A Specific RNA Structural Motif Mediates High Affinity Binding by the HIV-1 Nucleocapsid Protein (NCp7)

PATRICK ALLEN,1 BRIAN COLLINS,2 DAVID BROWN, ZDENEK HOSTOMSKY,* and LARRY GOLD

University of Colorado at Boulder, Department of Molecular, Cellular and Developmental Biology, Campus Box 347, Boulder, Colorado 80309-0347; and *Agouron Pharmaceuticals, Inc., 3565 General Atomics Court, San Diego, California 92121-1121

Received July 1, 1996; revised August 30, 1996; accepted September 10, 1996

Current research indicates that the nucleocapsid protein (NCp7) of human immunodeficiency virus type 1 (HIV-1) interacts with a variety of RNA substrates during the progression of the viral life cycle. The RNA features specifically recognized by the protein, however, have yet to be identified. SELEX was used to generate a set of RNAs whose affinities for nucleocapsid were on the order of $2 \times 10^{-10}$ M. Comparative analysis revealed that each RNA contains a highly conserved fourteen nucleotide sequence-block. Computer modeling and structure probing experiments indicate that the RNA ligands use the consensus sequence to fold into hairpins with an identical asymmetric bulge. The presence of the nucleocapsid protein protects the asymmetric bulge from ribonuclease attack, suggesting that it is the key element in protein recognition. A search for similar structural motifs within the HIV genome reveals several potential interaction sites for the nucleocapsid protein.

INTRODUCTION

One of the first events during the replication of HIV-1 and all other known retroviruses is conversion of genomic RNA to DNA using the viral reverse transcriptase (RT) and other host cell components (for review; Coffin, 1982; Varmus and Brown, 1989; Wain-Hobson, 1994). The viral RNA is reverse transcribed by RT with a host tRNA being used as the primer (reviewed in Wong-Staal, 1990; Vaihnnav and Wong-Staal, 1991). Retroviral nucleocapsid proteins are intimately involved in annealing of the tRNA primer to the genomic RNA. It has been reported that nucleocapsid proteins from Mason-Pfizer monkey virus (MPMV) and HIV-1 facilitate primer annealing (Dib-Hajj et al., 1993; Tsuchihashi and Brown, 1994; Lapadat-Tapsolsky et al., 1995). Moreover, it has been shown that the nucleocapsid proteins of Rous sarcoma virus (RSV) and murine leukemia virus (MuLV) are directly involved in annealing the tRNA primer onto the primer binding site under physiological conditions (Prats et al., 1988; Batat et al., 1989). NC proteins from the human immunodeficiency virus have also been implicated in dimerization of genomic RNA (Darlix et al., 1990; Sakaguchi et al., 1993), encapsidation of full-length genomic RNA (Rizvi and Panganiban, 1993; South and Summers, 1993; Zhang and Barklis, 1995; Tanchou et al., 1995), and interactions with reverse transcriptase which influence strand transfer and transcription rate (Weiss et al., 1992; Peliska et al., 1994; Rodriguez-Rodriguez et al., 1995; Ji et al., 1996).

Retroviral nucleocapsid proteins are precursor products (Mervis et al., 1988; Morellet et al., 1992). HIV-1 nucleocapsid protein contains two zinc finger binding domains with the general structure Cys-X2-Cys-X4-His-X4-Cys (Gorelick et al., 1993). Mutations in each of these finger domains result in different functional and structural defects (Gorelick et al., 1990, 1993; Julian et al., 1993). Similar types of mutations in RSV and MuLV nucleocapsid proteins result in defects in RNA packaging and dimer formation (Méric and Spahr, 1986; Méric and Goff, 1989). Results from these studies suggest a role for the nucleocapsid protein during reverse transcription (Méric and Goff, 1989; Weiss et al., 1992) and possibly during infection (Méric and Spahr, 1986).

Sequences near the 5’ end of the mature HIV-1 viral RNA are apparently involved in encapsidation (Aldovini and Young, 1990). It appears these sequences interact specifically with the nucleic acid binding zinc finger domains of HIV-1 NC. Mutations in either of the two Zn$^{2+}$ fingers or the encapsidation site result in similar defects. The 5’ end of retroviral RNAs have been rigorously studied and thus the structural and functional characterization is extensive (Bender et al., 1978; Murti et al., 1981; Darlix et al., 1982; Aldovini and Young, 1990; Harrison and Lever, 1992; Surovoy et al., 1993).

