

High-throughput sequence analysis reveals structural diversity and improved potency among RNA inhibitors of HIV reverse transcriptase

Mark A. Ditzler¹, Margaret J. Lange¹, Debojit Bose^{1,2}, Christopher A. Bottoms³, Katherine F. Virkler¹, Andrew W. Sawyer⁴, Angela S. Whatley¹, William Spollen³, Scott A. Givan^{1,3} and Donald H. Burke^{1,4,*}

¹Molecular Microbiology and Immunology, University of Missouri, Columbia, MO, USA,

²Proteomics and Structural Biology Unit, Institute of Genomics and Integrative Biology, CSIR, Delhi, India,

³Informatics Research Core Facility and ⁴Department of Biochemistry, University of Missouri, Columbia, MO, USA

Received August 2, 2012; Accepted October 29, 2012

ABSTRACT

Systematic evolution of ligands through exponential enrichment (SELEX) is a well-established method for generating nucleic acid populations that are enriched for specified functions. High-throughput sequencing (HTS) enhances the power of comparative sequence analysis to reveal details of how RNAs within these populations recognize their targets. We used HTS analysis to evaluate RNA populations selected to bind type I human immunodeficiency virus reverse transcriptase (RT). The populations are enriched in RNAs of independent lineages that converge on shared motifs and in clusters of RNAs with nearly identical sequences that share common ancestry. Both of these features informed inferences of the secondary structures of enriched RNAs, their minimal structural requirements and their stabilities in RT-aptamer complexes. Monitoring population dynamics in response to increasing selection pressure revealed RNA inhibitors of RT that are more potent than the previously identified pseudoknots. Improved potency was observed for inhibition of both purified RT in enzymatic assays and viral replication in cell-based assays. Structural and functional details of converged motifs that are obscured by simple consensus descriptions are also revealed by the HTS analysis. The approach presented here can readily be generalized for the efficient and systematic post-SELEX development of aptamers for down-stream applications.

INTRODUCTION

The indispensable role of reverse transcriptase (RT) in the replication of type I human immunodeficiency virus (HIV-1) and the appearance of drug-resistant virus motivate efforts to identify new RT inhibitors. Systematic evolution of ligands through exponential enrichment (SELEX) experiments have identified RNA and DNA aptamers that bind multiple epitopes of RT with a range of affinities and specificities in assays with purified enzyme (1–16). Among those aptamers, pseudoknot RNAs have received considerable attention and have been shown to bind HIV-1 RT with nanomolar affinity (1,2,16), to block all the major enzymatic steps catalyzed by RT (8) and to inhibit HIV replication in cell culture (16–18). Thus, aptamers that bind RT are promising reagents for use as therapeutics, diagnostics and research tools.

SELEX identifies aptamers by subjecting a diverse nucleic acid library to iterative cycles of enrichment and amplification. After several cycles, the population is typically evaluated by sequencing a few dozen members of the enriched population. Occasionally, a single sequence dominates the selected population, but more often there is only limited repeat sampling of individual sequences, indicating extensive untapped diversity (1–7,19–22). The abundance of sequences in the final populations derived from SELEX experiments reflects both the composition of the starting pool and all aspects of the intervening selection process, such as affinity for the molecular-binding target, non-specific nucleic acid binding to the target, ionic strength, temperature, the efficiency of the partitioning method(s) used and biases imposed by transcription, reverse transcription and polymerase chain reaction (PCR) amplification. The interplay of these multi-faceted

*To whom correspondence should be addressed. Tel: +1 573 884 1316; Fax: +1 573 884 9676; Email: burkedh@missouri.edu
Present address:

Mark A. Ditzler, Space Science Division, NASA Ames Research Center, Moffett Field, California.

selection pressures and the composition of the initial pool make interpreting SELEX output in terms of the desired phenotype non-trivial. Extensive validation and optimization of aptamers are often required (23).

High-throughput sequencing (HTS) provides an opportunity to describe SELEX populations more completely, leading to deeper structural insight, the identification of rare motifs and the ability to make functional inferences based on detailed comparisons between closely related populations. Initial efforts to evaluate differences between related aptamer populations using HTS have yielded promising results (24–27). In those efforts, it was the capacity of HTS to identify changes in sequence abundance between successive rounds of SELEX or between the starting pool and the single round of selection, which identified sequences too rare to be identified through cloning-based, low-throughput sequencing (LTS). Here, we applied additional levels of analysis that have not been exploited in prior HTS analysis of aptamer populations (24–27), including structural inferences based on the divergent cloud of mutations within individual lineages, identifying convergence on specific structural motifs and evaluation of the relative fitness among and within these structural motifs.

Transforming the large amount of HTS data into digestible descriptions of SELEX populations requires the development of specific computational tools, some examples of which have already been developed for evaluating changes in sequence abundance (24,25,28). In this work, we integrate multiple comparative sequence analysis tools (29–32) into an HTS analysis pipeline that identifies the functional structures present in multiple SELEX populations and reveals details about the structural motifs generally derived from follow-up post-selection analysis. The resulting high-resolution view of multiple SELEX populations illuminates the nature of nucleic acid recognition and streamlines the process of post-SELEX validation, prioritization and optimization, demonstrating the capacity of HTS analysis to accelerate this rate-limiting step in the development of aptamers for therapeutic, diagnostic and research purposes in general. In addition, for the RT-binding populations analyzed here, this high-resolution view identified the '(6/5)AL' motif, which inhibits RT enzymatic activity and HIV replication more potently than previously identified RNA pseudoknots (1,2,16–18), and simultaneously provided a detailed description of the sequence requirements of this new structural element.

