7, 9, 23a). These covariations were, in turn, used as distance constraints for a three-dimensional structural model of the RBE (13). In the current study, we selected Rex-binding RNAs from a partially randomized sequence population based on the XBE. Analysis of the selected sequences identified sequences and structures that are important for Rex recognition.

MATERIALS AND METHODS

Materials. Stem IID of the Rex-responsive element was transcribed from *Xho*I-cut pGEM-XRRE. This plasmid contains residues 144 through 223 of the XRRE (5); in addition to these residues, some polylinker sequences were also transcribed.

Bacterial expression of Rex. pGex-Rex, a glutathione-*S*-transferase (GST)–Rex fusion protein bacterial expression plasmid, was constructed by using a 252-nucleotide DNA fragment that encodes amino acids 1 to 80 of the HTLV-1 Rex protein and contains a TGA stop codon immediately following the coding sequence. The insert was prepared by PCR amplification and contains a unique *Bam*HI site at the 5′ end and a unique *Eco*RI site at the 3′ end. PCR products were digested with *Bam*HI and *Eco*RI, gel purified, and ligated into *Bam*HI-*Eco*RI-cut pGex-3X (Pharmacia, Piscataway, N.J.). Colonies containing inserts were identified, and a single clone containing an insert of the correct size was sequenced and used for protein expression.

GST-Rex induction, purification, and assay. A single colony was inoculated into 100 ml of 2× YT/Amp broth (19) and grown overnight at 37°C. A portion (25 ml) of this saturated culture was used to inoculate 1 liter of 2× YT/Amp broth. Cultures were incubated at 37°C with aeration for approximately 2.5 h (optical density at 595 nm, 0.65). Protein expression was induced by addition of isopropyl-β-D-thiogalactopyranoside to a final concentration of 1 mM, and growth continued for 3 h at 37°C. After this time, cells were collected by centrifugation, resuspended in 25 ml of RS buffer (20 mM Tris-Cl [pH 7.4], 1 M NaCl, 0.25 mM EDTA, 0.1 mM β-mercaptoethanol), and lysed by sonication at 10°C. Cell debris was removed by centrifugation, and 10 ml of the bacterial lysate was loaded directly onto a 1.8-ml glutathione-agarose column equilibrated in RS buffer. The initial eluate was reapplied to the column to optimize depletion of the fusion protein from the bacterial lysate. The column was washed with 10 column volumes of RS buffer and then eluted with 0.5 M glutathione, pH 8.0. The elution profile was monitored by using Bradford assays; fractions containing protein of the correct molecular weight were detected by denaturing sodium dodecyl sulfate-polyacrylamide gel electrophoresis followed by silver staining. Peak fractions were pooled and dialyzed for 4 h at 4°C in 500 ml of RBB buffer (50 mM *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid [HEPES; pH 7.9], 40 mM KCl, 10% glycerol) with two buffer changes. Following dialysis, the protein was analyzed by gel electrophoresis and judged to be 90% pure. Samples were aliquoted and frozen at −80°C; the same preparation was used for all of the experiments described in this report.

The GST-Rex fusion protein was assayed for the ability to bind the XRRE specifically. A number of individual ^{32}P-radiolabeled RNA probes were prepared by in vitro transcription with T7 RNA polymerase (vide infra). Each RNA probe was incubated with 25 ng of GST-Rex in 20 μl of RBB buffer containing 500 ng of yeast tRNA and 4 U of RNAsin (Promega, Madison, Wis.). Samples were incubated for 20 to 30 min at 4°C and then loaded directly onto a native gel (6% polyacrylamide, 60:1 acrylamide-bisacrylamide ratio; 0.5× TBE).

Random sequence RNA pools. A doped-sequence pool of RNA molecules centered on the wild-type XBE was used as a starting point for in vitro selection. A chemically synthesized DNA oligonucleotide was used to generate the doped-sequence RNA pool. This oligonucleotide contained constant sequences that corresponded to XRRE positions 153 to 168 (5) at its 5′ end, followed by XRRE positions 169 to 211 containing 70% wild-type residues and 10% of each non-wild-type residue, followed by constant-sequence XRRE positions 212 to 229 at its 3′ end (see Fig. 1A). The crude oligonucleotide (≈14.3 pmol or 8.6 × 10^{12} nucleotides) was amplified in a 100-μl PCR (10 mM Tris-Cl [pH 8.3], 50 mM KCl, 1.5 mM $MgCl_2$, 0.5 μM primer DX.1 [TAATACGACTCACTATAGGGA GATCTAGGTAAGTTTAAA], 0.5 μM primer DX.2 [CCGGCTGAGTCTAG GTAG], 200 μM deoxynucleoside triphosphates, 5% acetamide, 0.05% Nonidet P-40, 2.5 U of *Taq* polymerase) for eight cycles of 94°C for 1 min, 45°C for 1 min, and 72°C for 2 min. Approximately 0.5 μg (or 1 pool equivalent) of the amplified template was added to an Ampliscribe (Epicentre Technologies, Madison, Wis.) transcription reaction. Roughly 30 μg of the initial RNA pool was obtained following purification on a denaturing polyacrylamide gel (6). The initial sequence substitution bias was determined by sequencing 10 clones from the unselected population. The substitution frequencies were as follows: G→G, 76.5%; G→A, 8.0%; G→T, 8.0%; G→C, 7.4%; A→A, 65.0%; A→G, 17.5%; A→T, 10.0%; A→C, 7.5%; T→T, 67.4%; T→G, 14.6%; T→A, 6.7%; T→C, 11.2%; C→C, 59.0%; C→G, 21.5%; C→A, 9.4%; C→T, 10.1%. Skewing away from programmed frequencies has been previously observed (28).

TABLE 1. Selection conditions[a]

Cycle(s)	[tRNA] (μM)	[Doped XRRE] (μM)	Amt of Rex (ng)	Prefiltration[b]
1	10	1	~30	3×
2	6.6	1.2	~30	
3, 4	6.6	1.2	~30	3×

[a] The amounts and concentrations of protein and RNA used in each cycle of selection are shown.

[b] Passing the pool over the nitrocellulose membrane multiple times prior to each cycle of selection prevents the accumulation of any sequences that can interact with the filter.

Selection conditions. In each cycle, tRNA (Boehringer Mannheim, Indianapolis, Ind.) and Rex protein were incubated together for 10 min at ambient temperature in 1× binding buffer (50 mM Tris-Cl [pH 8.0], 50 mM KCl; 20 μl in the first cycle and 30 μl in subsequent cycles). Pooled RNAs in 1× binding buffer (20 μl in the first cycle and 30 μl in subsequent cycles) were separately heated to 90°C for 2 min, cooled to ambient temperature over a 10-min period (to equilibrate conformers), passed over HAWP 25 modified cellulose filters (Millipore, New Bedford, Mass.) to remove filter-binding sequences, and then added to the protein mixture. The final reaction mixture (40 μl in the first cycle and 60 μl in subsequent cycles) was left for an additional 60 min. The amounts of tRNA, Rex, and pooled RNA were varied in subsequent cycles as detailed in Table 1.

RNA-protein complexes were separated from free RNA by vacuum filtration (5 in. [1 in. = 2.54 cm; 1 mm Hg = 133.322 Pa] Hg) over modified cellulose filters. Following application of the RNA-protein mixtures, filters were washed twice with 500 μl of 1× binding buffer. Bound RNAs were eluted from filters with 400 μl of 2× PK buffer (0.2 M Tris-Cl [pH 7.6], 2.5 mM EDTA, 0.3 M NaCl, 2% sodium dodecyl sulfate) at 75°C for 15 min in every cycle. The eluate was extracted with phenol-chloroform and precipitated with ethanol, and the pellet was resuspended in 25 μl of water.

Amplification. RNA for subsequent cycles was synthesized by a combination of reverse transcription, PCR amplification, and in vitro transcription. A portion of the extracted RNA (10 μl) was reverse transcribed in a reaction mixture (20 μl) that contained 40 mM KCl, 50 mM Tris-Cl (pH 8.0), 6 mM $MgCl_2$, 0.4 mM deoxynucleoside triphosphates, 2.5 μM primer DX.2, 5 mM dithiothreitol, and 5 U of avian myeloblastosis virus reverse transcriptase (Seikagaku, St. Petersburg, Fla.). Nucleotides, dithiothreitol, and the enzyme were added only after an initial annealing step (3 min at 75°C followed by 5 min at ambient temperature). The reaction mixture was then incubated at 42°C for 45 min. For PCR amplification, 15 μl of this reaction mixture was diluted in 85 μl of a PCR mixture (10 mM Tris-Cl [pH 8.3], 50 mM KCl, 1.5 mM $MgCl_2$, 5% acetamide, 0.05% Nonidet P-40, 200 μM deoxynucleoside triphosphates, 2.5 U of *Taq* polymerase, 0.5 μM each DX.1 and DX.2). The reaction mixture was cycled at 94°C for 1 min, 45°C for 1 min, and 72°C for 2 min for as many cycles as needed to produce a product band of the correct size. RNA was synthesized from the PCR template in an Ampliscribe transcription reaction mixture and purified on a denaturing polyacrylamide gel for subsequent cycles of selection.

