High-Affinity ssDNA Inhibitors of the Reverse Transcriptase of Type 1 Human Immunodeficiency Virus†

Daniel J. Schneider,‡,§ Juli Feigon,*,∥ Zdenek Hostomsky,⊥ and Larry Gold*,‡,§

Department of Molecular, Cellular, and Developmental Biology, University of Colorado, Boulder, Colorado 80309, NeXstar Pharmaceuticals, Inc., 2860 Wilderness Place, Boulder, Colorado 80301, Department of Chemistry and Biochemistry and Molecular Biology Institute, University of California, Los Angeles, California 90095, and Agouron Pharmaceuticals, Inc., 3565 General Atomics Court, San Diego, California 92121

Received March 23, 1995; Revised Manuscript Received May 18, 1995⊗

ABSTRACT: The reverse transcriptase (RT) of HIV-1 is a plausible target for therapeutic agents aimed at inhibiting propagation of the virus. We have used "irrational drug design", that is, combinatorial chemistry with oligonucleotide libraries, to identify high-affinity ligands aimed at HIV-1 RT. The methodology, termed SELEX (*s*ystematic *e*volution of *l*igands by *ex*ponential enrichment), was employed with a single-stranded DNA library. The selected ssDNA ligands bind HIV-1 RT with K_d values as low as 1 nM and inhibit the RNA-dependent DNA-polymerase activity of the enzyme with K_i values as low as 0.3 nM. We also demonstrate the high specificity of one ligand able to selectively discriminate between the reverse transcriptases of HIV-1, AMV, and MMLV. These ssDNA molecules may be useful as inhibitors or as models for the design of small molecule inhibitors of HIV-1 RT *in vivo*.

The reverse transcriptase (RT) of type 1 human immunodeficiency virus (HIV-1) plays an indispensable role in the life cycle of the virus. Its primary function is the synthesis of a double-stranded DNA copy of the RNA genome for integration into the host chromosome. This is achieved by concerted activities including minus-strand DNA synthesis via an RNA-dependent DNA polymerase activity, concomitant degradation of the template RNA strand via an RNase H activity, and plus-strand DNA synthesis via a DNA-dependent DNA polymerase activity (Baltimore, 1970; Gilboa *et al.*, 1979; Goff, 1990; Peliska & Benkovic, 1992; Temin & Mizutani, 1970).

HIV-1 is generally accepted as the etiological agent of acquired immune deficiency syndrome (AIDS). The importance of RT in the life cycle of HIV-1 and the lack of a natural function in the host cell make RT a preferred target for antiviral agents. The X-ray crystal structures of HIV-1 RT complexed with a small non-nucleoside inhibitor (Kohlstaedt *et al.*, 1992) and a DNA duplex (Jacobo-Molina *et al.*, 1993) have been solved to 3.5 and 3.0 Å, respectively. Structurally, HIV-1 RT exists as a heterodimer consisting of a 66-kDa polypeptide chain (p66) complexed with a 51-kDa polypeptide chain (p51), identical in primary structure to p66 but lacking the C-terminal domain responsible for the RNase H activity intrinsic to p66 (Hansen *et al.*, 1988). As a heterodimer, p66 contains a groove where a duplex nucleic acid substrate can be accommodated between the polymerase and RNase H active sites (Arnold *et al.*, 1992; Jacobo-Molina *et al.*, 1993; Kohlstaedt *et al.*, 1992; Krug & Berger, 1991). First-strand DNA synthesis is initiated *in vivo* from a primer:template junction formed between the 3′ end of host cell $tRNA^{Lys,3}$ and a complementary region of the RNA genome (primer binding site) (Panet, *et al.*, 1975). This 18-nucleotide A-form helix is the correct size for an interaction with the groove of p66, and if they interact according to the model, the remainder of the tRNA would be positioned in close association with p51 (Kohlstaedt *et al.*, 1992). Specific cross-linking of the tRNA anticodon stem and loop to p51 (as well as p66) support the proposed proximity of the tRNA molecule and p51 (Barat *et al.*, 1989).

We have used SELEX (*s*ystematic *e*volution of *l*igands by *ex*ponential enrichment) (Tuerk & Gold, 1990) to identify high-affinity nucleic acid ligands that may be useful both as inhibitors of HIV-1 RT and for structural analysis of the enzyme. Previous selections identified RNA pseudoknots that bind specifically to HIV-1 RT and inhibit the RNA-dependent DNA polymerase activity (Tuerk *et al.*, 1992). Since the enzyme also recognizes duplex DNA as a substrate, a high-affinity DNA ligand may also be useful for comparing the recognition of DNA and RNA substrates by this enzyme. With HIV-1 RT as a target, we performed a SELEX experiment using a degenerate ssDNA library containing 35 random positions (35N) and identified a subset with an apparent affinity for HIV-1 RT almost 1000 times greater than that of the library from which it originated. Members of this subset, sharing little similarity with each other or with the RNA pseudoknot (Tuerk *et al.*, 1992) at the levels of primary and secondary structure, inhibit the RNA-dependent DNA polymerase activity of HIV-1 RT at very low concentrations. This inhibition is specific for HIV-1 RT; the polymerase activities of reverse transcriptases from avian myeloblastoma virus (AMV-RT) and Moloney murine leukemia virus (MMLV-RT) are unaffected by the presence of one of the ssDNA inhibitors. We believe the high affinity

† This work was supported by NIH Grants GM 19963 and GM 28685 to L.G., NIH Grant P01 GM 39558 and a Dreyfus Teacher-Scholar Award to J.F, and NIH Grant AI 3380 to Z.H. D.J.S. was supported by NIH Predoctoral Training Grant GM 07135.

* To whom correspondence may be sent.
‡ University of Colorado.
§ NeXstar Pharmaceuticals, Inc.
∥ University of California, Los Angeles.
⊥ Agouron Pharmaceuticals, Inc.

⊗ Abstract published in *Advance ACS Abstracts,* July 1, 1995.

0006-2960/95/0434-9599$09.00/0

and specificity for HIV-1 RT exhibited by the selected DNA ligands, along with their inhibitory effects, make them good candidates for structural investigations and potentially for therapeutic application.

MATERIALS AND METHODS

Source of HIV-1 RT. Recombinant HIV-1 RT overexpressed in *Escherichia coli* cells was purified according to the procedure described in Davies *et al.* (1991) and Hostomsky *et al.* (1991). Enzyme was aliquoted and stored at −70 °C in HRT buffer (200 mM KOAc, 50 mM Tris-HCl, pH 8.0, 6 mM $MgCl_2$, and 10 mM DTT). Aliquots were thawed and refrozen only once.

Generation of a Degenerate ssDNA Library. A library of synthetic DNA oligonucleotides (oligonucleotide 1) containing 35 random nucleotides flanked by invariant primer annealing sites was amplified by the polymerase chain reaction (PCR) using oligonucleotides 2 and 3 as primers (Figure 1). Oligonucleotide 3 had three biotin phosphoramidites covalently attached to its 5′ terminus. The 81-nucleotide double-stranded PCR product was size-purified on a 12% non-denaturing acrylamide gel, and 100−300 pmol was applied to 100 μL of a Pierce streptavidin−agarose bead matrix suspended in buffer A (50 mM NaCl, 10 mM Tris-HCl, pH 7.5, and 1 mM EDTA). After equilibration for 30 min at 20 °C to allow the biotinylated double-stranded DNA (dsDNA) to bind the streptavidin beads, unbound dsDNA was removed with five 500-μL washes of buffer A, and the matrix-bound dsDNA was denatured in 400 μL of 0.15 N NaOH for 15 min at 37 °C. As these conditions were not harsh enough to disrupt the biotin−streptavidin interaction, denaturation released only the non-biotinylated DNA strand from the bead complex. The free DNA was collected and precipitated, yielding 70−200 pmol of single-stranded DNA (ssDNA). Ten to 20 pmol was ^{32}P labeled at the 5′ end with T4 polynucleotide kinase, and the product was size-purified on an 8% denaturing acrylamide gel and combined with the remaining (unlabeled) ssDNA to comprise the degenerate ssDNA library used for the selection.