Although HIV-1 nucleocapsid plays a crucial role in so many different viral processes, the nature of its interaction with substrate RNAs has yet to be clearly defined.
To better understand the specific RNA features recognized by the nucleocapsid protein, we used SELEX to generate a set of RNA molecules with high affinity for the HIV-1 protein. Sequence comparison and structural characterization of the RNAs identified a consensus binding site of the nucleocapsid protein. A search of the HIV genome for structures resembling the SELEX-generated consensus revealed several potential targets for the protein.

MATERIALS AND METHODS

PCR amplification and selection

Amplification by PCR and the SELEX procedure were carried out essentially as described (Tuerk and Gold, 1990). Five picomoles of a random pool of DNA oligomers were used as template for PCR amplification for 8 cycles in the initial round. Complementary DNA of the selected pool of RNA from subsequent rounds of SELEX was PCR amplified 18 cycles. PCR reactions were carried out in 50-μl volumes containing 200 pmol of each primer, 2 mM dNTP’s, 5 units of Thermus aquaticus DNA polymerase (Perkin-Elmer-Cetus) in PCR buffer (10 mM Tris-Cl, pH 8.4, 50 mM KCl, 7.5 mM MgCl2, 0.05 mg/ml BSA). Conditions used for primer annealing, extension, and denaturation of DNA strands are identical to those previously described (Allen et al., 1995).

PCR products were transcribed using T7 RNA polymerase in vitro in a 200-μl reaction volume (Tuerk and Gold, 1990). T7 transcripts were purified from a 6% polyacrylamide, 7 M Urea gel and eluted by crushing gel pieces in a Sodium Acetate/EDTA solution. For each round of SELEX, 50 pmol of the selected pool of RNA was phosphatased using Calf Intestinal Alkaline Phosphatase (Biolabs) and 25 pmol was 5’ end-labeled using [γ-32P]-ATP with polynucleotide kinase (Boehringer) for 30 min. Kinased RNA was purified and about 150,000 cpm was used to follow the fraction of RNA retained on nitrocellulose filters during selection. Protein and nucleic acid conditions for selection and elution of RNA from nitrocellulose filters were as described (Allen et al., 1995).

Selection in the presence of nonamplifiable competitor RNA

Selections were done using two buffer conditions where the only difference between the buffers was the concentration of NaCl (Buffer A, 10 mM NaOAc (pH 5.3), 2 mM 2-mercaptoethanol, 0.1 mM ZnCl2, 5 mM MgCl2, 200 mM NaCl; Buffer B, 10 mM NaOAc (pH 5.3), 2 mM 2-mercaptoethanol, 0.1 mM ZnCl2, 5 mM MgCl2, 400 mM NaCl). The parent pool (500 pmol) of RNA was used in the first round of selection along with about 100 fmol of γ-labeled RNA from the same pool. RNA was recovered and cDNA synthesized and PCR amplified. In rounds 2 through 8, 250 pmol of the selected RNA was used with labeled RNA from that pool to follow the amount of recovery. In each round, other than round 1, we selected about 3% of the total input RNA or a smaller fraction if counts were greater than 10-fold above the control (cpm bound in the absence of protein). For rounds 9 and 10 we reduced the amount of RNA to 25 pmol. During the last 7 rounds (rounds 11 – 17), 25 pmol of cold-selected RNA was used along with 500 pmol of a nonamplifiable random pool of competitor RNA as described (Allen et al., 1995).