Cloning and sequencing. PCR products from the fourth cycle of selection (1 μg) were directly ligated into a TA cloning vector (Invitrogen, San Diego, Calif.) or pCRScriptSK(+) (Stratagene, La Jolla, Calif.) in accordance with the protocol provided. Clones were sequenced by using standard protocols for dideoxy incorporation. Since only a few deletions were present in the selected molecules, residues that were included within the partially randomized region were manually aligned. To determine whether the distribution of residues at a given position had been significantly skewed by selection, the probability of obtaining that distribution by chance (assuming no selection) was determined. For example, at position 176, 46 wild-type (U) residues and 14 non-wild-type residues were expected. In fact, 52 occurrences of uridine and 8 occurrences of nonuridine residues were observed. The probability of obtaining the observed distribution in

the absence of selection can be calculated by using a cumulative binomial distribution based on the formula

$$\sum_{x=\#\mathrm{wt}}^{\mathrm{total}} (f\mathrm{wt})^{\#\mathrm{wt}}(f\mathrm{non\text{-}wt})^{\#\mathrm{non\text{-}wt}}\{\mathrm{total}!/[(\#\mathrm{wt})!\,(\#\mathrm{non\text{-}wt})!)]\},$$

where *f*wt is the observed sequence bias for a wild-type nucleotide in the initial pool, expressed as a fraction (e.g., 0.67 of the initial population contained uridines at a position that was uridine in the wild type); #wt is the number of clones containing a wild-type nucleotide at a given position; *f*non-wt is the observed sequence bias for non-wild-type nucleotides in the initial pool, expressed as a fraction, which is 1 − the original frequency (e.g., 1 − 0.67 = 0.33 of the initial population contained nonuridines at a position that was uridine in the wild type); #non-wt is the number of clones containing a non-wild-type nucleotide at a given position; and total is the total number of clones analyzed from the final cycle of selection (e.g., 60). For uridine, these values are

$$\sum_{x=52}^{60} (0.67)^x(0.33)^8[60!/(x!8!)] \approx 0.0006,$$

which means that this result is expected to occur in 6 of 10,000 times by chance. Observed distributions that would have occurred at $P \leq 0.05$ (i.e., less than or equal to 1 in 20 times by chance) were assumed to be significant for this study.

Minimal XBE constructs. To generate minimal XBEs, DNA templates with extensive self-complementarity had to be synthesized. PCR amplification of sufficient quantities of these templates for in vitro transcription reactions proved to be difficult. Therefore, the following strategy was employed. First, small amounts of short, double-stranded oligomers were cloned into plasmids. Next, larger (>1 μg) amounts of longer, double-stranded oligomers were generated from individual plasmid clones via PCR. Finally, these oligomers were enzymatically cleaved to produce suitable templates for runoff transcription of short RNAs.

As a specific example, a chemically synthesized template for bulge-right (9.3 μM final concentration; 5′ GGGCAATTGCTCAGGAAGAGCATGCTCCCT TGGAGCAATTGAATT 3′) and its primer (94 μM final concentration; 5′ AAT TCAATTGCTCCAAG 3′) were mixed in 1× Klenow buffer (20 μl; New England Biolabs, Beverly, Mass.), heated to 100°C for 2 min, and slowly cooled to 40°C over a 30-min period. After annealing was complete, Klenow fragment (5 U; New England Biolabs), *Escherichia coli* single-stranded binding protein (1 μg; U.S. Biochemicals, Cleveland, Ohio), and deoxynucleoside triphosphates (90 μM) were added and the reaction mixture was incubated at 25°C for 15 min. Full-length, double-stranded oligomers were gel isolated and incubated with *Taq* polymerase (5 U) in a standard PCR mixture (40 μl). Any incompletely elongated oligomers were filled in during this procedure and, more importantly, nontemplated deoxyribosyladenine residues were added to the 3′ ends of the DNA. A portion (1/40) of the reaction mixture was cloned into the TA vector. Clones with the correct insert were identified by sequencing. Double-stranded DNAs were generated from clones by PCR amplification in the presence of a 5′ primer containing the T7 promoter (5′ TAATACGACTCACTATAGGGCAAT TGCTCAGGAAG 3′) and a 3′ primer that was over 70 bases beyond the 3′ end of the insert (5′ ATTTAGGTGACACTATA 3′). An internal *Mse*I site was present at the junction between the insert and vector DNAs. PCR products were restricted with *Mse*I, and RNAs were synthesized by runoff transcription.

Filter-binding assays. Pools and minimal RNAs were assayed for the ability to be coretained with the Rex protein on modified cellulose filters. Unselected and selected RNA pools were internally labeled by including 1 μl of [α-^{32}P]UTP (3,000 Ci/mmol; Dupont NEN, Boston, Mass.) in the Ampliscribe reaction mixture. Following thermal equilibration (90°C, 2 min; ambient temperature, 10 min), pooled RNAs (0.8 μM) were added to a mixture of tRNA (8 μM final concentration) and Rex protein (~30 ng) in 50 μl of 1× binding buffer, and the binding reaction was allowed to proceed for 1 h at room temperature. The mixture was then filtered with modified cellulose, and the filters were washed twice with 500 μl of 1× binding buffer. The amount of radioactive RNA that was coretained with protein was quantitated with a PhosphorImager (Molecular Dynamics, Sunnyvale, Calif.). Background binding to the filter alone (generally less than 1% of the applied pool) was independently determined and subtracted from the protein-dependent signal.

In competition experiments, pooled or cloned RNA (0.6-μM final concentration) was incubated with an equimolar amount of stem IID of the XRRE (0.6-μM final concentration), a 10-fold excess of tRNA (6 μM), and ~30 ng of active Rex protein in 1× binding buffer (60 μl). The reaction mixture was incubated, filtered, and eluted as in the selection experiments. After precipitation, samples were resuspended in 4 μl of stop dye (7 M urea, 1× TBE, 0.1% bromphenol blue) and electrophoretically separated on a denaturing 10% polyacrylamide gel. The amount of radioactivity in each band was determined with a PhosphorImager. Binding ratios were calculated by using the formula [(no. of counts filtered$_{\mathrm{pool}}$ − background)/(no. of counts unfiltered$_{\mathrm{pool}}$ − background)]/[(no. of counts filtered$_{\mathrm{XRRE}}$ − background)/(no. of counts unfiltered$_{\mathrm{XRRE}}$ − background)]. The relative specific activities of samples cancel one another in this equation and do not influence the binding ratio. If the binding reaction has gone to equilibrium, the binding ratio represents a ratio of dissociation constants for different RNA-protein complexes; the samples were equilibrated for multiple half-lives of the wild-type complex. Minimal XBEs were assayed in a similar fashion, except that the RNAs were labeled with polynucleotide kinase (New England Biolabs) and [γ-^{32}P]ATP (3,000 Ci/mmol; Dupont NEN, Boston, Mass.).

RESULTS

Interactions between GST-Rex and XRRE RNA are specific. To determine whether GST-Rex binds specifically to the high-affinity site within the XRRE, its ability to bind both cognate and noncognate RNAs was examined. The fusion protein retarded the mobility of the XRRE (residues 144 through 223) efficiently but failed to form a complex with either the stem II domain of the HIV-1 RRE (residues 36 to 106, numbered in accordance with reference 16), the HIV-1 TAR RNA element, or antisense XRRE RNA (data not shown).

Additionally, the XRRE RNA was tested for the ability to interact with several unrelated sequence-specific RNA-binding proteins by using a similar assay. When XRRE RNA was incubated with either purified HIV-1 Rev protein, mammalian splicing factor U2AF^{65}, or a synthetic peptide containing amino acids 46 to 60 of HIV-1 Tat, no stable complex formation was observed (data not shown). Similar results were obtained when purified GST was used in this assay. Taken together, these results suggest that GST-Rex binds specifically to the XRRE in vitro.