Selection for HIV-1 RT Affinity. The affinity of the random ssDNA library for HIV-1 RT was determined using a protein-excess nitrocellulose filter binding assay as described in Carey *et al.* (1983). Although higher in ionic strength than a physiological buffer, HRT buffer was chosen for the binding assays and selections because of its proven success in RNA SELEX experiments targeting HIV-1 RT (Tuerk *et al.*, 1992). Selections were performed with a saturating ssDNA concentration to promote competition among DNA ligands for a limited number of available target binding sites. The percent of target-dependent DNA retention was minimized each selection to ensure maximum enrichment of the library for target binders; however, to avoid propagation of members with high affinity for nitrocellulose, selections were repeated if target-free (background) retention was greater than 10% of target-dependent retention. For the first selection, 500 nM HIV-1 RT and 2 μM ssDNA (100 pmol, or about 10^{14} different molecules) were equilibrated at 37 °C for 5 min in HRT buffer and filtered through nitrocellulose to sequester target-bound ligands. Target-free selections were done in duplicate to measure and correct for background binding levels. The fraction of total DNA retained by the filters was calculated by measuring Cerenkov radiation in a scintillation counter. Ligands were harvested from the filter as described in Tuerk and Gold (1990), amplified by PCR, denatured from the biotinylated complementary strand, and end-labeled as described previously to regenerate the library. The affinity of the pool for HIV-1 RT was measured prior to selections 6, 8, 10, and 12 and was estimated for all other selections. The affinity value (dissociation constant, K_d) was used to determine the ligand concentration necessary for saturation during each selection. A ligand concentration at least 10-fold above the K_d value was used to saturate binding to the target. As the affinity of the population for HIV-1 RT increased, the concentrations of ligand and RT were reduced accordingly to increase selection stringency.

Cloning and Sequencing Isolates. Following round 15, 1 pmol of the library was amplified by PCR using oligonucleotide 3 without the biotins (containing a *Pst*I restriction endonuclease cleavage site) and oligonucleotide 4 (containing a *Bam*HI site) as primers (see Figure 1). Double-stranded products were digested with *Pst*I and *Bam*HI and subsequently ligated into pUC19, similarly digested prior to the ligation. The vectors were electroporated into *E. coli* DH1α cells, and isolates were sequenced by dideoxy extension of end-labeled oligonucleotide 5 with AMV reverse transcriptase. For a detailed description of these techniques, refer to Schneider *et al.* (1993). Large quantities of individual DNA ligands were prepared by amplifying the vector inserts by PCR using oligonucleotides 2 and 3 as primers and following the streptavidin matrix purification technique described previously to isolate ssDNA.

Assay for Inhibition of RNA-Dependent DNA Polymerase Activity. A substrate for the RNA-dependent DNA polymerase activity of HIV-1 RT was assembled by annealing an 18-nucleotide, 5′ end-labeled DNA primer to a 30-nucleotide RNA template with a complementary 3′ end (see Figure 5A) and purifying the duplex on a 12% non-denaturing acrylamide gel. The primer sequence matched the 3′-terminal 18 nucleotides of $tRNA^{Lys,3}$, responsible for priming minus-strand DNA synthesis of the HIV-1 genome *in vivo*, while the template sequence paralleled the HIV-1 genomic primer binding site and downstream 12 nucleotides. A dilution series of inhibitory ssDNA ligand (to give a final concentration of 0, 1, 3, 9, 27, or 81 nM) was denatured in HRT buffer at 70 °C for 5 min and allowed to renature slowly at 20 °C. The primer:template substrate was added to a final concentration of 40 nM, along with dNTPs at 400 μM. The 10-μL reaction was initiated with the addition of either HIV-1, AMV, or MMLV RT (to give a final concentration of 10 nM), allowed to proceed for 5 min at 37 °C, and terminated with 1 vol of formamide. Extension products were separated on an 8% denaturing acrylamide gel and quantitated with an Ambis radioanalytic imager.

Antagonists as Substrates for Polymerase Activity. In a 10-μL reaction mixture, 0.1 pmol of 5′ end-labeled ssDNA ligand was denatured and slowly renatured as above, combined with 400 μM dNTPs and a saturating concentration of enzyme (200 nM HIV-1 RT, 100 nM AMV RT, or 0.6 unit/μL Sequenase T7 DNA polymerase, all determined to have equivalent activity), extended for 30 min at 37 °C, and terminated with 1 vol of formamide. To determine the precise location of the annealed 3′ end, extensions were also done with Sequenase in the presence of 25 μM ddATP.

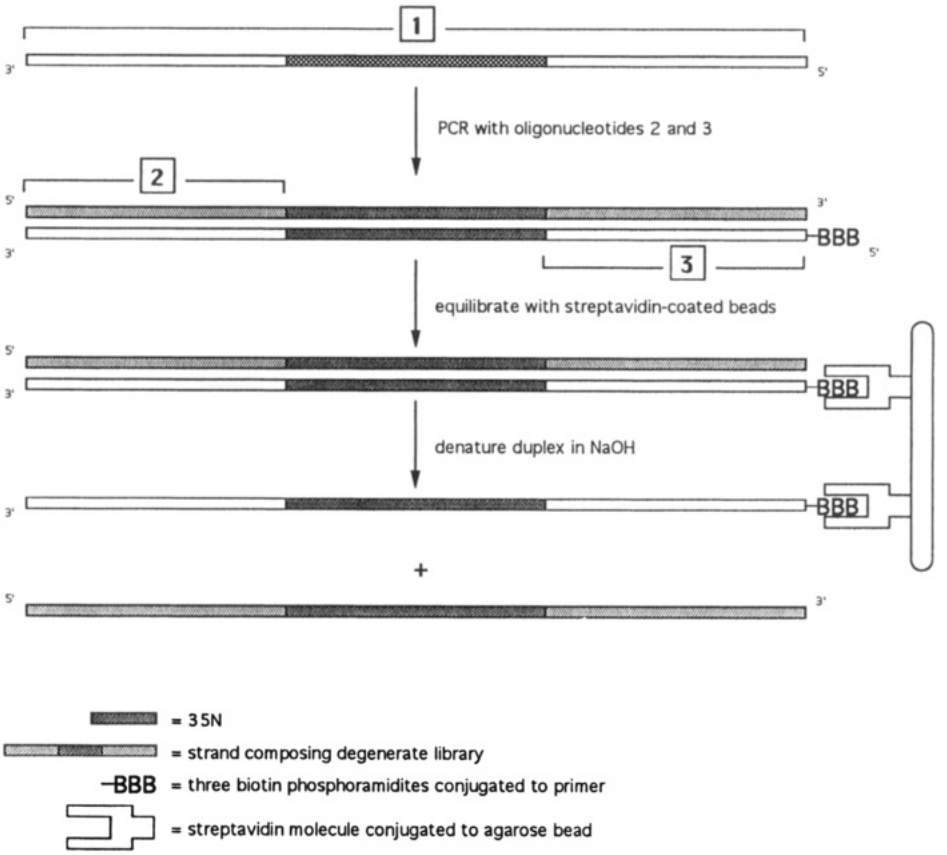

FIGURE 1: Experimental design and oligonucleotide sequences. A degenerate double-stranded DNA library was created using the polymerase chain reaction to amplify oligonucleotide 1, using oligonucleotides 2 and 3 as primers. Oligonucleotide 1 had a central cassette consisting of 35 random positions, while oligonucleotide 3 had three biotin phosphoramidites added to its 5′ end during chemical synthesis. The size-purified double-stranded PCR product was attached to agarose beads coated with streptavidin by virtue of the biotinylated strand, and the non-biotinylated strand was isolated by denaturation of the duplex in 0.15 N NaOH followed by removal of the beads. This population of non-biotinylated strands comprised the degenerate single-stranded DNA library used for the first selection. Oligonucleotide 3 contains a *Pst*I restriction endonuclease site (CTGCAG) and was used (without the biotins) for cloning individuals of the fully enriched library into pUC19 along with oligonucleotide 4, containing a *Bam*HI site (GGATCC). Oligonucleotide 5, complementary to 16 nucleotides of the pUC19 polylinker region, was used as a primer for dideoxy sequencing of the cloned inserts. Refer to Materials and Methods for detailed descriptions of each of these procedures.

Extension products were separated on an 8% denaturing acrylamide gel.

Biased Randomization Selections. A library of ligand RT1 variants was chemically synthesized, incorporating the "wild-type" nucleotide at a frequency of 0.625 and each of the "mutant" nucleotides at a frequency of 0.125 at each position in the 35N cassette. Selections for HIV-1 RT affinity were performed as described above; however, during the amplification step, [α-^{32}P]dATP was incorporated into both strands of the duplex. The strands were separated on an 8% denaturing acrylamide gel by virtue of the retarded migration of the strand possessing the three biotins, and the internally

labeled, non-biotinylated strand was recovered and used for the next selection.