Expression constructs

The capsid – nucleocapsid (CA – NC) and nucleocapsid (NC) segments of the HIV-1 gag protein were amplified from the clone pBH10 (Ratner et al., 1985) and inserted downstream of the phage T7 promoter of the multicopy expression vector pGZ/H6, derived from the phagemid pTZ18R (Mead et al., 1986). The plasmid pGZ/H6CANC amplified 18 cycles. PCR reactions were carried out in 6 μl volumes containing 200 pmol of each primer, 2 mM hexahistidine tag (MKLHHHHHGY) was followed by dNTP’s, 5 units of Thermus aquaticus DNA polymerase residues Gln127 through Asn432 of HIV-1 gag. The plasmid pGZ/NC55-1 expressed a 55-residue form of NC starting from the naturally occurring Met378 through Asn432 of HIV-1 gag. An Escherichia coli K12 strain AP401K harboring plasmids pGZ/H6CANC6 or pGZ/NC55-1 was induced for expression with bacteriophage mGP1-2 (gift from S. Tabor), an M13 derivative containing the gene for T7 RNA polymerase, as described (Hostovsky et al., 1989).

Purification of nucleocapsid p7

The E. coli cells, harvested 3 hr after induction, were resuspended in 50 mM Tris/HCl, pH 8.0, and lysed in a microfluidizer (Microfluidics Corp.). After removal of cell debris by centrifugation, the nucleic acids were precipitated from the soluble lysate by stepwise addition of Polymin P to a final concentration of 0.5%. Powdered ammonium sulfate was added to the supernatant to a final concentration of 40%. Upon removal of pellet by centrifugation, more ammonium sulfate was added to the supernatant until its concentration reached 80%. The pellet was taken up in 50 mM Tris/HCl, pH 8.0, 5 mM 2-mercaptoethanol, 10% glycerol, and 50 mM NaCl (Buffer C). After dialysis against buffer C, the sample was applied to a FPLC MonoQ column (Pharmacia). The flow-through fractions were pooled and applied directly to an FPLC MonoS column (Pharmacia). The protein was released by a linear gradient of 50 mM - 1 M NaCl in buffer C. The pooled nucleocapsid containing fractions were dialysed against buffer D (25 mM Tris/HCl, pH 8.0, 50 mM NaCl, 2 mM 2-mercaptoethanol, and 0.1 mM ZnCl2) and concentrated to 1.7 mg/ml.
Purification of capsid-nucleocapsid fusion protein

Soluble lysate of E. coli harboring pGZ/H6CANC-6, prepared as in the NC purification described above, was applied to a Ni$^{2+}$-NTA-agarose column (Quiagen). The column was washed extensively with 20 mM Tris/HCl, pH 8.0, containing 1 M NaCl, followed by 20 mM Tris/HCl, pH 8.0. The bound CANC fusion protein was eluted with a linear gradient of 1–200 mM imidazole in 20 mM Tris/HCl buffer (pH 8.0). The CANC (released at approximately 60 mM imidazole concentration) was dialysed against buffer D.

Cloning and sequencing

PCR amplified DNA from the round 16 selected-pool of RNA was phenol and chloroform extracted and ethanol precipitated. The extracted PCR DNA was digested using BamH1 and HindIII (Biolabs) and subcloned into pUC18. Ligation was carried out at room temperature for 2 hr, after which the reaction was phenol and chloroform extracted and used to electroporate competent cells. Twenty-five transformants from the SELEX performed in the NC purification described above, was prepared as in the NC purification described above, was applied to a Ni$^{2+}$-NTA-agarose column (Quiagen). The column was washed extensively with 20 mM Tris/HCl, pH 8.0, containing 1 M NaCl, followed by 20 mM Tris/HCl, pH 8.0. The bound CANC fusion protein was eluted with a linear gradient of 1–200 mM imidazole in 20 mM Tris/HCl buffer (pH 8.0). The CANC (released at approximately 60 mM imidazole concentration) was dialysed against buffer D.

Cloning and sequencing

PCR amplified DNA from the round 16 selected-pool of RNA was phenol and chloroform extracted and ethanol precipitated. The extracted PCR DNA was digested using BamH1 and HindIII (Biolabs) and subcloned into pUC18. Ligation was carried out at room temperature for 2 hr, after which the reaction was phenol and chloroform extracted and used to electroporate competent cells. Twenty-five transformants from the SELEX performed in the low salt buffer and 30 from the high salt experiment were picked and their DNAs sequenced.