In vitro selection of Rex-binding RNAs. The boundaries of the primary XBE of the XRRE have been mapped by deletion analysis and modification interference (1, 4, 22; Figure 1B). To examine which residues in the XBE supported interactions with the Rex protein, the entire XBE and flanking sequences were partially randomized. The level of randomization that was chosen, 30% non-wild-type substitution per position, allowed all possible single mutational variants to nearly all ($P < 0.01$) hextuple mutational variants of the XBE to be simultaneously assessed for function (3, 9).

RNA molecules that could bind to Rex were iteratively selected from the randomized pool. Only 0.6% of this original pool was coretained with Rex on a modified cellulose filter, indicating that most of the random sequence substitutions altered wild-type RNA sequences or shapes necessary for binding to Rex. Nonetheless, the small fraction of the RNA population that bound to Rex was eluted, amplified, and reselected in subsequent cycles. After only four cycles of selection and amplification, 5.4% of the RNA population was coretained with Rex on the filter; protein-independent binding to the filter alone did not similarly increase.

To assess whether internal competition between RNA variants had produced sequences that could bind Rex as well as the wild-type XBE, competitive filter-binding assays were carried out with the selected pool and stem IID of the XRRE. The affinity of the pool for Rex improved by 20-fold during the course of the selection and bound ninefold better than the wild-type (Fig. 2). Assays with individual clones from the fourth cycle also showed that selected sequences could effectively compete with the wild-type XBE for binding to Rex (Fig. 3).

Residues critical for Rex recognition. The selected RNA population represented a subset of the original pool that retained a common function, the ability to interact with the Rex protein. To determine what sequence or structural motifs were important for Rex recognition, individual RNAs from the unselected pool and from the fourth cycle of selection were cloned, sequenced, and compared with one another. All of the sequences were still similar enough to the wild type that they

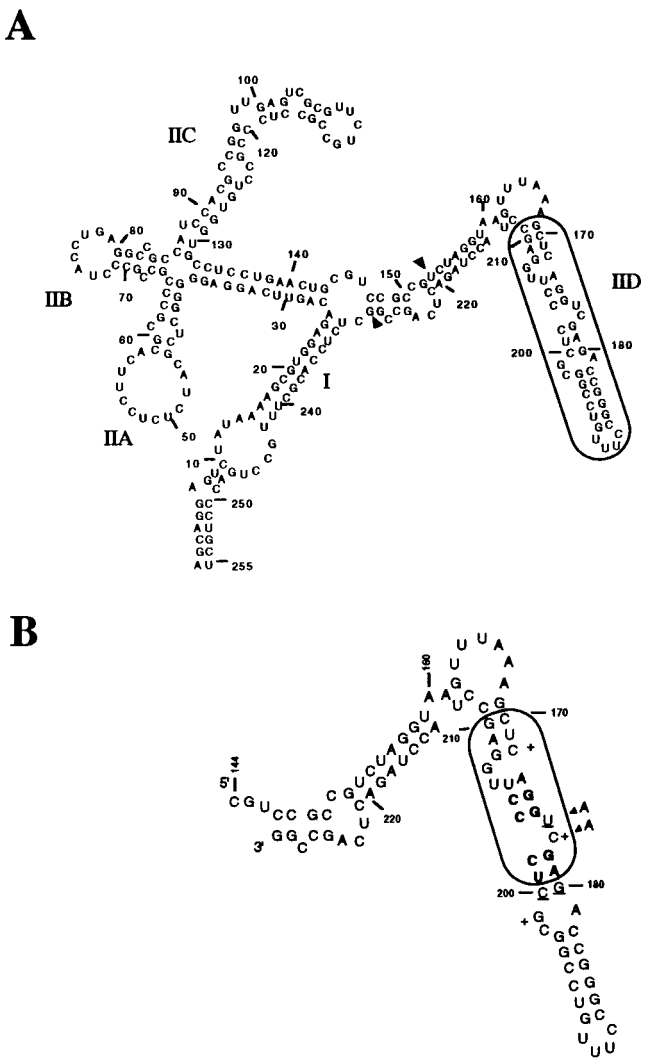

FIG. 1. Predicted secondary structures of the XRRE and the XBE. (A) Predicted secondary structure of the XRRE. A secondary-structural fold of the HTLV-1 XRRE based on that of Bogerd et al. (5) is shown. The region of the XRRE which was randomized in the initial pool (positions 169 to 211 [numbered in accordance with reference 5]) is boxed. The arrows at positions 153 and 229 indicate the start and stop of the in vitro transcription products. (B) Predicted secondary structure of the XBE. Based on available data from deletion studies and modification interference analyses, the boxed region of stem IID represents the best estimate of the primary XBE of the XRRE. This region is essential for function in vivo (1, 10). Modification of boldface residues with either hydrazine or diethylpyrocarbonate strongly interfered with Rex-binding ability, while modification of underlined residues interfered weakly with binding. Plus signs indicate positions where modification enhances binding (5). Arrows indicate a double substitution that drastically reduces function in vivo (22).

could be easily aligned (Fig. 3). As expected, clones from the unselected pool were found to contain numerous randomly distributed sequence substitutions. Many positions in the 60 selected clones also contained multiple different sequence substitutions relative to the wild type, indicating that these positions are functionally neutral (Fig. 4). For example, position 173 is A in the wild-type binding element but varies readily between A (39 times; 65%), G (12 times; 20%), C (6 times; 10%), and U (3 times; 5%) in selected clones. However, even a cursory examination of the alignment showed that a few of the positions in the selected sequences were highly conserved between different clones (e.g., G-178); these positions were expected to be most important for binding to Rex.

Determination of the statistical significance of conservation or substitution of each position in the selected sequences was based on comparisons with unselected sequences. Many of the 43 partially randomized positions were found at significantly higher frequencies than expected by chance. Alteration of these residues presumably disrupts the formation of stable RNA-protein complexes, and hence, sequences containing substitutions at these positions were lost during the course of the selection. For example, G-178 is present in 57 of 60 molecules; since G-178 was originally present in the randomized population at a frequency of roughly 0.76 per position (slightly higher than the programmed value of 0.7), the probability that this residue was conserved among the various clones by chance is less than

$$\sum_{x=57}^{60} (0.76)^x (0.24)^8 [60!/(x!3!)] \approx 0.00009.$$

In addition, a number of other positions were significantly conserved ($P \leq 0.05$ for occurrence of the observed distribution by chance), including G-169, U-171, C-172, U-176, G-178, G-196, C-202, C-203, U-206, and G-208. A summary of these conserved positions is shown in Fig. 4.

Secondary structures critical for Rex recognition. A number

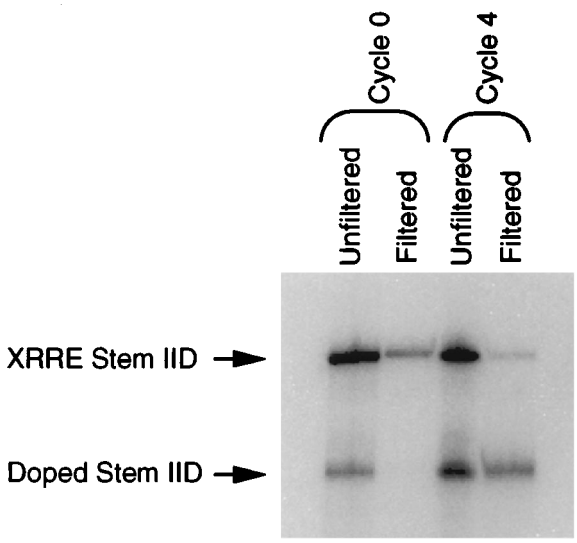

FIG. 2. Improvement in the Rex-binding ability of the randomized pool during the course of selection. The binding ability of the randomized pool relative to that of stem IID of the XRRE was determined as a function of the cycle. Pooled RNA and stem IID were mixed with the Rex protein and allowed to reach equilibrium. Bound RNAs were separated from unbound RNAs by filtration, and the two types of RNA were separated from one another on a denaturing polyacrylamide gel. The positions of the two RNAs are indicated by arrows. Since RNA was in excess over Rex during the binding reactions, the pool should have competed with the XRRE for available protein. Therefore, the relative intensities of the bands are a rough indication of how well the RNAs could compete with one another for binding. A more precise estimate of the binding ratio was obtained by determining the ratio of the two RNAs before and after filtration with the equation [(no. of counts filtered$_{pool}$ − background)/(no. of counts unfiltered$_{pool}$ − background)]/[(no. of counts filtered$_{stem\ IID}$ − background)/(no. of counts unfiltered$_{stem\ IID}$ − background)] (7). This value should represent the relative dissociation constants of pooled and wild-type RNAs for Rex. At the beginning of the selection (cycle 0), the binding ratio is 0.46, similar to the background binding observed with other RNAs. At the conclusion of the selection (cycle 4), the binding ratio is 9, indicating substantial improvement in the ability of the final population to bind Rex.