CLUSTER Algorithm. CLUSTER is a program that performs multiple sequence alignment with reoptimization of gap placement within the growing consensus (written by B. Javornik and D. A. Zichi, manuscript in preparation). The algorithm consists of two parts: sequence alignment and clustering. Sequence alignment uses the dynamic programming algorithm of Altschul and Erickson (Altschul & Erickson, 1986) with a weight vector selected on an *a priori* statistical basis, namely, a match = 1.0, a mismatch = −1/3, a gap opening = −1.0, and a gap extension = −1/3. The total cost of alignment is the sum of each pairwise alignment within the consensus, utilizing the quasi-natural gap costs of Altschul (Altschul, 1989). Normalization of alignment costs allows for comparison between alignments that contain different numbers of sequences. The normalization used in CLUSTER compares an alignment to the best possible one in which every position matches. A normalized score is the cost of alignment divided by the cost of the best possible alignment. The K-Means algorithm clusters sequences into families. Here, the algorithm is modified slightly from the original version (Tou & Gonzales, 1974) to accommodate cost of alignment as the distance measure. The steps to the algorithm are the following:

1. Let the N sequences be the initial cluster centers.
2. Group the two clusters that are closest, i.e. the two clusters that are most similar in sequence on the basis of their alignment.
3. Compute the new cluster center by optimizing the alignment.
4. If not converged, then go to step 2.

Convergence occurs when there is only one family or when the cost to combine any two clusters is beyond a threshold. Optimization (step 3) pulls out subsets of sequences and realigns them as described by Subbiah and Harrison (1989).

RESULTS

Selected Individuals Classified into Six Families. Following the selection guidelines described in Materials and Methods, we were able to enrich the DNA library for HIV-1 RT binders from an initial apparent K_d value of 1400 nM to a final value of 2 nM in 15 SELEX cycles, a total increase in affinity of 700-fold (data not shown). Forty-six individuals were cloned from the selected library, of which 30 had unique sequences. These 30 isolates were classified into families according to common primary sequence elements using the alignment-based computer algorithm CLUSTER (see Materials and Methods), as illustrated in Figure 2. Ten of the 30 sequences contain a conserved pentamer, CCCCT, in the central 35-nucleotide cassette. The sequence of the invariant 3′ end of these molecules is AGGGG, which can pair with the internal CCCCT to form a structure mimicking a primer:template junction, a natural substrate for the enzyme *in vivo*. Additional conserved elements flanking the CCCCT separate these 10 individuals into families I and II. The remaining 20 individuals lack the CCCCT, but possess other conserved elements unique to their respective families (III through VI).

HIV-1 RT Binders Characterized by Interrupted Helices. The primary sequence diversity between families suggests that if there is a common element responsible for the high affinity of the selected ligands, it is structural. Unfortunately, a reliable set of rules characterizing the folding of ssDNA molecules has not been described. We used two available tools to make secondary structure predictions for each family. The first was an algorithm which reveals conserved base-pairing patterns and positional covariation among an aligned sequence set (Davis *et al.*, 1995); the second was an algorithm using free energy minimization parameters to predict RNA secondary structure (Jaeger *et al.*, 1989; Zuker, 1989). Although neither of these tools was sufficient alone, structural credibility was increased when base-pairing patterns were supported by both the phylogenetic and thermodynamic predictors. For each family, a potential secondary structure offered by these tools, reflecting only base-pairing patterns common to all of the family members, is illustrated in Figure 3.

There is no secondary structure element common to all of the families; however, several small conserved motifs do appear in subsets of families. As mentioned earlier, families I and II have in common an element mimicking a primer:template junction: a five base pair duplex at the 3′ end with a 5′ flanking single strand that could act as a template for nucleotide addition to the 3′ end. Family IV possesses a variation of this duplex element but with the 3′-terminal G residue unpaired. Families I and III each contain an internal loop motif with an AA opposite a CG, flanked by specific helices. Finally, families II, V, and members of VI all pair their 5′ termini in a similar fashion, with an unpaired A followed by five conserved base pairs. Isolates shown in Figure 2 that do not conform to the conserved structures mentioned above adopt unique secondary structures: the remaining members of family VI form a pseudoknot structure, and ligand RT16 forms an imperfect helix. Aside from the conserved elements mentioned above, possibly representing specific protein binding components, the families show only a general preference for perturbed helices. From the selected library, six of the most highly represented ligands and two additional ligands (indicated with asterisks in Figure 2) were chosen to further investigate the characteristics of the families they represent.

Binding Curves Confirm High Affinity of Individual Ligands. For each of the eight chosen ligands, affinity values for HIV-1 RT were measured using the filter binding assay described in Materials and Methods (see Table 1). The RNA pseudoknot inhibitor reported in Tuerk *et al.* (1992) was also measured and had a K_d value of 5 nM under our conditions. Isolates RT1 and RT26 exhibited the highest affinity, having K_d values of approximately 1 nM, while the others ranged from 2 to 11 nM. No significant correlation was observed between the affinity of a ligand and the family into which it was classified in Figure 2, as the three highest affinity ligands (RT1, RT12, and RT26) were each classified into different families. However, both RT1 and RT26 can adopt the internal loop structure shaded in Figure 3, suggesting a possible participation of this motif in conferring high affinity upon ligands that possess it.

Antagonists as Substrates for Polymerase Activity Terminate Prematurely. The isolates able to form an intramolecular primer:template junction (RT10, RT12, and RT26) were assayed for the ability to be extended from their 3′ termini by the DNA-dependent DNA polymerase activity of a variety of enzymes. Results for ligand RT26 are shown

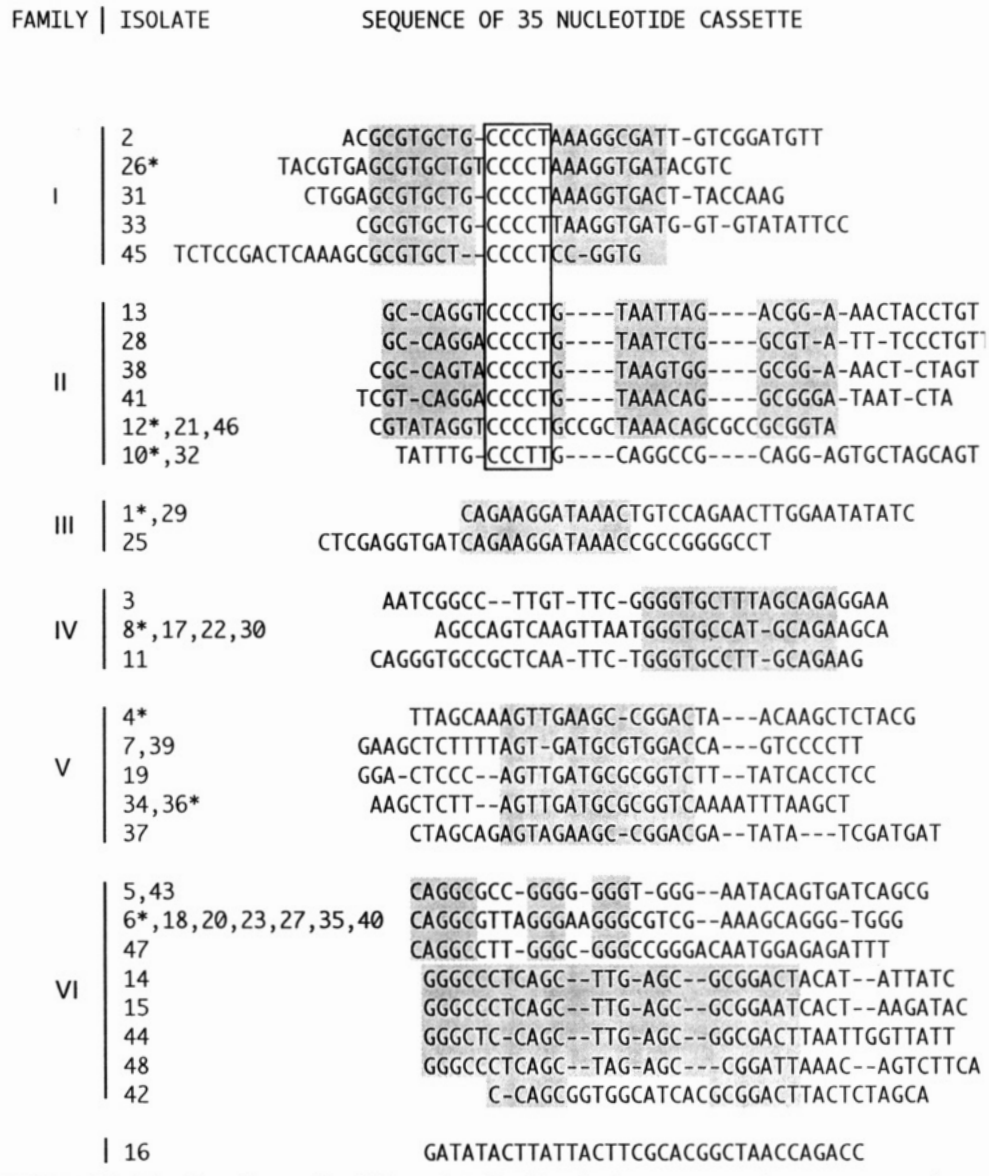

FIGURE 2: Classification of isolates from the enriched library into families sharing common primary sequence elements. The 30 isolates were classified into subsets according to aligned primary structure similarities within their originally randomized 35N region using the CLUSTER algorithm (see Materials and Methods). Only the 35 positions originally randomized are shown for each numbered individual. Regions of similarity among isolates are shaded. Individuals marked with an asterisk were chosen for further characterization. Ligands will be referred to in the text as RT*N*, where *N* represents the ligand number corresponding to the sequence shown in this figure.

in Figure 4A. For the greater part of the substrate, extension with a saturating concentration of HIV-1 RT proceeded only 5−8 nucleotides before premature termination occurred. AMV-RT initiated less efficiently, but extension proceeded to completion (when the end of the template was reached). With Sequenase T7 DNA polymerase, both initiation and extension were efficient and proceeded to completion. The sequence pattern created by extension with Sequenase in the presence of ddATP supported the proposed annealing of the 3′-terminal AGGGG of RT26 to the internal CCCCT. We believe that a common structure for RT26 is recognized by all three enzymes and that the primer:template junction structure confirmed by Sequenase extension is also the structure recognized by HIV-1 RT, although this has not been tested directly. Similar HIV-1-specific extension anomalies were found for ligands RT10 and RT12 (data not shown), and the results are summarized in Figure 4B with arrows indicating premature termination positions.