Binding assays

Binding assays were done by adding 5 µl of HIV-1 nucleocapsid p7 protein, at the appropriate concentrations, to 45 µl of binding buffer (10 mM NaOAc (pH 5.3), 2 mM 2-mercaptoethanol, 0.1 mM ZnCl2, 5 mM MgCl2, 100 mM NaCl) on ice, then adding 80,000 cpm of kinased RNA (<200 fmol) in a volume of 3 to 4 µl. This mix was incubated at 37°C for 20 min. The reactions were passed over nitrocellulose filters that had been preequilibrated in the same buffer. The protein bound filters were washed five times with one milliliter of 50 mM Tris–Cl, pH 7.5, solution. Filters were dried and counted incocktail. The protein used in these experiments was frozen and thawed only once. Dissociation constants were determined as previously described (Carey et al., 1983).

Determination of the minimum binding domain

Boundary determination experiments were carried out essentially as previously published (Tuerk et al., 1990) with the following exceptions. Our transcripts were treated with calf intestinal alkaline phosphatase (Biolabs). For each reaction, 25 pmol of partially hydrolyzed RNA was incubated with nucleocapsid protein in binding buffer in a final reaction volume of 50 µl. Nucleocapsid concentrations were 0.5 × 10$^{-9}$, 1 × 10$^{-9}$, and 5 × 10$^{-9}$ in each case. Reactions were quenched as previously described (Chen and Gold, 1994). RNA fragments retained on filters were eluted using 150 µl of 1% SDS/0.3 M NaOAc/1 mM EDTA solution, then phenol extracted twice, and chloroform extracted, and the fragments were precipitated with ethanol.

Chemical and enzymatic modification of selected RNAs

RNAs were partially digested using ribonuclease S1 (0.2 units/reaction). For each enzymatic digestion carried out with NC present in the reaction mixture, the concentration of protein varied from 0.05 µM to 10 µM and the amount of RNA used was 4 pmol per reaction. Nucleocapsid and RNA was incubated together in binding buffer at 37°C for 20 min then transferred to ice for 5 min before adding RNase S1 for an additional 5 min. The RNase S1 was preincubated at 4°C before adding to reactions. The mix was phenol extracted three times, chloroform extracted twice, ethanec precipitated, and stored at −20°C. Primer extensions (Stern et al., 1988) were done the day of nuclease treatment to identify positions that are accessible to the enzyme.

In vitro inhibition assay

The 5’ end of the synthetic DNA oligonucleotide, Con+, (5’-CAATGACCGCATGGGATCCGTTGGAAAAATCTCTAGCAGT-3’) was labeled with [γ-32P]ATP using polynucleotide kinase (Boehringer). Labeled con+ (10 nM) was incubated at 37°C with equimolar ratio of the complementary strand, con−, in binding buffer containing various concentrations of NaCl for 6 min and then run on an 8% native acrylamide gel to determine the salt concentration that provides approximately 50% duplex. Binding buffer containing 200 mM NaCl converted 40 to 70% labeled con+ to duplex. This buffer was used and the concentration of each oligo was reduced keeping the reaction time constant to find an oligo concentration that would give less than 10% duplex formation under these conditions.

The conditions in the reaction used for the inhibition assay were as follows. Four hundred picomoles of each DNA oligo (con+/-) was mixed with 10 nM nucleocapsid for 6 min at 37°C in binding buffer containing 200 mM NaCl in a reaction volume of 30 µl. After this time, reactions were stopped by placing on ice and loaded on gel after 2 min. Gels were run at 250 V for 3.5 hr. In reactions where random RNA was added as inhibitor, the concentrations ranged from 6500 nM to 9 nM. Where selected RNAs were used, the range was 720 nM to 0.11 nM.

Synthesis of RNA truncates

Primers that anneal at approximately 10 nucleotide increments from the 3’ end of the 400-23 RNA were used to synthesize cDNAs. Primers and RNAs were heated to 90°C and cooled slowly to 45°C to anneal and primer extension was carried out at 48°C. cDNAs were then PCR amplified using the same 3’ primers with the original 5’ primer.
TABLE 1
Sequences of RNAs That Resulted from SELEX Using HIV-1 Nucleocapsid Protein p7