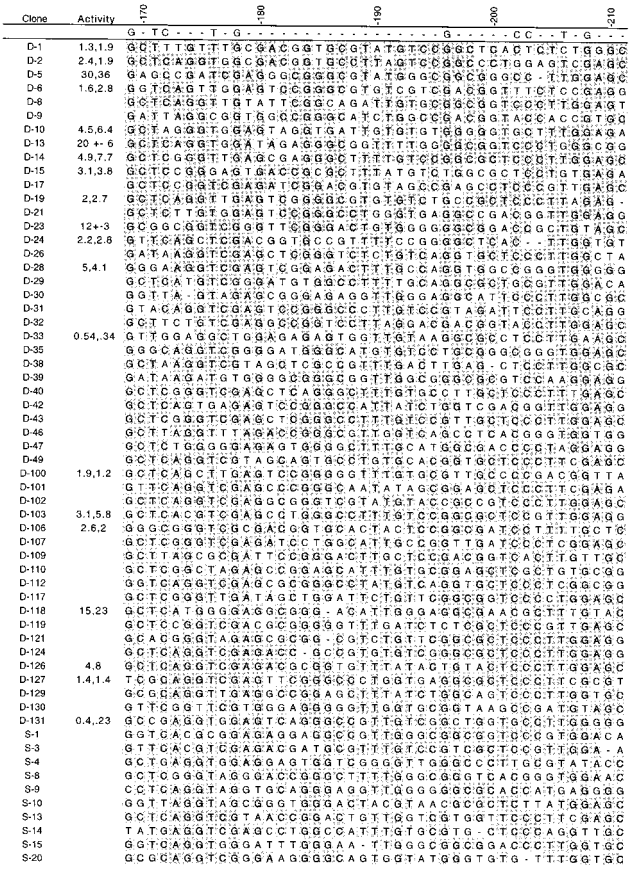

FIG. 3. Sequences of selected RNAs. Manual alignment of 60 clones from cycle 4. A clone is the name of a selected sequence. Activity is the activity of that sequence relative to the wild-type XBE. The results of duplicate assays are shown separately; the results of three assays were averaged, and the deviation is shown. The position of a residue within the XRRE is shown along the top. Residues that were significantly conserved ($P \leq 0.05$) between clones are also shown along the top of the alignment; residues that were not significantly conserved are represented by dashes. Within the sequences themselves, residues that are identical to the wild type are shaded and dashes indicate deletions.

of different secondary-structural models for the XBE have been advanced (5, 11, 22). On the basis of these models, many of the positions that have been shown to be conserved among selected RNAs are potentially involved in Watson-Crick base pairs. The artificial phylogeny of selected Rex-binding sequences allowed the functional significance of a predicted base pair between positions to be assessed in much the same way that the functional significance of individual residues was assessed. Many of the predicted base pairings occurred at frequencies greater than those expected by chance ($P \leq 0.05$; Fig. 4 and Table 2). In some cases, residues seemed to be functional only in terms of the ability to support a particular secondary-structural model. For example, while neither G-174 nor C-204 was significantly ($P \leq 0.05$) conserved as an individual residue, the residues at these positions occurred together frequently enough that the probability of the base pair was statistically significant. Similarly, while G-175 and A-209 are not conserved as individual residues, they could nonetheless form statistically significant base pairings (e.g., G-175 · C-203 and U-171 · A-209). Finally, statistically significant base pairings can form between residues that are also important individually, such as G-178 · C-202 and C-172 · G-208.

The rate of mutagenesis was high enough that double substitutions between positions could be observed in the selected clones; in many cases, the identities of these double substitutions confirmed the prediction that a Watson-Crick pair was formed. For example, while the preponderance of bases found at positions 171 and 209 could form a wild-type U · A base pair, in two clones a G · C base pair could have formed and in one clone a C · G base pair could have formed (Table 2). Unfortunately, in some cases the level of conservation (and the corresponding lack of sequence variation) made it difficult to confirm a postulated base pair (e.g., G-178 · C-202).

Testing of structural models with minimal binding elements. The sequence and structural features of the XBE defined by in vitro selection (Fig. 4) were further assessed by constructing a series of site-directed mutations of the RNA. In particular, since the selection data clearly revealed the XBE to be symmetrical with a twofold axis of symmetry through G-174 · C-204, the relative contribution of each half molecule to Rex recognition was probed in two ways. First, the halves of the XBE were separately synthesized. Each half molecule contained the central (AGG. . .CCU) paired stem and sequences either to the left or to the right of this stem (orientation as in Fig. 5). The left-half molecule spanned wild-type residues 169 to 175 and 203 to 211 (Fig. 5, min-left), while the right-half molecule spanned residues 173 to 180 and 200 to 205 (Fig. 5, min-right). Next, variants of the full-length XBE in which structural features such as paired stems or bulge loops were asymmetrically altered were constructed (Fig. 5, stem-left, stem-right, bulge-left, and bulge-right). These RNAs were all tested for the ability to compete with the wild-type stem IID for binding to Rex.

RNAs that contained the right half of the XBE bound Rex better than did molecules that contained the left half. For example, the right half of the XBE (min-right) bound as well as the wild-type control (min-wt) and almost threefold better than a construct that contained only the left half of the binding site (min-left). Similarly, alterations of right-half stem or bulge regions (stem-right and bulge-right, respectively) had a greater effect on binding than did disruptions of left-half stem or bulge regions (stem-left and bulge-left, respectively). However, the apparent functional asymmetry of the XBE is tempered by the fact that almost all of the RNA variants retained substantial Rex-binding activity. Only mutation of the right-half bulge loop (bulge-right) severely interfered with Rex recognition. Surprisingly, deletion of the right-half bulge loop was less deleterious than its alteration (compare bulge-right and min-left).

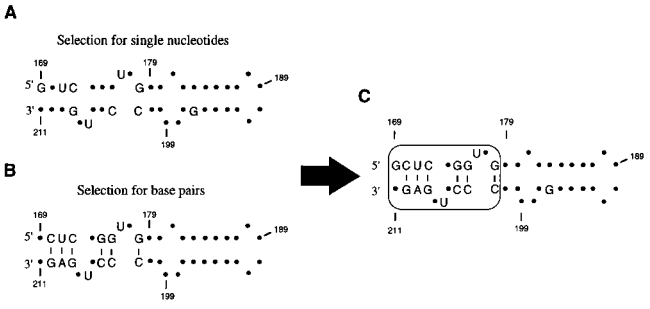

FIG. 4. Summary of selected positions and pairings. (A) Selected positions. Residues within the randomized region that were shown to be significantly conserved ($P \leq 0.05$) in Fig. 3 are mapped onto a proposed secondary structure. (B) Selected pairings. Base pairings within the randomized region that were shown to be significantly conserved ($P \leq 0.05$) are mapped onto a proposed secondary structure. (C) The two sets of data are superimposed. The boxed region spans the primary XBE of the XRRE.

TABLE 2. Canonical and wobble pairs from the final cycle of selection

WT sequence position	No. of occurrences[a]														
	G-169 · C-211	C-170 · G-210	U-171 · A-209	C-172 · G-208	A-173 · U-205	G-174 · C-204	G-175 · C-203	G-178 · C-202	A-179 · U-201	G-180 · C-200	C-182 · G-197	C-183 · G-196	G-184 · C-195	G-185 · C-194	G-186 · U-193	C-187 · G-192
Watson-Crick pairs																
G · C	31	1	2	0	1	35	40	45	1	26	0	1	27	20	4	1
C · G	1	35	1	41	2	2	1	0	0	0	19	26	1	1	0	29
A · U	0	1	0	0	28	0	0	0	31	1	0	0	0	0	1	0
U · A	0	0	33	0	0	0	0	0	0	0	1	1	1	1	0	1
Wobble pairs																
G · U	2	1	2	3	9	4	1	5	2	1	0	2	7	5	31	0
U · G	0	5	2	6	0	2	1	0	1	3	6	1	1	1	1	3

[a] Numbers indicate the number of times each type of pairing was observed in the final cycle of selection based on the secondary structure of Bogerd et al. (5). Significantly conserved ($P \leq 0.05$) pairings are boxed. Cutoff numbers for $P \leq 0.05$, based on the observed bias found in the initial pool, are as follows: G · C, 34; A · U, 33; G · U, 38.