Both of the RT26 extension products (premature and complete) were isolated and found to have 100-fold lower affinity for HIV-1 RT than the unextended RT26 (data not shown). We believe the products represent abortive extensions from the junctions predicted in Figure 3. Addition of the 5−8 templated nucleotides to the 3′ end of RT26 may create a low-affinity product released by the enzyme more frequently than addition of the next nucleotide, or a trapped product unable to release enzyme or be further extended. It is unlikely that incomplete extension was caused by impeding secondary structures, as product lengths for RT10, RT12, and RT26 extensions did not correlate with positions at the bases of stable hairpins or other extraordinarily stable secondary structures (see Figure 4B). The short products seen when extending RT10, RT12, and RT26 with HIV-1 RT might simply reflect alternative annealing patterns of the 3′ termini recognized by this enzyme as primer:template junction substrates. However, no base-pairing patterns allowing the 3′ terminus to form an extendible primer: template junction were consistent with the lengths of the extension products seen for each of the three ligands.

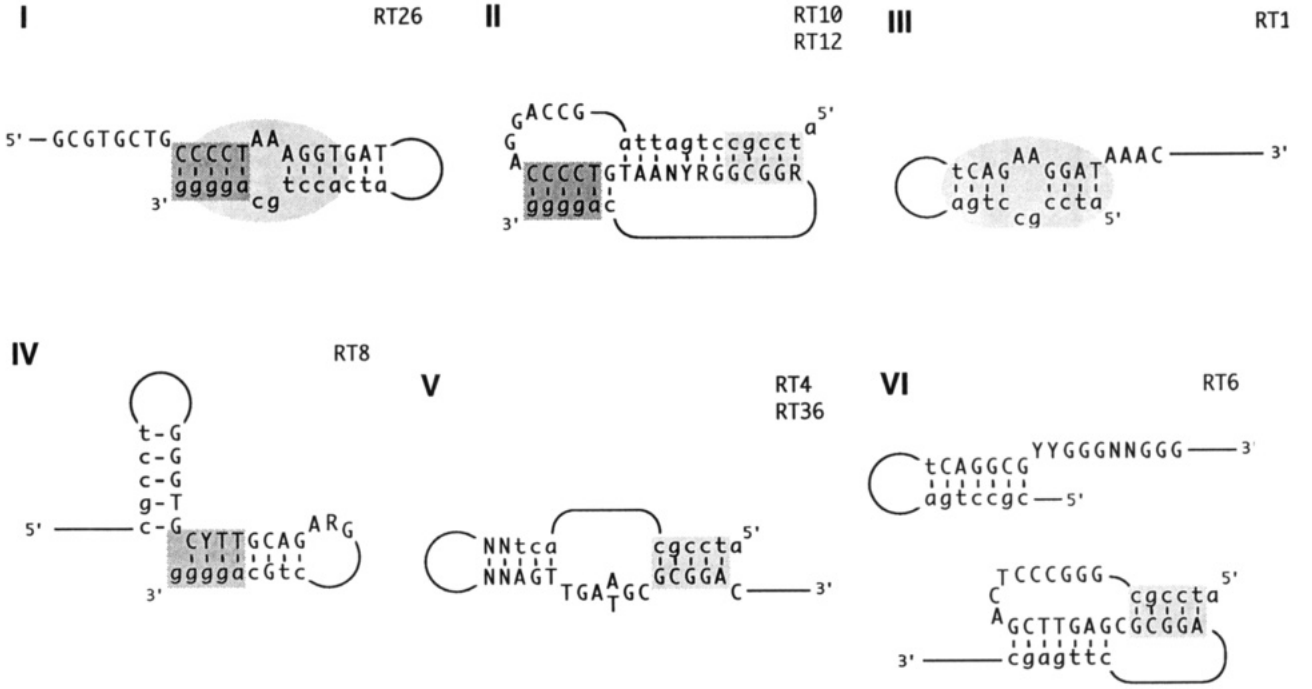

FIGURE 3: Conserved elements of predicted secondary structure for each family. Thermodynamically optimal structures for each of the isolates in Figure 2 were generated using an RNA folding algorithm (Jaeger *et al.*, 1989; Zuker, 1989). Optimal and suboptimal structures were compared within each group, and structural elements conserved in all members of a family were used to predict functional binding motifs for that family. The family number is indicated with a roman numeral in the top left corner of each panel, and the individuals marked with an asterisk in Figure 2 are listed in their respective families in the top right corner of each panel. Boxing and shading patterns reflect elements conserved among different families. In this and all subsequent figures, positions originally randomized are indicated with capital letters, while invariant positions are indicated with lowercase letters. R represents a purine, Y a pyrimidine, and N a purine or a pyrimidine.

Table 1: Comparison of K_d and K_i Values for Indicated Ligands[a]

ligand	K_d (nM)	K_i (nM)
degenerate library	1400	>3000
RNA pseudoknot	5	4.7
RT1	1	<0.3
RT4	8	4
RT6	5	30
RT8	5	13
RT10	11	62
RT12	2	
RT26	1	2.7
RT36	4	6.5

[a] Dissociation constants were determined using the assay described in Carey *et al.* (1983), and inhibition constants were measured according to the assay described in Materials and Methods and Figure 5.

Inhibition of Polymerase Activity Suggests Interaction at Substrate Binding Site and/or Active Site. The ability of seven of the isolates to inhibit the RNA-dependent DNA polymerase activity of HIV-1 RT was assayed by measuring the decrease in extension product formation from a primer: template substrate as a function of inhibitor concentration (Figure 5). From these data, K_i values were determined and were found to exhibit the same general trend as their respective K_d values measured independently with a filter binding assay (see Table 1). Almost no inhibition was seen with as high as 81 nM of the degenerate ssDNA library present (R0, $K_i > 3$ μM). The RNA pseudoknot (RNA pk) inhibited this activity of HIV-1 RT with a K_i value of 4.7 nM under our conditions, consistent with the K_d value shown in Table 1 and Tuerk *et al.* (1992). Of the seven ligands tested, RT1 and RT26 were the most potent inhibitors of the RNA-dependent DNA polymerase activity of HIV-1 RT, with K_i values of 0.3 and 2.7 nM, respectively. Because we have not assayed each component of polymerization for inhibition by these ligands, the K_i values we report reflect an inhibition that could occur at any step of the pathway. However, the similarity between the K_i and K_d values suggests that the mechanism of inhibition may be a competition between the inhibitory ligand and the substrate for the nucleic acid binding site and/or the polymerase active site of RT, although we have not tested this directly.

Selection Using Partially Randomized RT1 Library Identifies Essential Elements. From a library of partially randomized RT1 mutants synthesized as described in Materials and Methods, we selected those maintaining a high affinity for HIV-1 RT. In six SELEX cycles the affinity of the library increased almost 1000-fold, from 1500 nM to approximately 2 nM (data not shown). The results of this experiment are summarized by the consensus illustrated in Figure 6. On the basis of these results, truncated versions of ligand RT1 were synthesized and affinities for HIV-1 RT were measured with a filter binding assay (see Figure 7). RT1t30, composed of the first 30 nucleotides of RT1 containing the proposed internal loop duplex, showed no significant binding below 1 μM HIV-1 RT. However, addition of 19 nucleotides downstream produced a 49-mer (RT1t49) that bound HIV-1 RT with an affinity of 4 nM, nearly as high as that of the full-length 81-nucleotide RT1. The relative affinities of RT1, RT1t30, and RT1t49 suggest that all of the specific binding components of RT1 exist in the first 49 nucleotides and that, while the proposed internal loop motif is insufficient alone, it likely participates in the interaction with HIV-1 RT in combination with other specific binding components.