<table>
<thead>
<tr>
<th>Name</th>
<th>Number of Isolates</th>
<th>RNA Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>400-18</td>
<td>12</td>
<td>GGGACUCAGAAUAAAACUCUAUAUAACCTGAGACCACCCUCUCUCUCUAAUACUGUAGGCGCCGGAUCCGCGC</td>
</tr>
<tr>
<td>400-42</td>
<td>10</td>
<td>GGGACUCAGAAUAAAACUCUAUAUAACCTGAGACCACCCUCUCUCUCUAAUACUGUAGGCGCCGGAUCCGCGC</td>
</tr>
<tr>
<td>400-31</td>
<td>8</td>
<td>GGGACUCAGAAUAAAACUCUAUAUAACCTGAGACCACCCUCUCUCUCUAAUACUGUAGGCGCCGGAUCCGCGC</td>
</tr>
<tr>
<td>400-17</td>
<td>7</td>
<td>GGGACUCAGAAUAAAACUCUAUAUAACCTGAGACCACCCUCUCUCUCUAAUACUGUAGGCGCCGGAUCCGCGC</td>
</tr>
<tr>
<td>400-23</td>
<td>4</td>
<td>GGGACUCAGAAUAAAACUCUAUAUAACCTGAGACCACCCUCUCUCUCUAAUACUGUAGGCGCCGGAUCCGCGC</td>
</tr>
<tr>
<td>400-13</td>
<td>4</td>
<td>GGGACUCAGAAUAAAACUCUAUAUAACCTGAGACCACCCUCUCUCUCUAAUACUGUAGGCGCCGGAUCCGCGC</td>
</tr>
<tr>
<td>400-15</td>
<td>3</td>
<td>GGGACUCAGAAUAAAACUCUAUAUAACCTGAGACCACCCUCUCUCUCUAAUACUGUAGGCGCCGGAUCCGCGC</td>
</tr>
<tr>
<td>400-38</td>
<td>1</td>
<td>GGGACUCAGAAUAAAACUCUAUAUAACCTGAGACCACCCUCUCUCUCUAAUACUGUAGGCGCCGGAUCCGCGC</td>
</tr>
<tr>
<td>400-27</td>
<td>1</td>
<td>GGGACUCAGAAUAAAACUCUAUAUAACCTGAGACCACCCUCUCUCUCUAAUACUGUAGGCGCCGGAUCCGCGC</td>
</tr>
<tr>
<td>400-12</td>
<td>1</td>
<td>GGGACUCAGAAUAAAACUCUAUAUAACCTGAGACCACCCUCUCUCUCUAAUACUGUAGGCGCCGGAUCCGCGC</td>
</tr>
</tbody>
</table>

Note. Bases in italics are fixed positions used for PCR amplification. Nucleotides that are highly conserved are underlined. In addition to being underlined, the fourteen nucleotide block is also in bold face. Numbers to the left of sequences represent the number of isolates bearing this sequence. All sequences are presented in a 5' to 3' orientation.

The resulting product was used to transcribe 3' truncated RNAs. In addition to in vitro transcriptions, the shortest truncate (T-23) was also made synthetically.

RESULTS

Sequence analysis

Sequencing a total of 55 clones isolated after 16 complete rounds of SELEX resulted in multiple copies of 14 distinct sequences (Table 1). Each sequence includes a 14-nucleotide block which is nearly identical. This highly conserved 14-base sequence-block occurs at different positions within the randomized region. Six of the isolates have single-base substitutions within the 14-nucleotide sequence-block (Table 1). Five of these are single base substitutions (U to A) at the ninth position of the block, while the other is a C to U base change at the fourteenth position. All other positions in the conserved block are identical among the clones. Besides the conserved sequence-block there is no other sequence homology shared among the selected sequences with the exception of a highly conserved cytosine at the last position of the random region. In general, the conserved sequence is located proximal to the 5' end of the random region.

Filter binding studies

The dissociation constants for 12 of the 14 RNA species were determined by nitrocellulose filter binding. These experiments were carried out in duplicates. All binding curves were done using γ-32P-labeled RNA in binding buffer containing 100 mM NaCl. Dissociation constants (Kd's) ranged from 2 to 6 nM (Fig. 1). Kd's of the two round 16 selected pools of RNA for nucleocapsid protein were likewise under 10 nM (data not shown). The Kd for the initial random pool of RNA in the same buffer could not be determined since there was no significant binding above background at the highest protein concentration (10^-7 M).