DISCUSSION

Few mutational studies of the XBE have been carried out, and only a few residues that contribute to Rex recognition have previously been identified. Similarly, the limited number of phylogenetic variants of HTLV-1 that have been sequenced do not firmly establish either the sequence or the structure of the XBE. Therefore, to generate a wide spectrum of sequence substitutions for analysis, the XBE was partially randomized. The resultant RNA population spanned more than 10^{12} different sequences and contained all possible single to nearly all hextuple mutational combinations of the wild-type XBE. This approach allowed many more sequence variants to be generated than would have been possible by using site-directed mutagenesis. For example, while the 129 single substitutions of the XBE could have been individually synthesized, it would have been very difficult to create all of the 8,127 double substitutions, much less the higher mutational combinations.

RNA variants that could bind to Rex were iteratively selected from the randomized population. The selection primarily removed RNAs from the population that contained deleterious substitutions. However, since RNA molecules were present in at least 10-fold molar excess over Rex in each cycle (Table 1), species competed with one another for binding to Rex and the best binding variants were preferentially retained. As the selection proceeded, the population should have been skewed towards those RNA variants with the highest binding affinities. This was found to be the case. Although the original RNA population bound poorly to Rex, after four cycles of selection and amplification the aggregate binding ability of the population had improved by 20-fold and the population as a whole bound Rex ninefold better than does the wild-type XBE. This value is similar to that obtained by averaging the Rex-binding activities of 20 individual, selected sequences (sevenfold).

The functional improvements observed during the selection of Rex-binding RNAs are similar to those that have been found in selections that targeted other viral regulatory proteins. For example, a Rev-binding RNA population selected from a completely random sequence pool showed 31-fold improvement in binding during the course of the selection, and the final population bound 3-fold better than the wild-type RRE (7, 23a). A Rev-binding RNA population selected from a partially randomized RNA pool also bound approximately threefold better than the wild-type sequence (3). These values are also consistent with selections for other types of RNA-binding proteins, such as T4 polymerase (23).

The final population should represent a subset of XBE variants that retain the ability to bind Rex. When individual sequences were assayed, 18 of 20 bound Rex as well as or better than wild-type XBE. Moreover, since the selected pool and many individual clones bind Rex better than does wild-type stem IID, some residues in selected sequences may form additional or different contacts with Rex. The selection of Rex-binding motifs other than the wild-type XBE would not be unprecedented: in vitro selections have previously identified novel nucleic acid binding motifs in addition to wild-type sequences (7, 23, 23a). In a separate study, we have found that Rex-binding aptamers derived from completely random sequence pools give rise to a variety of RNA sequences and shapes, many of which bind Rex better than does the XBE (3a).

Selected sequences were aligned to form an artificial phylogeny of binding species. Just as natural RNA sequences have been aligned and compared to probe RNA structure and function, so too can Rex-binding RNAs be compared to determine

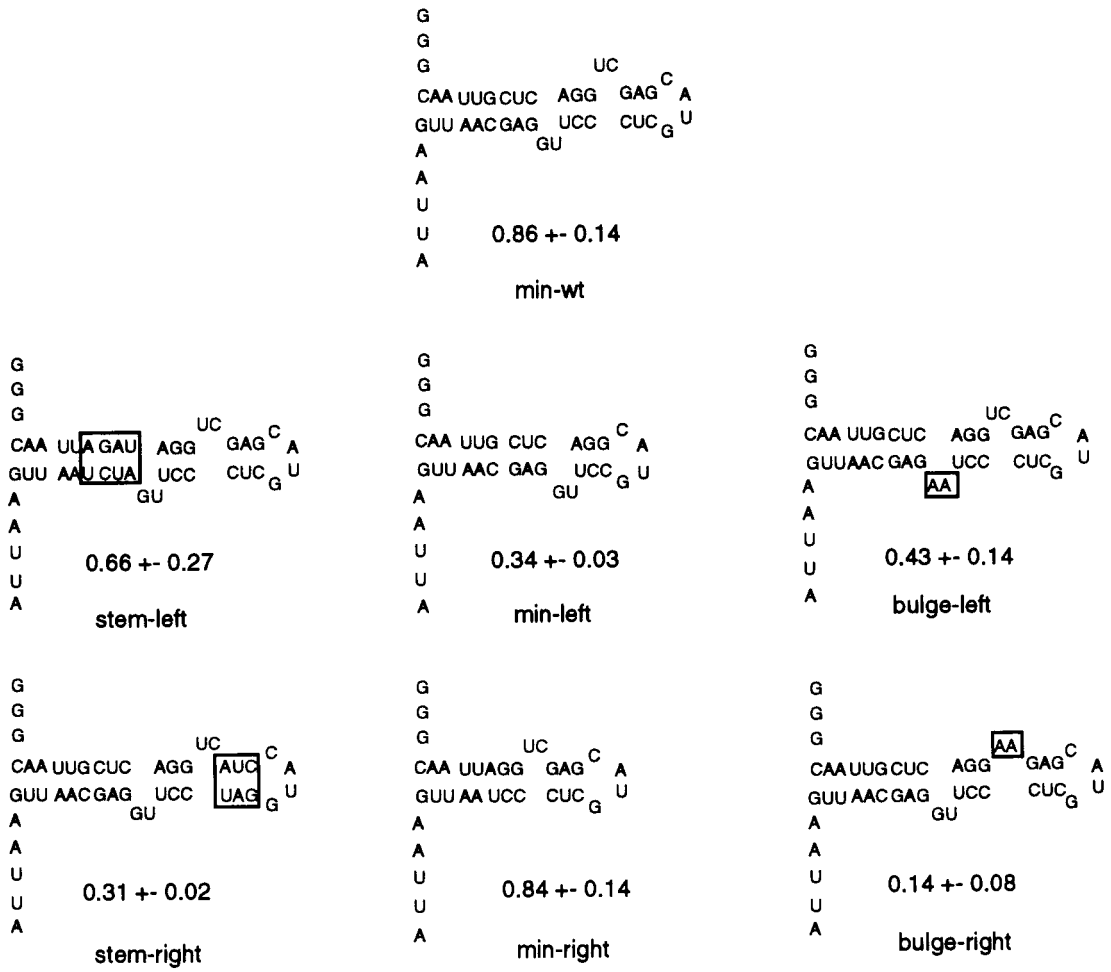

FIG. 5. Site-directed substitutions of the XBE. Minimal elements were synthesized and tested for the ability to compete with stem IID of the XRRE for binding to Rex. Sequences different from the wild-type XBE are boxed. The name of each variant and its binding ratio are shown beneath the predicted secondary structure. Standard deviations are the result of at least three trials for each sequence.

sequence and structural features of the XBE. A potential caveat to this analysis is that sequences with improved binding function might significantly differ from the wild type. However, the most prominent feature of selected sequences was the retention of a subset of wild-type residues (Fig. 3 and 4). No systematic sequence substitution or set of substitutions could be identified that led to improvements in Rex-binding ability. Thus, the sequences derived from this selection can be used to map the wild-type element, just as Rev-binding sequences with both wild-type binding and improved binding were superimposed to map the RBE (3). By determining what positions in the XBE were conserved, the boundaries of the XBE can be established. By determining which residues predominate in selected RNAs, bases that may interact with Rex can be defined. Finally, by determining how the selected sequences may fold, structural models for the XBE can be developed. Each conclusion will be considered in turn.

Previous studies, using a variety of genetic and biochemical techniques, have mapped the boundaries of the XBE to stem IID of the XRRE. The in vitro selection data complement and extend these analyses. In particular, deletion, modification interference, and mutational studies search for sequence changes that lead to loss of Rex-binding function. These techniques can define what residues or structures cannot be lost without concomitantly losing binding activity but are less adept at defining what residues or structures actually contribute to binding activity. Conversely, in vitro selection experiments search primarily for conserved residues that support wild-type activity or substitutions that lead to an increase in Rex-binding activity.

Sequence positions that were highly conserved ($P \leq 0.05$) during the course of the selection establish the extent of the XBE. Most (9 of 10) of these conserved positions were localized within the 22-residue region covering positions 169 to 178 and 202 to 210 (Fig. 4). This region is similar to the minimal XBE defined by deletion analysis (1, 10, 22). For example, while an XRRE that lacks residues 180 to 200 can still function in localizing mRNAs to the cytoplasm, further deletion of residues between positions 170 and 210 (inclusive) abolishes this activity (1). Similarly, an XBE that is bounded on the left by positions 169 and 211 retains the ability to bind Rex in vitro (10). Interestingly, two copies of this minimal XBE (169 to 211) can act as an XRRE and mediate the appearance of mRNAs in the cytoplasm (10). A comparison of the boundaries of the XBE established by deletion studies and in vitro selection data is shown in Fig. 6A.