49-mer Inhibits HIV-1 RT Specifically. The inhibition assay described previously was used to determine the

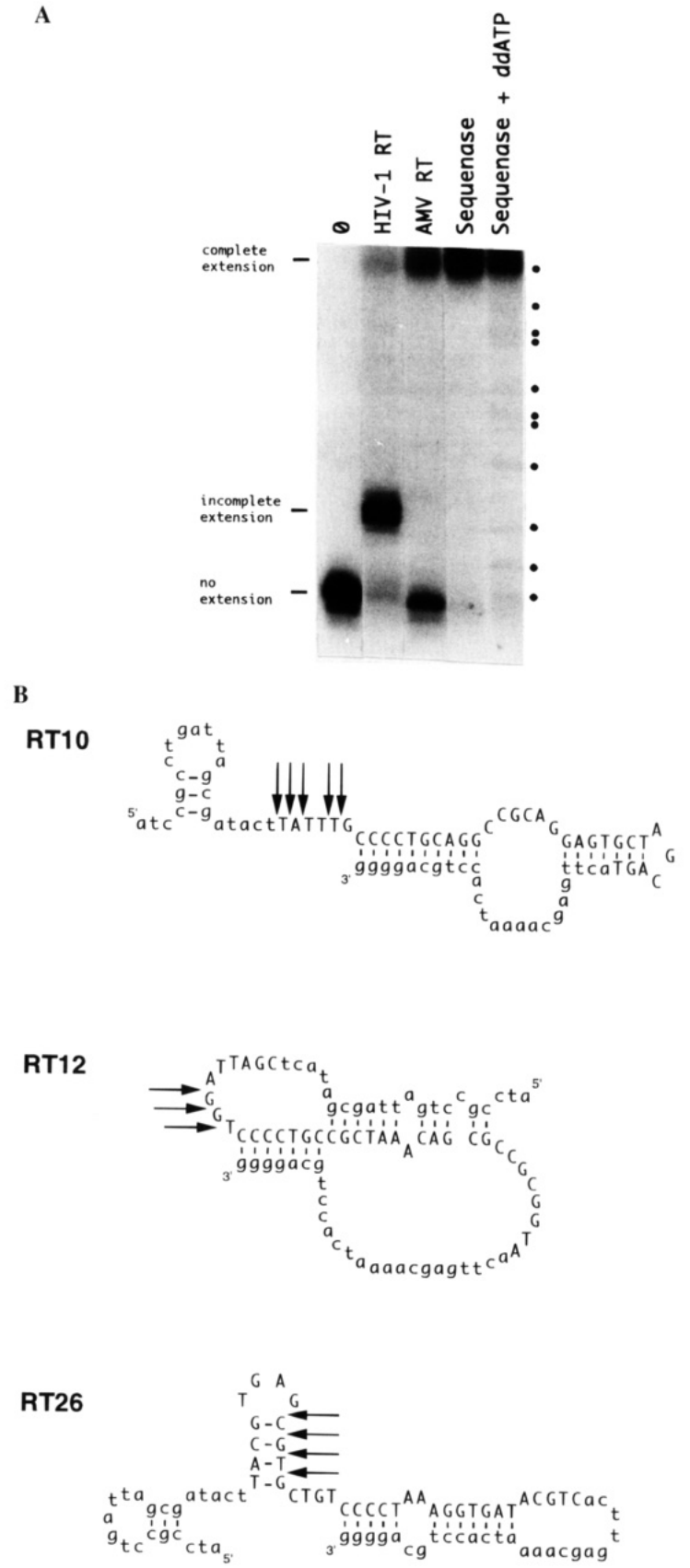

FIGURE 4: (A) Intramolecular extension of RT26. End-labeled RT26 was extended with a saturating concentration of either HIV-1 RT, AMV RT, or Sequenase T7 DNA polymerase (see Materials and Methods). Terminations at positions of ddATP incorporation are indicated with bullets. (B) Predicted secondary structures of RT10, RT12, and RT26, with arrows indicating positions where premature termination occurred in the extension assay described above and in Materials and Methods.

specificity of inhibition of ligand RT1t49 by comparing in parallel its ability to inhibit the polymerase activities of HIV-1 RT, AMV RT, and MMLV RT. As illustrated in Figure 8, inhibition of primer extension was seen when performed with HIV-1 RT, but was not detectable when performed with AMV RT and MMLV RT, even at inhibitor

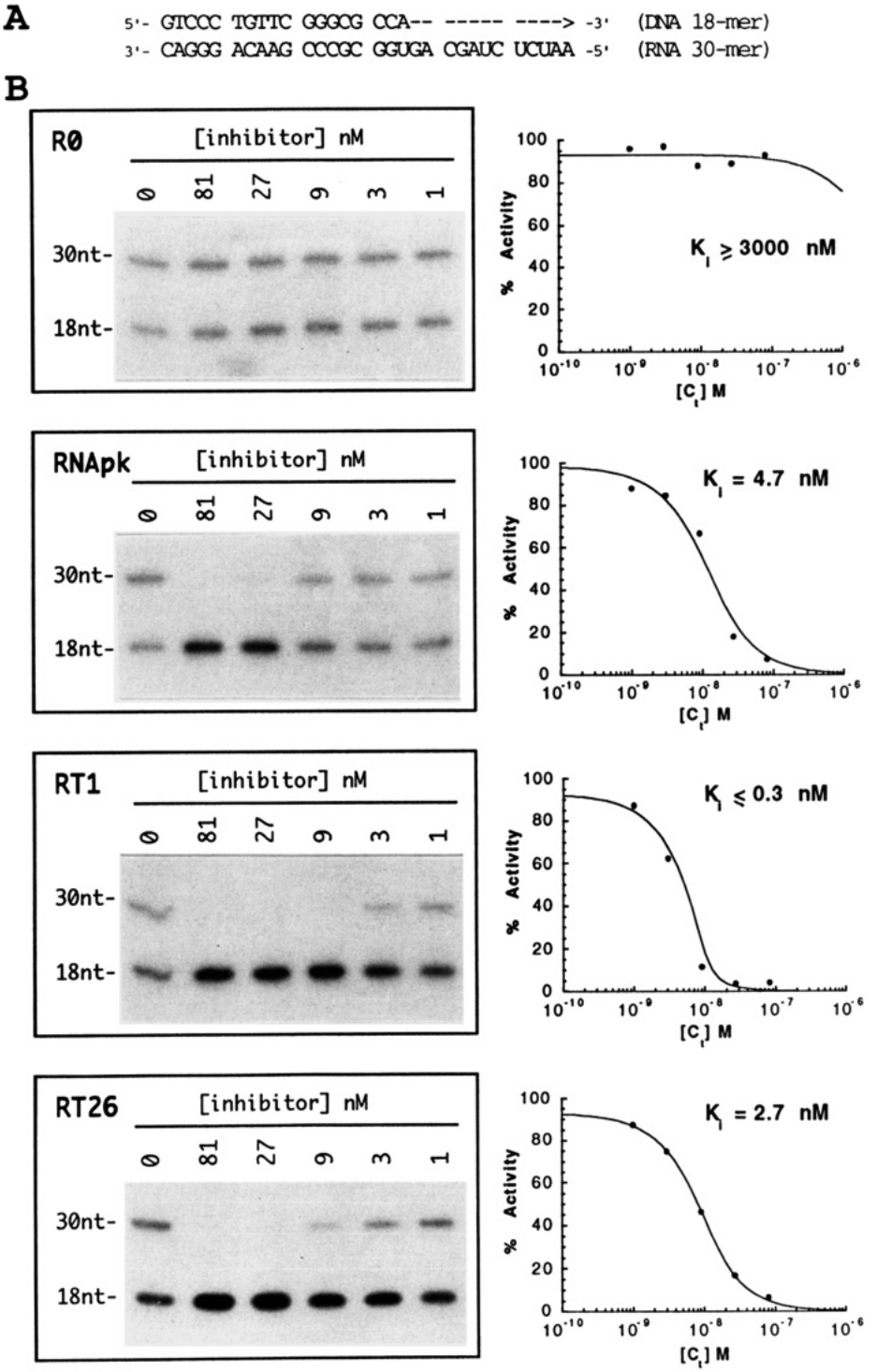

FIGURE 5: Inhibition of RNA-dependent DNA polymerase activity of HIV-1 RT. (A) The substrate for the inhibition assay was a DNA:RNA heteroduplex consisting of an 18-nucleotide end-labeled DNA primer identical in sequence to the 3′ end of $tRNA^{Lys,3}$ annealed to a 30-nucleotide RNA template whose sequence matches the genomic primer binding site and the first 12 transcribed nucleotides. **B.** Extension reactions were performed as described in Materials and Methods in the presence of 0, 81, 27, 9, 3, and 1 nM inhibitor as indicated. The two bands on the gels are the unextended DNA primer migrating as an 18-mer and the extended DNA product migrating as a 30-mer. The percent of primer extended as a function of inhibitor concentration is plotted for each inhibitor. K_i values were determined from the shape of the function using the algorithm described in Gill *et al.* (1991).

concentrations as high as 81 nM.