Binding affinities were determined for 6 of the 12
isolates for a precursor fusion protein derived from the HIV-1 gag region [CA-NC; Capsid (p24) fused to Nucleocapsid (p7)]. The 322 amino acid polypeptide product, which includes the entire 55 amino acid sequence of the NC used in the SELEX experiments, had no detectable binding affinity for these RNAs (data not shown). Two possible explanations for this result are that the RNA binding site of nucleocapsid is blocked in the fusion protein, or the conformation of nucleocapsid within the fusion is altered yielding a protein that lacks the RNA binding site. SELEX was carried out using the same gag fusion protein and the winning ligands from those experiments bear no resemblance to the NC ligands (Allen, Liss, and Gold, unpublished).

RNA structure analysis

RNA transcripts of the individual clones were sequenced directly and entered into the Zuker folding program. The algorithm predicted that all 55 clones folded into the same secondary structure. The energy minimization for each RNA molecule was calculated to be $-11 \text{kcal/mol}$ or better. The consensus structure was a stem formed between the first 10 nucleotides at the fixed 5’ end and 10 of the 14 nucleotides from the conserved block with a loop that varied in size from 13 nucleotides (clone 400-23) to 23 nucleotides (clone 400-18). In the center of the 10-bp stem is a 4-base internal bulge whose primary sequence, 5’AACU3’, is highly conserved (Fig. 2). The balance of the nucleotides that were originally randomized are modeled as an unstructured region. The highly conserved cytosine at the last position in the random region is paired with a G in the 3’-fixed region. The rest of the 3’-fixed nucleotides fold into two hairpin structures with the terminal 3’ C unpaired. There are two noncanonical G-U pairs in the consensus structure. Approximately 75% of the modeled Watson–Crick interactions are G-C base pairs. There is good support for this structure from experiments using structure-sensitive probes and primer extension (Fig. 5). Because all of the stem regions involve base-pairing with bases that were fixed, there is no comparative evidence to support our model.

Boundary analysis

Boundary experiments were carried out in order to determine the minimum RNA structure required for binding to HIV-1 nucleocapsid protein. The 5’-end-labeled RNAs were partially alkaline hydrolyzed and the fragments were incubated with nucleocapsid protein and passed over nitrocellulose filters. Fragments with affinities high enough to survive washes were retained. These RNAs were eluted from the filters and analyzed on acrylamide gels. In every case, the deletion of nucleotides from the 3’ end is tolerated up to, but not including, the last nucleotide in the sequence block (Fig. 3). In other words, full-length RNAs and molecules bearing sequential deletions from the 3’ end are bound by nucleocapsid and retained on filters except for deletions into the conserved 14-nucleotide block (Fig. 3). This indicates that the stem between the 5’ end and the nucleotides of the 3’ end of the conserved block is required for high affinity binding. When the 3’ end is labeled and the same hydrolysis and binding experiments are carried out, only intact RNAs were retained on filters (data not shown). Thus, any deletion at the 5’ end results in failure to bind the nucleocapsid protein.

Binding studies of RNA truncates

The minimum binding domain predicted by the boundary analysis for the RNA identified as 400-23 was used for further characterization of the RNA binding site. This
RNA RECOGNITION BY HIV-1 NUCLEOCAPSID

Fig. 3. Autoradiograph showing results from boundary experiments. 17, 400-17; 23, 400-23; 13, 400-13. T1, partial RNase T1 digest; AH, partial alkaline hydrolysis; lane 1, contains 0.5 \( \times 10^{-9} \) M nucleocapsid; lane 2, contains 1 \( \times 10^{-9} \) M nucleocapsid; lane 3, contains 5 \( \times 10^{-9} \) M nucleocapsid; G's are T1-specific cleavage sites and the numbers are the position in the RNA from the 5' end. Arrows indicate the 3' terminal C in the conserved 14 nucleotide block.

RNA fragment, referred to as T-23, is 37 nucleotides long. In addition to this fragment, we made an RNA where the loop was substituted with a UUCG tetra-loop resulting in a fragment 28 nucleotides long (BC2). Both of these RNAs were synthesized on an Applied Biosystems Model 394 DNA/RNA synthesizer as well as transcribed in vitro from PCR products. Neither of the truncates bound the nucleocapsid protein as well as the full-length RNA, therefore other truncates were made and their \( K_d \) values were determined.