The minimal XBE defined by deletion and in vitro selection is slightly larger than a Rex-binding site defined by modification interference analysis (5). Modification of positions 174 to

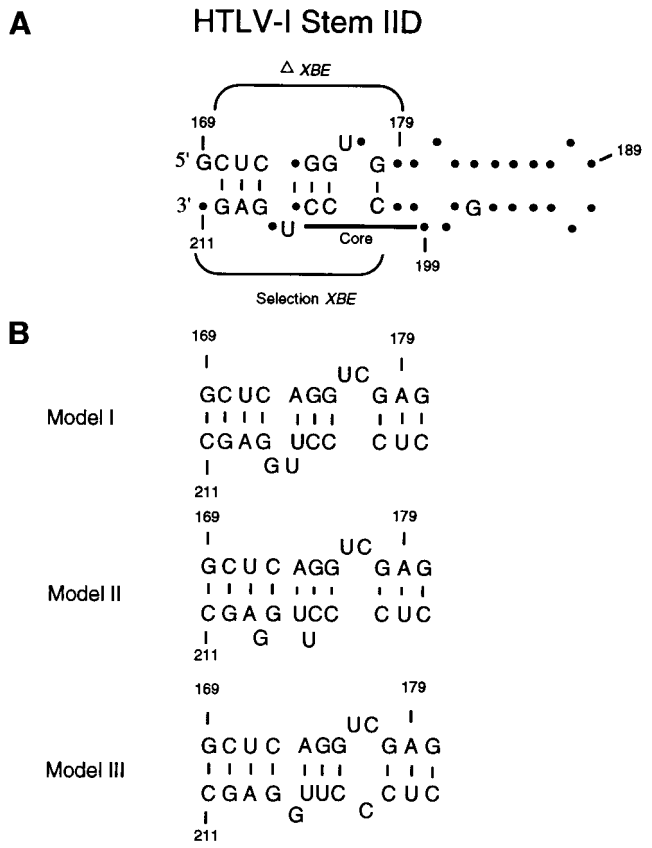

FIG. 6. Models for the secondary structure of the XBE. (A) Data from selection experiments, deletion studies, and modification interference analyses are overlaid. Δ XBE indicates the extent of the functional XBE identified by deletion studies. The Core consists of residues whose modification interferes with Rex binding. Selection XBE indicates the extent of the functional XBE identified by selection experiments. The residues and base pairings that are represented are those that were found to be significantly conserved during in vitro selection. Dots represent residues that were not found to be significantly conserved during in vitro selection. (B) Three different models for the secondary structure of the XBE, model I (5), model II (11), and model III (22), are shown.

180 and 200 to 204 most strongly influenced the ability of the XBE to bind Rex. The researchers termed this 12-residue region the Rex-binding core. In addition to being smaller than the 22-residue XBE defined by deletion and selection, the right boundary of the Rex-binding core has two additional base pairs: that is, the last base pair of the XBE is G-178 · C-202, whereas the last base pair of the Rex-binding core is G-180 · C-200. Figure 6A shows the modification interference data overlaid on the results of deletion and selection experiments.

Selection data can be used to reveal the relative functional significance of individual residues, as well as to define sequence boundaries. The substitutions that occur at a given position in selected sequences fit three types of patterns. First, positions that are conserved between different selected sequences are likely important for function. Second, positions that vary freely between different selected sequences are likely neutral for function. Third, positions that vary in a skewed manner may signal the existence of important structures (e.g., if residues at different positions covary so as to fit a pattern of Watson-Crick interactions, a Watson-Crick pairing may exist between these positions) or chemical features (e.g., the conservation of specific donors or acceptors in the major groove may indicate the importance of a specific contact along that surface). Each of these classes of sequence variations was observed in the current study.

Residues that were highly conserved during selection likely contribute to Rex binding, either by directly contacting the protein or by establishing a structural context for those residues that directly contact the protein. The in vitro selection data show that four of the residues within the XBE core defined by modification interference were also significantly conserved ($P \leq 0.05$) in the selected RNA population (U-176, G-178, C-202, and C-203). Since the modification interference data reveal the proximity of the XBE to Rex and the selection data show what bases in the XBE contribute to Rex-binding function, these four residues are the most likely candidates for the formation of direct contacts with the Rex protein. Additional evidence for the functional significance of these positions can be garnered by observing the skewing of the non-wild-type residues that were found. For example, while position 178 was most frequently a G (57 times), it was sometimes a U (3 times) and never an A or C. Since both guanosine and uridine have a keto moiety in the major groove while adenosine and cytidine present amino groups, this pattern of substitutions may represent the functional conservation of a contact point (O-6 of guanosine 178 in the wild-type element) with Rex.

A number of residues outside of the modification interference core were also found to be conserved in the final pool (G-169, U-171, C-172, U-206, and G-208). Although these residues do not fall within the modification interference core, they may nonetheless form sequence- or structure-specific contacts with Rex. Deletions that retain the core sequence and structure but extend past the defined left-hand boundary (residues 169 and 211) of the XBE have little activity in vivo (10). Similarly, chemical modifications that influence Rex binding do not occur exclusively within the core; for example, modification of C-172 with hydrazine resulted in enhanced interactions with Rex (5). These results are also consistent with the hypothesis that the flanking residues do not directly contact Rex but might aid in establishing or stabilizing the modification interference core structure. In this respect, site-directed sequence substitutions involving the left half of the XBE (orientation as in Fig. 5) tend to be less deleterious than those within the modification interference core, the right half of the XBE.

Results from loss-of-function deletion and modification interference analyses and gain-of-function selection experiments can be reconciled by assuming that residues outside of the modification interference core form a sequence-specific scaffold that supports structure-specific interactions, such as a matrix of charge-charge interactions between the sugar-phosphate backbone of the XBE and the arginines of the Rex ARM. These contacts would likely have been missed by the base modification interference experiments but would have been included among the functional sequence motifs identified by in vitro selection. In a similar study, an analysis of selected Rev-binding sequences led to the identification of a non-Watson-Crick interaction in which the backbone structure, rather than the bases themselves, was recognized by Rev (3, 7, 13).

Many of the positions in the mutagenized sequence segment remained randomized following selection and appear to be essentially neutral towards binding function. Most positions outside of the XBE showed no significant conservation following selection and served as an internal control for sequences that did show significant conservation. Some positions within the boundaries of the XBE were also found to vary randomly following selection and thus are presumably less important for Rex-binding function. However, this conclusion may be dependent on what kinds of variants are assayed. For example, G-180

is not conserved in the selection, but carboxyethylation of this residue can lead to disruption of binding. In this instance, selection and modification interference results can be rationalized by assuming that Rex closely approaches, but does not contact, position 180. Changing G to a residue similar in size, such as A, C, or U, does not affect binding, but addition of a bulky chemical adduct to G may sterically hinder interactions between the protein and the RNA. In contrast, while changing C-177 to A, G, or U was found to be relatively unimportant for the interaction of the XBE with Rex, chemical modification of this residue with hydrazine has been found to improve interactions with Rex (5). As discussed below, these results are most consistent with the involvement of C-177 in a particular tertiary structural motif in the XBE.

In a few instances, positions in the selected molecules were neither random nor conserved relative to the wild-type sequence but, instead, contained a significant number of non-wild-type residues. For example, the distribution of residues at positions 188, 194, 200, and 209 was skewed away from randomness and away from the wild-type residue as well. Wild-type cytidines at positions 188, 194, and 200 all moved toward guanosines, while a wild-type adenosine at position 209 moved toward a pyrimidine. In the most extreme case, only 24 cytidines remained at position 188 following selection (35 were expected), while 20 guanosines occurred instead ($P < 0.03$; 13 were expected). It is possible that these skewings are merely statistical aberrations. However, since many positions outside the boundaries of the XBE were shown to conform to the random distribution expected for neutral residues, it can be argued that these skewed distributions are functionally significant.

The XBE can potentially be folded into several different secondary structures, depending on what bases are assigned to bulge loops and what bases are assigned to base pairs. The statistical significance of functional base pairs, as well as functional residues, is embedded in the artificial phylogeny of Rex-binding sequences. The expected probability that a given wild-type base pair will be observed between two positions is the product of the expected probability that the wild-type bases are present in the same molecule; the probability that the observed distribution of base pairings is due to chance (as opposed to selection) can thus be calculated. Base pairings that were significantly conserved ($P \leq 0.05$ for occurrence by chance) are shown in Fig. 4B. To further solidify this model, sequence covariations between positions predicted to pair were also examined. As a simple example, if substitutions generally varied with one another such that potential Watson-Crick pairings were supported (e.g., A and U at two positions made concerted changes to, respectively, C and G), then this was interpreted as evidence for the existence of a structurally important base pair in the wild-type RNA. Numerous covariations that supported the model shown in Fig. 4 were identified, including three G · U substitutions for C-172 · G-208, two C · G substitutions for A-173 · U-205, and two C · G substitutions for G-174 · C-204 (Table 2).