RT1 Competes with RNA Pseudoknot for RT Binding. The specific inhibition characteristics exhibited by both the RNA pseudoknot (RNA pk) and RT1t49 posed the question of whether two apparently dissimilar molecules, at least in composition and secondary structure, interact with HIV-1 RT at a common nucleic acid binding site. We addressed this by measuring the ability of the ssDNA ligand RT1 to maintain its specific binding contacts with HIV-1 RT in the presence of high concentrations of the RNA pk ligand. As shown graphically in Figure 9, when a stoichiometric equivalent of RNA pk was added, approximately one-half of the complexed RT1 was displaced, and nearly all was displaced when a large excess of RNA pk was added. The K_c value for RNA pk (3 nM) was determined using the algorithm described in Gill *et al.* (1991), and was consistent

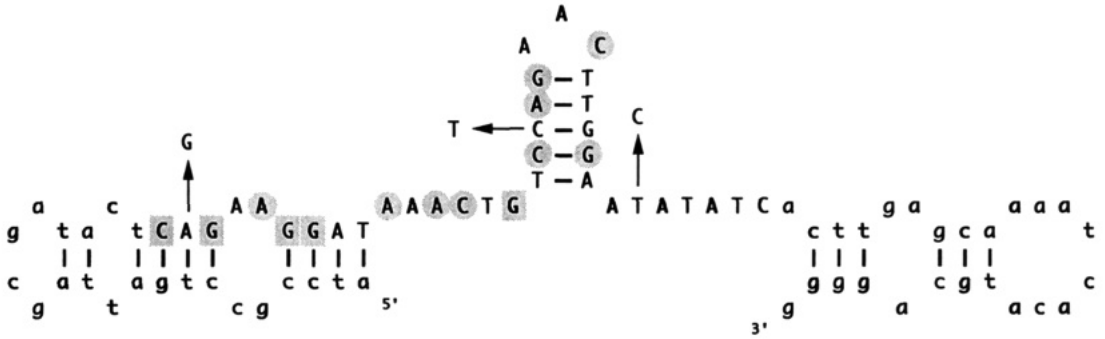

FIGURE 6: Summary of biased randomization SELEX of RT1. The proposed secondary structure of RT1 is illustrated with originally randomized positions indicated with capital letters and the 5′-invariant positions indicated with lowercase letters. Following six selections as described in Materials and Methods, 16 isolates from the enriched library were sequenced and aligned for an analysis of positional variation. Positions absolutely conserved are shaded with a square, those partially conserved (fewer individuals possess a substitution than would be expected at random) are shaded with a circle, and those preferring a substitution are indicated with an arrow. The proposed secondary structure of the internal loop motif is supported by these results, as nearly all of the positions comprising this element showed a high level of conservation.

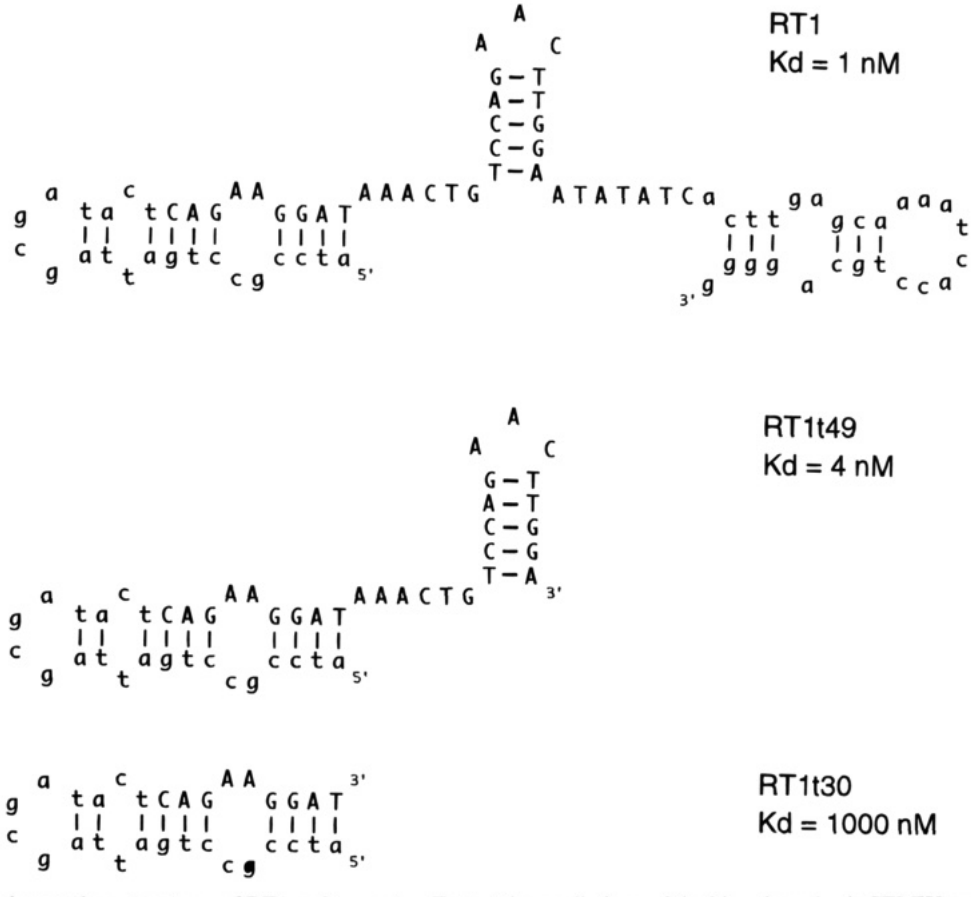

FIGURE 7: Predicted secondary structures of RT1 and truncates. To test the predictions of the biased synthesis SELEX experiment, truncated versions of RT1 were synthesized and assayed for affinity. Ligand RT1t49, a 49-mer oligonucleotide lacking the 3′-terminal 31 nucleotides, had a K_d value of 4 nM, only 4-fold worse than that of RT1 (1 nM). However, removal of an additional 19 nucleotides from the 3′ end to create RT1t30 eliminated specific binding (K_d = 1000 nM).

with the K_d value (5 nM) measured using the nitrocellulose filter binding assay described in Materials and Methods. These results indicate that binding of RT1 and RNA pk to HIV-1 RT are mutually exclusive. We are, however, unable to distinguish between an interaction of these ligands at a common site, at overlapping sites, or at spatially distinct sites.

DISCUSSION

A Diverse Set of ssDNA Ligands Inhibits HIV-1 RT Activity. Selection and amplification experiments of the kind described in this work are becoming an increasingly popular method of identifying a nucleic acid binding motif with high affinity for a target protein (Gold *et al.*, 1995). This selection yielded many different ssDNA motifs able to bind HIV-1 RT with high affinity and inhibit the polymerase activity of the enzyme. Ligands classified in families I and II appear to mimic a primer:template junction, while those in families III−VI possess different conserved helices with the potential to stack on each other to form longer flexible helices interrupted by loops, bulges, or mismatched base pairs. Perhaps the sequence nonspecific binding required of the enzyme *in vivo* to perform its activities is reflected in the

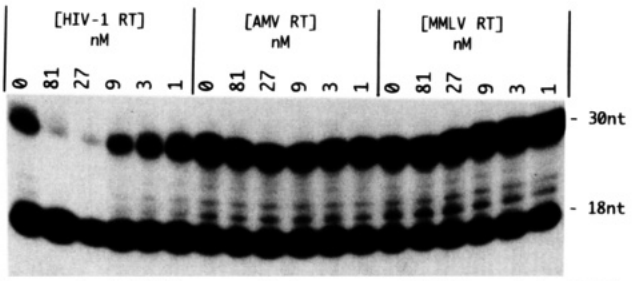

FIGURE 8: Inhibition specificity assay. Inhibition of the RNA-dependent DNA polymerase activity of three reverse transcriptases was performed as described in Figure 5 and Materials and Methods, with inhibitor RT1t49 present at the indicated concentrations.