Truncated molecules containing the 5' 68, 48, and 37 nucleotides (Fig. 4) were generated by transcribing synthetic oligonucleotides. Comparing the \( K_d \) values of these truncates for the nucleocapsid protein to that of the full-length RNA revealed that deletions of nine (BC68) and 29 (BC48) nucleotides from the 3' end resulted in slight, though significant decreases in affinity (Fig. 4). Removal of an additional 11 nucleotides (T-23; minimum binding domain) caused binding to drop drastically. Although T-23 binds to the nucleocapsid protein tighter than the parent pool (30N RNA), its dissociation constant is surprisingly greater than 400-23 (Fig. 4). All truncated RNA fragments were sequenced directly to confirm their size and sequence (data not shown). The fragment containing the tetra-loop (BC2) bound with an affinity similar to that of T-23.

Nucleocapsid footprints on the RNA

The physical binding site of the nucleocapsid protein on one of the SELEX-generated RNAs (400-23) was determined by nuclease footprinting. Four regions of the RNA are shielded from digestion by S1 nuclease by varying concentrations of the nucleocapsid protein. At NC concentrations as low as 50 nM, sites in the conserved AACU bulge are protected from cleavage while remote nuclease-sensitive sites require up to 10 \( \mu \)M nucleocapsid to display protection (Fig. 5). Nucleocapsid footprints that are observed at high concentrations are seen both 5' and 3' of the bulge. If protein–protein interactions are important for binding to the low affinity sites, then the highly conserved four base bulge may serve as the nucleation site for cooperative binding. Our results are consistent with the estimated occlusion size of seven to eight nucleotides for HIV nucleocapsid (L. Henderson, personal communication). The apparent binding of multiple proteins to each RNA might explain the decreased affinities observed for the RNAs when they are truncated. The binding of the longer RNAs could be enhanced either through cooperative protein binding or by an increase in

Fig. 4. Nitrocellulose filter binding curves for mature 400-23 RNA, truncated 400-23 RNAs and 30N1 RNA. open circle, mature 400-23 RNA; closed square, 400-23 minus 9 nucleotides from the 3' end (BC68); open squares, 400-23 minus 29 nucleotides from the 3' end (BC48); diamonds, 400-23 minus 40 nucleotides from the 3' end (T-23); filled circles, 30N1 RNA. Nucleocapsid concentrations are in molar. Binding reactions were carried out in binding buffer containing 100 mM NaCl.
the filter retention efficiency brought about by multiple proteins binding to a single RNA.

In vitro inhibition of annealing function

If the RNA ligands are binding a site on NCp7 that is used for substrate recognition, then the ligands should act as inhibitors. An assay for one of the proposed activities of nucleocapsid, stimulation of nucleic acid annealing, was developed to test the inhibition efficiency of the ligands. Complementary nucleic acids, con+ and con−, are converted to the con+/− double-stranded form in 6 min when nucleocapsid is present in the reaction at 100 nM or greater (Fig. 6). Nucleocapsid concentrations of 10 nM result in about a 60% conversion while 1 nM nucleocapsid results in a mild increase in double-stranded DNA formation above the minus-nucleocapsid background level.

Annealing assays, as described above, were carried out in the presence of 30N, 400-17, and 400-23 RNAs to determine the extent to which they inhibited the function of the nucleocapsid protein. All three RNA samples were able to inhibit HIV-1 nucleocapsid protein-dependent annealing, but the two selected RNAs inhibited at concentrations which were orders of magnitude lower than that of the parent pool of RNA (Fig. 7). It took concentrations of 30N RNA of 722 nM or greater to inhibit the annealing activity of nucleocapsid in binding buffer containing 100 mM NaCl. For clones 400-17 and 400-23, the extent of inhibition was much greater. 400-17 was able to inhibit approximately 50% of the annealing activity of nucleocapsid at a concentration of 0.5 nM, while the Kᵣ for 400-23 was 3 nM (Fig. 7). The observation of Kᵣ’s for these ligands that are lower than their observed Kᵢ’s might be explained by the apparent cooperative binding observed in the footprinting experiments above, as a single ligand could bind and inhibit multiple proteins.