Three different secondary-structural models of the XBE have previously been advanced (Fig. 6B). As detailed above, our data are most consistent with the model of Bogerd et al. (5). However, for comparison, the relative base-pairing probabilities for each of the models were examined. The differences between the proposed secondary structures can be localized to alternative pairings for four residues: C-172 pairs with either G-207 or G-208, A-173 pairs with either U-205 or U-206, G-174 pairs with either C-204 or U-205, and G-175 pairs with either C-203 or C-204. These possibilities and the models they support are shown in Fig. 6. Model II (11) is less likely than model I (5) on the basis of the larger number of C-170 · G-210 base pairs that could form, i.e., 41, relative to C-172 · G-207 base pairs, i.e., 28. However, model III (22) cannot be as readily discounted. While there are five additional base pairs for G-175 in model I versus model III, there are three fewer base pairs for A-173. When double substitutions are examined for a pattern of Watson-Crick interactions, model I is preferred over model III, but only slightly. Thus, while the sequence data strongly support model I, they cannot completely rule out model III.

Distinguishing between the different secondary-structural models aids in determining what structural features of the viral genome Rex recognizes. These data can, in turn, be used for the design of discrete XBEs. A similar structural definition of the RBE of HIV-1 has led to the design and synthesis of minimal RBEs that have been successfully used as decoys for Rev (14). In addition, the three-dimensional structure of the XBE is largely dependent on what base pairs and bulges are formed in the secondary structure. In this respect, it is interesting that model I closely resembles the TAR element of HIV-1.

The TAR element is a stem-bulge-stem structure that specifically interacts with HIV-1 Tat by forming a binding pocket for arginine. The pocket is, in turn, composed of two critical structural elements: (i) a G · C base pair in which the guanosine forms a network of hydrogen bonds with the guanidino group of arginine and (ii) a base triple formed by an A · U base pair and an unpaired uridine residue. The A · U base pair is adjacent to the conserved G · C pairing, while the unpaired U is located in a bulge loop 5′ to these pairings. The base triple forms in the presence of arginine and positions the negatively charged phosphates within the bulge loop adjacent to the positively charged head group of arginine. In its simplest form, the TAR element consists of the sequence $UN_{1\text{-}2}GA\ldots$ UC embedded in a stem-loop (Fig. 7A).

The essential residues and secondary structure of the XBE identified by in vitro selection are similar to those that would be found in a functional TAR element. In fact, model I for the XBE resembles a dimeric, head-to-head TAR element in which three central base pairs form an axis of symmetry that separates the two arginine-binding pockets (Fig. 7B). Highly conserved potential G · C pairings are observed in both half molecules in the right half at position G-178 · C-202 ($P < 3 \times 10^{-6}$) and in the left half at position C-172 · G-208 ($P < 3 \times 10^{-4}$), potential A · U base pairs are found adjacent (A-179 · U-201 and U-171 · A-209), and bulge loops are found 5′ to these base pairings.

However, the bulge loops in model I of the XBE are only two residues long, as opposed to three residues in the wild-type TAR element. Mutational studies have shown that TAR variants containing a two-residue loop can bind peptides from Tat as well as the wild-type TAR element (26). In addition, structural studies of TAR suggest that it should be possible for a two-base bulge loop to fold into the same TAR-like tertiary structure as a three-base bulge loop (19). If the XBE and TAR bulge loops are structurally and functionally equivalent, then two trends would be expected in the selected sequences. First, uridine residues should be present at the 5′ end of the bulge; in fact, conserved uridines (U-176 and U-206, corresponding to U-23 of Tar; Fig. 7A) are at the beginning of the bulge loops. Conversely, the 3′-most residues in the bulge loops would be predicted to be irrelevant to the formation of TAR-like arginine-binding pockets; in fact, the 3′-most bulged residues (C-177 and G-207) vary randomly.

The proposed U · A · U base triple found in the XBE variants is supported by sequence conservation and covariation.

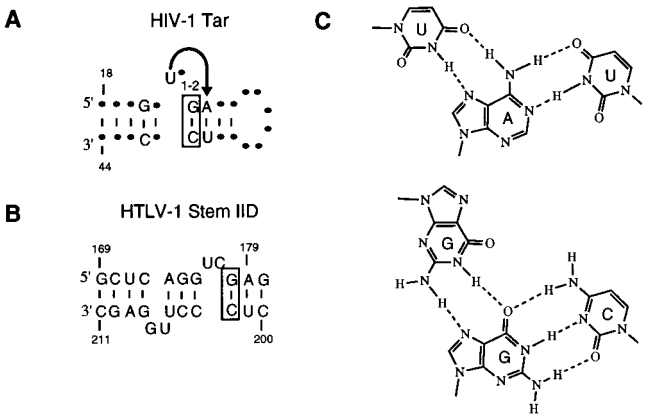

FIG. 7. Similarities between HIV-1 TAR and HTLV-I XBE. (A) Sequence and structural features of HIV-1 TAR. The structure was redrawn from reference 26 by using the numbering of reference 19. Residues and secondary structures that contribute to the ability of TAR to bind arginine, a peptide from Tat, or Tat itself are shown. The arrow highlights a triple-base-pair interaction; the guanosine of the boxed G · C pairing interacts with the guanidino group of arginine (19). (B) Sequence and structural features of HTLV-I XBE. Residues and secondary structures of the XBE are shown. The boxed G · C pairing was the most significantly conserved pairing observed in clones from the final cycle of selection ($P < 3 \times 10^{-6}$ for occurrence by chance). (C) A triple-base substitution observed in selected XBEs. The U · A · U base triple represents the structure observed for the arginine-binding motif of TAR and is hypothesized to occur among residues U-176, A-179, and U-201 of the XBE. The G · G · C base triple is isomorphic to the U · A · U base triple. Residues that would support this hypothesized base triple (G-176, G-179, and C-201) occur in one of the selected clones.

Residue U-176 is present in the bulged region in 29 of the 31 clones containing A-179 · U-201. Similarly, U-206 is found in 28 of 33 clones containing the U-171 · A-209 pairing. Just as concerted changes between two positions provide support for the existence of Watson-Crick base pairings, the identities of triple sequence substitutions may support the proposed base triples. An examination of the selected XBE variants led to the identification of two clones that were triply substituted in the right half of the XBE, at positions 176, 179, and 201. One of these clones (D-33, containing U-176→G, A-179→G, and U-201→C) could potentially form a G · G · C base triple that is predicted to be isomorphic to the U · A · U triple observed in TAR (Fig. 7C). A G · G · C base triple has previously been proposed to occur in group I introns (18).

Further support for the existence of at least one TAR-like tertiary structure comes from a comparison of available modification interference data. Weeks et al. (25) have shown that modification of the 3′-most base (U-25) in the bulge loop of TAR with hydrazine enhances binding to a fragment of Tat. As discussed above, this residue may be functionally equivalent to the 3′-most bases in the XBE bulge loops, C-177 and G-207. Thus, it is interesting that Bogerd et al. (5) show that hydrazine modification of right-half bulge residue C-177 enhances the Rex-binding ability of the XBE.

To further test both the model I secondary-structure prediction and the dual-TAR tertiary-structure prediction, site-specific substitutions were introduced into the XBE. First, to determine if the TAR-like sequences could individually bind Rex, the left half of the XBE (min-left) was transcribed separately from the right half (min-right). Both sequences were found to bind Rex, although the left half was less active than the right (Fig. 5). This result is consistent with modification interference analyses that localize the core binding structure in the right half (Fig. 6A). The apparent asymmetry of the element is also in accord with the activities and phenotypes of a limited set of deletion and substitution mutants studied by Gröne et al. (10). Second, the stem and bulge sequences in each half molecule were separately mutagenized (stem-left, stem-right, bulge-left, and bulge-right). With one exception, the binding ability of the half-molecule variants was reduced by two- to threefold. However, when the right-half bulge that contained U-176 and C-177 was mutated to AA (bulge-right), activity was essentially eliminated. The same mutation also impaired the ability of the full-length XRRE to transport mRNAs in cell-based assays (22). This result again demonstrates the functional asymmetry of the XBE halves.