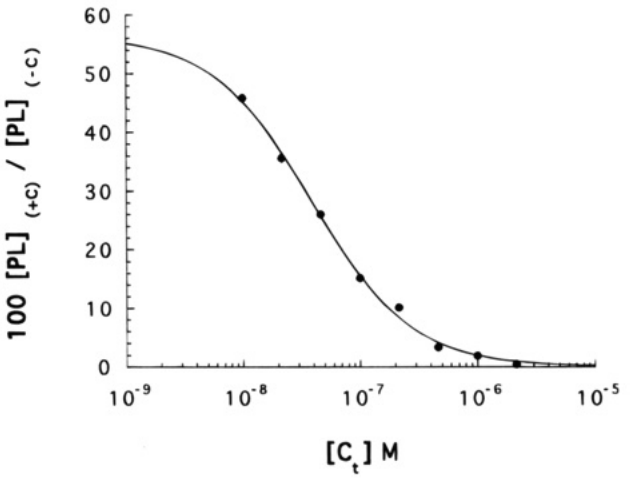

FIGURE 9: Competitive binding of RT1 and the RNA pseudoknot (RNA pk). The percent of RT1 bound in the presence of competitor ($[PL]_{(+C)}$) relative to the percent bound in the absence of competitor ($[PL]_{(-C)}$) is plotted as a function of RNA pk concentration ($[C_t]$).

sequence (and structural) variability of the high-affinity families. However, relative to nonspecific binding of ssDNA to HIV-1 RT, the selected ligands have up to 1000 times greater affinity for HIV-1 RT.

Nucleic Acid Binding Groove of HIV-1 RT Prefers a Bent Helix. A groove on the surface of HIV-1 RT, shown by the X-ray structure to extend from the polymerase catalytic site to the RNase H active site (Arnold *et al.*, 1992; Jacobo-Molina *et al.*, 1993; Kohlstaedt *et al.*, 1992; Krug & Berger, 1991), is the best candidate for the protein region contacted by the selected DNA ligands. The ability of RT10, RT12, and RT26 to be extended by HIV-1 RT demands that their 3′ termini be present in the polymerase active site when bound, likely positioned there by interactions between the proposed helix and the protein groove. Between 15 and 18 base pairs of a B-form helix with a bend in the middle are predicted to span the distance between the two active sites (Furfine & Reardon, 1991; Kohlstaedt *et al.*, 1992). The structure of an 18 base pair DNA duplex complexed to HIV-1 RT has been solved at 3 Å resolution, and it showed that the bound helix was bent to fit the binding groove (Jacobo-Molina *et al.*, 1993). A similar analysis of the *E. coli* DNA polymerase I Klenow fragment revealed a helical bend of about 80° in a DNA duplex bound by the active enzyme (Beese *et al.*, 1993; Freemont *et al.*, 1988). Each of the members of the six families shown in Figure 2 has multiple helices that could adopt structures conforming to the criteria above if stacked upon or aligned with each other in the groove. The distortion of DNA helices during interaction with protein is not uncommon, as structures of the nucleosome core particle (Richmond, 1984), the λ *cro* repressor (Brennan *et al.*, 1990), and the *E. coli* catabolite activator protein (CAP) (Schultz, *et al.*, 1991) bound to bent DNA helices have been reported. In our SELEX experiment, HIV-1 RT had the opportunity in the first selection to sample as many as 10^4 different helices with 18 consecutive Watson−Crick base pairs, as well as a large number of imperfect helices with intrinsic kinks and bends determined by their unique sequences. Imperfect helices with specific helix geometries may have higher affinities than perfect helices because the interaction energy required for distortion of a perfect helix is not needed when binding helices with intrinsic bends. Protein residues responsible for helix distortion would be available for contact with the already-bent selected ligands, resulting in a higher affinity relative to perfect helices.

Selected Ligands Inhibit Polymerase Activity of HIV-1 RT. The ability of the ssDNA ligands to inhibit the RNA-dependent DNA polymerase activity of HIV-1 RT provides further evidence for an interaction between RT and the selected ligands in the nucleic acid binding groove RT uses to recognize primer:template junction substrates. The ssDNA ligands could inhibit by competing with substrate for the nucleic acid binding site or by interfering with a step in the catalytic cycle. K_i values measured by the inhibition assay reflected K_d values measured by the filter binding assay, consistent with a competitive binding mechanism. Inhibition by RT1t49 is specific to the RT of HIV-1, as the RTs of AMV and MMLV were resistant to its inhibitory effects. Unique contacts must be made by this ssDNA ligand in the nucleic acid binding groove to confer such specificity. We dismiss as unlikely (given the intramolecular extension properties of ligands RT10, RT12, and RT26) the possibility that each of the high-affinity ssDNA ligands interacts with HIV-1 RT on its surface distant from the nucleic acid binding groove in a fashion prohibiting binding of the primer:template substrate.

Conserved Motifs Contribute to High Affinity. The internal loop motif shaded in Figure 3 is a secondary structure component able to form in multiple ligands classified into different primary structure families. The existence of this proposed motif in various contexts and the intolerance to mutation revealed by the biased randomization experiment support its existence and possible importance in RT binding. Although the low affinity of the motif in isolation negates the hypothesis that it alone confers tight binding, in combination with other components this motif contributes to the specific binding and inhibitory qualities of a ssDNA ligand. In ligand RT1t49, for example, the internal loop motif in combination with a small hairpin sufficiently represents the high affinity exhibited by ligand RT1. Ligand RT26 contains a GCGTGCTG element that contributes to its high affinity, as replacement of this conserved octamer with $(dA)_8$ abolishes high-affinity binding (data not shown). Although it may be sequestered in a small hairpin, the lack of secondary structure in this region in other individuals containing this octamer (RT2, RT31, RT33, and RT45; data not shown) suggests it is unpaired when interacting with HIV-1 RT. This is consistent with the structure of a primer:template junction where the genomic (template) strand is single-stranded and immediately downstream of the double-stranded priming region. Direct contact has been proposed

between the enzyme and the template strand near the polymerase active site (Arnold *et al.*, 1992). RT26 might mimic these interactions, achieving a large portion of its binding energy through specific contacts between unpaired nucleotides of the conserved octamer and amino acid side groups of HIV-1 RT.

HIV-1 RT Selected Imperfect Helices from a Degenerate ssDNA Library. The degenerate library containing 10^{14} different molecules was insufficient in number to saturate all possible 35-mers, but large enough to represent every possible 25-mer. Individuals with 35N regions complementary to the invariant end sequences were present in the degenerate library but were not selected, suggesting that a perfect helix defined by either of these sequences had an affinity too low to be enriched. In fact, every perfect helix of 11 base pairs or fewer (and some as large as 21 base pairs) was present in the degenerate library, yet none of the selected ligands had a perfect helix 18 base pairs long. Of the eight starred ligands in Figure 2, RT10 contains the largest continuous helix, 10 base pairs long (8 short of the 18 predicted to fill the groove). Our results support the premise that a perfect 18 base pair helix is not the ideal ligand conformation recognized by HIV-1 RT *in vitro* and that combinations of helices with numerous mismatches might create shapes more conducive for the interaction. HIV-1 RT shows a dramatic loss in its ability to bind template:primers comprising simple rA·dT duplexes fewer than 15 nucleotides in length (Reardon et al. 1991). This lends further support to the idea that the high-affinity ssDNA inhibitors afford a specific structural geometry not offered by linear duplex DNA template:primers.

The paradox that primer:template junctions recognized as natural substrates for the polymerase activity of RT *in vivo* are always perfectly matched helices and yet were not selected in our experiment can be justified if the structure of the substrate plays a role in a kinetic mechanism regulating its occupancy time in the nucleic acid binding site. As mentioned earlier, for the substrate to fit snugly in the groove, some amount of energy must be provided by the interaction to compensate for the strain created by bending the duplex. The enzyme may have evolved to have a specific off-rate when bound to a strained helix, promoting translocation after addition of a nucleotide. Our ligands were selected for high affinity alone and would not be influenced by regulation of this function. Specific bends in the helices defined by mismatched base pairs may have been selected because additional energy was not necessary to contort the helices to fit the groove. The incomplete extension of RT26 (Figure 4A) might reflect a ligand conformation that kinetically traps the enzyme in a bound form after addition of 5−8 nucleotides, preventing further translocation.

A Model for the Interaction of RT26 with HIV-1 RT. It is tempting to speculate on the orientation of the high-affinity ssDNA ligands when they are bound to HIV-1 RT. As mentioned earlier, the ability of RT26 to be extended implies positioning of the 3′-terminal G in the catalytic site precisely oriented for addition of the next nucleotide. This requirement would be satisfied by placing the helix of RT26 into the groove of HIV-1 RT, with the 3′ end of the priming strand adjacent to the three aspartic acid residues marking the proposed polymerase catalytic site (Larder *et al.*, 1987, 1989). This is consistent with the orientation adopted by the duplex crystallized with HIV-1 RT (Jacobo-Molina *et al.*, 1993). In the crystallized complex the DNA helix shows A-form character near the polymerase active site and B-form character near the RNase H active site with a 40−45° bend in the helix at the junction separating the A- and B-form duplexes. The most extensive protein−DNA contacts involved the sugar−phosphate backbone of the DNA duplex and amino acid side groups of the "palm, thumb, and fingers" domains comprising the polymerase active site and nucleic acid binding groove of HIV-1 RT (Kohlstaedt *et al.*, 1992). The palm and thumb are believed to act as a "clamp" to position the primer:template junction in the polymerase active site.