DISCUSSION

As a first step toward understanding the RNA substrate binding specificity of the HIV-1 nucleocapsid protein p7, we generated RNA ligands to the viral protein using the SELEX protocol. All of the ligands that resulted from the selection contained a highly conserved 14-nucleotide block. Computer modeling and structure probing experiments are consistent with the RNAs adopting a secondary structure comprising two 5-bp stems with a 4-base internal bulge of the consensus sequence 5’-AACU-3’.

The invariance of the structural motif among the selected ligands suggests that it is the primary recognition site for the HIV-1 nucleocapsid protein p7.
for the nucleocapsid protein. This idea is corroborated by boundary and protein footprinting data. The fact that all of the ligands utilized the 5' most nucleotides in the RNA molecules does not necessarily reflect a need for the structure to occur at the 5' terminus of an RNA in order to generate high affinity binding. Molecules using the 5'-fixed region to form one-half of the hairpin are expected to be more than a thousand-fold more prevalent in the initial RNA population than those molecules that form the structure using only nucleotides within the random region.

Two regions of the HIV-1 genome that are thought to interact with the nucleocapsid protein, the primer binding site (PBS) where tRNA^Lys^ anneals prior to cDNA synthesis and a domain implicated in encapsidation (Aldovini and Young, 1990; Clavel and Orenstein, 1990), were searched for potential nucleocapsid binding sites. AACU sequences or minor variations of this sequence were sought out within or near these two areas, and the surrounding sequences were checked for their capacity to base-pair to form helices on either side of the AACU. The search identified several primary sequences that could form structures that were very similar to the SELEX-generated consensus (Fig. 8). We have no physical evidence for the existence of these structures in the PBS or encapsidation region. However, it is tempting to spec-

![Image](https://example.com/image.png)

**FIG. 7.** Autoradiograph showing in vitro inhibition of HIV-1 nucleocapsid annealing function. Lanes labeled Con+/Con− have both complementary strands in the absence of nucleocapsid; Con+/Con−/NC, contains nucleocapsid but no RNA. All lanes labeled with a RNA concentration also contains con+, con−, and nucleocapsid (see methods). 7A, 30N1 RNA; 7B, 400-17; 7C, 400-23. RNA concentrations are given in nanomolar.

![Image](https://example.com/image.png)

**FIG. 8.** Three potential secondary structures from the HIV-1 viral genome that are similar to the motif that was identified as a high affinity site for the nucleocapsid protein. The numbers indicate the nucleotide position relative to the transcriptional start site of the HIV-1 genome. The primer binding site spans the RNA from positions 107–125; the encapsidation domain is suspected to occur between nucleotides 223 and 337 (nucleotide positions taken from Ratner et al., 1985). Structures A and B flank the primer binding site, while structure C is located in the putative encapsidation region.
ulate on a model where nucleocapsid protein binds these elements during the life cycle of the virus and exert its effects on the various NC-mediated events. In our model, nucleocapsid protein binds these structures and localize the involved components to these regions of the genomic RNA. Since our SELEX-generated ligands are inhibitors of the annealing activity of the NC in vitro, these RNAs could prove to be useful therapeutic agents. It would be interesting to learn about the effects of these ligands on other NC-mediated activities.

ACKNOWLEDGMENTS

We thank members of the Gold lab and NeXstar Pharmaceuticals, Inc. for many valuable discussions. We are grateful to Zuzana Hostom-kska for her supervision of the NC preparation. We thank Bryn Weiser and the Noller lab for secondary structure pictures. G. Hudson, E. Nguyen, J. Dutra, M. Liss, M. Sotelo, and T. Weikl for technical assistance, and A. K. Sotelo for careful reading of the manuscript. This work was supported in part by a fellowship to P.A. from the Jane Coffin Childs Memorial Fund, NeXstar Pharmaceuticals, Inc., W. M. Keck Foundation, and NIH Grants GM28685, GM19963, Al-39186, and AI-33380.

REFERENCES


RNA RECOGNITION BY HIV-1 NUCLEOCAPSID