Taken together, the in vitro selection and site-specific mutational studies have led to a comprehensive sequence and structural definition of the XBE. The boundaries of the minimal XBE are consistent with results from previous studies. Residues that contribute to recognition of the Rex ARM are localized within these boundaries. Sequence conservation and covariation provide evidence for a secondary-structural model of the XBE. The primary and secondary structural features of the XBE identified by selection foster the construction of tertiary structural models. One such model suggests that the XBE contains dimeric, TAR-like arginine-binding pockets. However, sequence conservation and covariations and the activities of site-directed mutations show that there is a functional asymmetry in the XBE half molecules and that right-half core sequences originally identified by modification interference analysis are more important for Rex binding. Our results are most consistent with a model in which the left half of the XBE forms an RNA structure that either stabilizes the right half or makes structure-specific contacts with the Rex ARM, while the right half of the XBE, spanning residues 174 to 179 and 201 to 204, forms a sequence-specific TAR-like arginine-binding pocket.

ACKNOWLEDGMENTS

This work was supported by National Institutes of Health grant R01 GM48175 (A. Ellington), a National Science Foundation National Young Investigator Award (A. Ellington), a Scholar Award from The American Foundation for AIDS Research (A. Ellington), and The Pew Scholar Award in the Biomedical Sciences (A. Ellington).

We thank Amy Boles for excellent technical assistance.

REFERENCES

1. **Ahmed, Y. F., S. M. Hanly, M. H. Malim, B. R. Cullen, and W. C. Greene.** 1990. Structure-function analyses of the HTLV-1 Rex and HIV-1 Rev RNA response elements: insights into the mechanism of Rex and Rev action. Genes Dev. **4:**1014–1022.
2. **Ballaun, C., G. K. Farrington, M. Dobrovnik, J. Rusche, J. Hauber, and E. Böhnlein.** 1991. Functional analysis of human T-cell leukemia virus type I *rex*-response element: direct RNA binding of Rex protein correlates with in vivo activity. J. Virol. **65:**4408–4413.
3. **Bartel, D. P., M. L. Zapp, M. R. Green, and J. W. Szostak.** 1991. HIV-1 Rev regulation involves recognition of non-Watson-Crick base pairs in viral RNA. Cell **67:**529–536.

3a. **Baskerville, S., et al.** Unpublished data.

4. **Bogerd, H. P., G. L. Huckaby, Y. F. Ahmed, S. M. Hanly, and W. C. Greene.** 1991. The type I human T-cell leukemia virus (HTLV-I) Rex trans-activator binds directly to the HTLV-I Rex and the type 1 human immunodeficiency virus Rev RNA response elements. Proc. Natl. Acad. Sci. USA **88:**5704–5708.
5. **Bogerd, H. P., L. S. Tiley, and B. R. Cullen.** 1992. Specific binding of the human T-cell leukemia virus type I Rex protein to a short RNA sequence located within the Rex-response element. J. Virol. **66:**7572–7575.
6. **Ellington, A. D.** 1989. Purification of oligonucleotides by denaturing polyacrylamide gel electrophoresis. *In* F. M. Ausubel et al. (ed.), Current protocols in molecular biology. John Wiley & Sons, Inc., New York.
7. **Giver, L., D. Bartel, M. Zapp, A. Pawul, M. Green, and A. D. Ellington.** 1993. Selective optimization of the Rev-binding element of HIV-1. Nucleic Acids Res. **21:**5509–5518.
8. **Grassmann, R., S. Berchtold, C. Aepinus, C. Ballaun, E. Boehnlein, and B. Fleckenstein.** 1991. In vitro binding of human T-cell leukemia virus *rex* proteins to the *rex*-response element of viral transcripts. J. Virol. **65:**3721–3727.

9. **Green, R., A. D. Ellington, D. P. Bartel, and J. W. Szostak.** 1991. In vitro genetic analysis: selection and amplification of rare functional nucleic acids. Methods **2:**75–86.
10. **Gröne, M., E. Hoffmann, S. Berchtold, B. R. Cullen, and R. Grassmann.** 1994. A single stem-loop structure within the HTLV-1 Rex responsive element is sufficient to mediate Rex activity *in vivo*. Virology **204:**144–152.
11. **Hanly, S. M., L. T. Rimsky, M. H. Malim, J. H. Kim, J. Hauber, M. Duc Dodon, S.-Y. Le, J. V Maizel, B. R. Cullen, and W. C. Greene.** 1989. Comparative analysis of the HTLV-I Rex and HIV-1 Rev *trans*-regulatory proteins and their RNA response elements. Genes Dev. **3:**1534–1544.
12. **Inoue, J.-I., M. Yoshida, and M. Seiki.** 1987. Transcriptional ($p40^{X}$) and post-transcriptional ($p27^{X\text{-}III}$) regulators are required for the expression and replication of human T-cell leukemia virus type I genes. Proc. Natl. Acad. Sci. USA **84:**3653–3657.
13. **Leclerc, F., R. Cedergren, and A. D. Ellington.** 1994. A three-dimensional model of the Rev-binding element of HIV-1 derived from analyses of aptamers. Nat. Struct. Biol. **1:**293–300.
14. **Lee, S.-W., H. F. Gallardo, E. Gilboa, and C. Smith.** 1994. Inhibition of human immunodeficiency virus type 1 in human T cells by a potent Rev response element decoy consisting of the 13-nucleotide minimal Rev-binding domain. J. Virol. **68:**8254–8264.
15. **Lee, T. C., B. A. Sullenger, H. F. Gallardo, G. E. Ungers, and E. Gilboa.** 1992. Overexpression of RRE-derived sequences inhibits HIV-1 replication in CEM cells. Nat. New Biol. **4:**66–74.
16. **Malim, M. H., J. Hauber, S. Y. Le, J. V. Maizel, and B. R. Cullen.** 1989. The HIV-1 rev transactivator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. Nature (London) **338:**254–257.
17. **Mattaj, I.** 1993. RNA recognition: a family matter? Cell **73:**837–840.
18. **Michel, F., A. D. Ellington, S. Couture, and J. W. Szostak.** 1990. Phylogenetic and genetic evidence for base-triples in the catalytic domain of group I introns. Nature (London) **347:**578–580.
19. **Puglisi, J. D., T. Ruoying, B. J. Calnan, A. D. Frankel, and J. R. Williamson.** 1992. Conformation of the TAR RNA-arginine complex by NMR spectroscopy. Science **257:**76–80.
20. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
21. **Tan, R., L. Chen, J. A. Buettner, D. Hudson, and A. D. Frankel.** 1993. RNA recognition by an isolated alpha helix. Cell **73:**1031–1040.
22. **Toyoshima, H., M. Itoh, J.-I. Inoue, M. Seiki, F. Takaku, and M. Yoshida.** 1990. Secondary structure of the human T-cell leukemia virus type 1 *rex*-responsive element is essential for *rex* regulation of RNA processing and transport of unspliced RNAs. J. Virol. **64:**2825–2832.
23. **Tuerk, C., and L. Gold.** 1990. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science **249:**505–510.

23a. **Tuerk, C., and S. MacDougal-Waugh.** 1993. In vitro evolution of functional nucleic acids: high affinity RNA ligands of HIV-1 proteins. Gene **137:**33–39.

24. **Unge, T., L. Solomin, M. Mellini, D. Derse, B. K. Felber, and G. N. Pavlakis.** 1991. The Rex regulatory protein of human T-cell lymphotropic virus type I binds specifically to its target site within the viral RNA. Proc. Natl. Acad. Sci. USA **88:**7145–7149.
25. **Weeks, K. M., C. Ampe, S. C. Schultz, T. A. Steitz, and D. M. Crothers.** 1990. Fragments of the HIV-1 Tat protein specifically bind TAR RNA. Science **249:**1281–1285.
26. **Weeks, K. M., and D. M. Crothers.** 1991. RNA recognition by Tat-derived peptides: interaction in the major groove? Cell **66:**577–588.
27. **Yoshida, M., I. Miyoshi, and Y. Hinuma.** 1982. Isolation and characterization of retrovirus from cell lines of human adult T-cell leukemia and its implication in the disease. Proc. Natl. Acad. Sci. USA **79:**2031–2035.
28. **Zon, G., K. A. Gallo, C. J. Samson, K.-I. Shao, M. F. Summers, and R. A. Byrd.** 1985. Analytical studies of 'mixed sequence' oligodeoxyribonucleotides synthesized by competitive coupling of either methyl- or β-cyanoethyl-N,N-diisopropylamino phosphoramidite reagents, including 2′-deoxyinosine. Nucleic Acids Res. **13:**8181–8196.