If the 3′ G of RT26 were positioned in the polymerase active site, the nucleotides of the proposed internal loop motif would be correctly oriented for direct contact with the α-helix of the thumb domain (Kohlstaedt *et al.*, 1992) via the minor groove. We have modeled RT26 onto the nucleic acid binding groove of the HIV-1 RT dimer using the crystal structure coordinates of Jacobo-Molina *et al.* (1993) and placing RT26 to conform to the backbone phosphate coordinates in the crystal structure (not shown). If RT26 were bound in a fashion similar to the crystallized complex, the internal loop motif would be at the site of the helix bend, and it may have been selected for this purpose. When complexed, one α-helix of the enzyme's thumb domain would contact the backbone of the DNA primer strand, while the other α-helix would contact the backbone of the template strand, as seen in the crystal structure. The consensus octamer (GCGTGCTG) could then interact with HIV-1 RT at the periphery of the polymerase active site where the template strand ahead of the primer would reside.

The dependence of RT26 affinity on ionic strength (an affinity increase of nearly 10-fold results from reducing the KOAc concentration from 250 to 50 mM; data not shown) illustrates the importance of electrostatic interactions. This is not surprising for an enzyme that recognizes DNA duplexes whose surfaces are covered with backbone phosphate anions. However, to be discriminated among all of the other DNA duplexes present in the degenerate library, RT26 must possess a characteristic surface that is preferentially recognized by HIV-1 RT. A unique helical twist or kink might create a preferred shape by placing a phosphate group in a stereospecific orientation ideal for contacting a lysine or an arginine or by making a base in the major or minor groove available for hydrogen bonding. We propose that the function of the internal loop motif might be the provision of a flexible hinge that facilitates bending of the helix and stereospecifically presents the backbone phosphates for direct contact by protein side groups of the thumb domain. The other selected ligands possessing this motif would make similar contacts with RT at this element, in addition to ligand-specific contacts at distant regions of the protein.

RNA Pseudoknot and Selected ssDNA Ligands Share Little Structural Similarity. Unlike the RNA ligands selected for HIV-1 RT affinity, each containing a consensus pseudoknot with a specific surface geometry recognized by the enzyme (Tuerk *et al.*, 1992), the recognition elements conferring specificity of the ssDNA ligands are more subtle. These ligands appear to achieve at least a portion of their high affinity by forming uniquely shaped helices able to lie across the nucleic acid binding groove of the enzyme. We have identified a library of ssDNA ligands with very high affinity for HIV-1 RT but sharing no single consensus motif

responsible for the affinity or specificity of the interaction. Only the overall structures of the ligands may be similar, potentially forming long helices with distortable positions able to provide kinks and bends. Perhaps recognition specificity is provided primarily by a preferred helix geometry, with stereospecific kinks bending the helix in such a way that it fits snugly in the nucleic acid binding groove of the enzyme and exposing core nucleotides to protein side groups. An X-ray crystal structure analysis of ligand RT1t49 complexed with HIV-1 RT has been initiated and will provide the most direct evidence for the nature of these interactions. We have also begun a more rigorous analysis of the binding requirements of RT1 and RT26 to minimize their size and make them more amenable to structure determination using nuclear magnetic resonance. Solving the three-dimensional structures of the free DNA or the DNA−protein complex not only will determine the accuracy of our model for specific recognition by HIV-1 RT but will provide additional clues to the more universal question of how HIV-1 RT interacts with duplex DNA. Understanding these fundamental activities of HIV-1 RT might provide a foundation for rational design of therapeutic agents.

ACKNOWLEDGMENT

The authors thank David J. Schneider and Naomi Fields for experimental help and Dom Zichi, Brenda Javornik, and Jeff Davis for computational help. J.F. thanks the members of the Gold lab for help during her sabbatical visit.

REFERENCES

Altschul, S. F. (1989) *J. Theor. Biol. 138*, 297−309.
Altschul, S. F., & Erickson, B. W. (1986) *Bull. Math. Biol. 48*, 603−616.
Arnold, E., Jacobo-Molina, A., Nanni, R. G., Williams, R. L., Lu, X., Ding, J., Clark, A. D., Zhang, A., Ferris, A. L., Clark, P., Hizi, A., & Hughes, S. (1992) *Nature 357*, 85−89.
Baltimore, D. (1970) *Nature 226*, 1209.
Barat, C., Lullien, V., Schatz, O., Keith, G., Nugeyre, M. T., Gruninger-Leitch, F., Barre-Sinoussi, F., LeGrice, S. F. J., & Darlix, J. L. (1989) *EMBO J. 8*, 3279−3285.
Beese, L. S., Derbyshire, V., & Steitz, T. A. (1993) *Science 260*, 352−355.
Brennan, R. G., Roderick, S. L., Takeda, Y., & Mathews, B. W. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 8165.
Carey, J., Cameron, V., de Haseth, P., & Uhlenbeck, O. C. (1983) *Biochemistry 22*, 2601−2609.
Davies, J. F., Hostomska, Z., Hostomsky, Z., Jordan, S. R., & Matthews, D. A. (1991) *Science 252*, 88−95.
Davis, J. P., Janjic, N., Pribnow, D., & Zichi, D. A. (1995) *Nucleic Acids Res.* (in press).
Freemont, P. S., Friedman, J. M., Beese, L. S., Sanderson, M. R., & Steitz, T. A. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 8924−8928.
Furfine, E. S., & Reardon, J. E. (1991) *J. Biol. Chem. 266*, 406−412.
Gilboa, E., Mitra, S. W., Goff, S., & Baltimore, D. (1979) *Cell 18*, 93−100.
Gill, S. C., Weitzel, S. E., & von Hippel, P. H. (1991) *J. Mol. Biol. 220*, 307−324.
Goff, S. P. (1990) *J. Acquired Immune Defic. Syndr. 3*, 817.
Gold, L., Polisky, B., Uhlenbeck, O., & Yarus, M. (1995) *Annu. Rev. Biochem.* (in press).
Hansen, J., Schulze, T., Mellert, W., & Moelling, K. (1988) *EMBO J. 7*, 239−243.
Hostomsky, Z., Hostomska, Z., Hudson, G. O., Moomaw, E. W., & Nodes, B. R. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 1148−1152.
Jacobo-Molina, A., Ding, J., Nanni, R., Clark, A., Lu, X., Tantillo, C., Williams, R. L., Kamer, G., Ferris, A. L., Clark, P., Hizi, A., Hughes, S., & Arnold, E. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 6320−6324.
Jaeger, J. A., Turner, D. H., & Zuker, M. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 7706−7710.
Kohlstaedt, L. A., Wang, J., Friedman, J. M., Rice, P. A., & Steitz, T. A. (1992) *Science 256*, 1783−1790.
Krug, M. S., & Berger, S. L. (1991) *Biochemistry 30*, 10614−10623.
Larder, B. A., Purifoy, D. J., Powell, K. L., & Darby, G. (1987) *Nature 327*, 716.
Larder, B. A., Kemp, S. D., & Purifoy, D. J. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 4803.
Panet, A., Haseltine, W. A., Baltimore, D., Peters, G., Harada, F., & Dahlberg, J. E. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 2535−2539.
Peliska, J. A., & Benkovic, S. J. (1992) *Science 258*, 1112−1118.
Reardon, J. E., Furfine, E. S., & Cheng, N. (1991) *J. Biol. Chem. 266*, 14128−14134.
Richmond, T. J. (1984) *Nature 311*, 532.
Schneider, D., Gold, L., & Platt, T. (1993) *FASEB J. 7*, 201−207.
Schultz, S. C., Shields, G. C., & Steitz, T. A. (1991) *Science 253*, 1001−1007.
Subbiah, S., & Harrison, S. C. (1989) *J. Mol. Biol. 209*, 539−548.
Temin, H. M., & Mizutani, S. (1970) *Nature 226*, 1211.
Tou, J. T., & Gonzales, R. T. (1974) in *Pattern Recognition Principles* (Kalaba, R., Ed.) pp 86−104, Addison-Wesley Publishing Co., Reading, MA.
Tuerk, C., & Gold, L. (1990) *Science 249*, 505−510.
Tuerk, C., MacDougal, S., & Gold, L. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 6988−6992.
Zuker, M. (1989) *Science 244*, 48−52.

BI950659V