MATERIALS AND METHODS

RT expression and purification, *in vitro* transcription and plasmids

RT from HIV-1 strain HXB2 was expressed in *Escherichia coli* strain BL21(DE3) and purified as previously described (15). Individual aptamers were transcribed *in vitro* from either fully or partially double-stranded DNA templates with a T7 promoter by incubation in buffer (50 mM Tris-HCl pH 7.5, 10 mM NaCl, 30 mM MgCl₂, 2 mM spermidine, 40 mM DTT) with 4 mM of each NTP and T7 RNA

polymerase for 12–18 h at 37°C. Transcripts were then purified using denaturing polyacrylamide gel electrophoresis (PAGE). All full-length aptamers share the same 5' and 3' constant sequences for amplification that flank the ~70 nt random region; the 5' sequence is 5'-GGGAAAAGCGAAUCAUACACAAGA and the 3' sequence is GGGCAUAAGGUAAUUAAUCCAUA-3'. For the five representative full-length sequences identified in this work, the 70 nt sequences between these two constant regions are as follows: 240.1.1:ACGUUGUCGAAAGC CUAUGCAAUUAAGGACUGUCGACGAAACCU UGCGUAGACUCGCCACGCUUGGUGU, 240.3.1: AAGUCGUGUGAUUGAAUUGUCCCGACAGGUAU CCGAAAGGGCACGGGAUAAGCUCUGUAUGGA GUCGAAU, 240.5.1:CGAGACAAGUACCGAAAAAG AGAUCUGGCAGUGUCACAACCAGGAAAAAGA CACGACGAACACGCCGCAC, 240.88.1:CAUUGAC CAAAAGGGUCGAUAGCGUCGUGUAGAUUGCA CCCAUGACUGAGCUACUGCCAAAUCCACCCAC and 240.94.1:UUCGUAGAAUUGGGGAACAAAGAA AUCACGCCAAGGCCCAUAGGGCAAUUGCUCC AUGGGUGGAACCAC. DNA oligonucleotides were synthesized by Integrated DNA Technologies, Inc (Coralville, IA). The aptamer-expressing plasmid used in viral inhibition assays is driven by a human cytomegalovirus (CMV) immediate early promoter and contains the full-length aptamer flanked by two catalytically inactivated hammerhead ribozymes (18). The HIV-1_{NL4-3}-derived CMV-enhanced green fluorescent protein (EGFP) plasmid (pNL4-3-CMV-EGFP) used in single-cycle infectivity assays was kindly provided by Vineet KewalRamani (National Cancer Institute, Frederick, MD). This proviral vector lacks the genes encoding *vif*, *vpr*, *vpu*, *nef* and has a CMV immediate early promoter-driven EGFP in place of *nef*. The vesicular stomatitis virus glycoprotein-expressing plasmid (pMD-G) used for viral pseudotyping was obtained from Invitrogen (Carlsbad, CA).

Competitive SELEX

The additional round of competitive SELEX was performed by PCR amplifying the dsDNA product from the 14th round of SELEX followed by transcription. Approximately 2 pmol of RNA pool spiked with trace [³²P]-5' end-labeled RNA was pre-incubated in 45 μl binding buffer (150 mM KCl, 10 mM MgCl₂ and 50 mM Tris-HCl pH 7.5) with a 1 cm × 3 cm strip of nitrocellulose and vortexed briefly, then a 3.3 μl aliquot of the sample was passed through a nitrocellulose filter under vacuum (NoRT). Nitrocellulose filters were pre-wet with buffer and washed with 1 ml buffer immediately before and after adding sample to the filter. To the remaining sample, 2 pmol of RT was added along with 250 ng of bovine serum albumin. The RNA-protein mixture was incubated at 23°C for 10 min, after which a 3.3 μl aliquot was passed through a nitrocellulose filter. Then the sample was diluted into a total volume of 200 μl of the same buffer containing the DNA aptamer R1T at a final concentration of 2 μM. Twenty microliter aliquots of diluted sample were passed through nitrocellulose filters

15, 60 and 240 s following dilution. The bound RNA was recovered, reverse transcribed and then PCR amplified for sequencing (Supplementary Text).

High-throughput sequencing and analysis pipeline

HTS data were obtained for six SELEX populations using a HiSeq2000 instrument (Illumina). Flanking sequences required for Illumina sequencing (including indexing sequences to allow the multiplexing of multiple populations in a single-sequencing run) were appended during two sequential PCR amplification steps using Pfu DNA Polymerase (Supplementary Text). Indexing was used to sequence multiple populations within a single lane, resulting in very low sequencing costs for this work; therefore, cost does not represent a barrier to HTS analysis at the level used in this work. Final amplified products from the six populations were pooled, loaded onto a single lane, bridge amplified and then run through 100 sequencing cycles. Illumina's analysis software was used to generate fastq files with sequence calls and associated quality scores for >3 million raw sequence reads per population. Sequences were quality filtered down to ~1 million reads per population and then aligned, clustered and used to identify converged structures (Supplementary Text). Scripts associated with our analysis have been deposited in CPAN (Bio::App::SELEX::RNAmotifAnalysis). Consensus secondary structures in Figure 3 were modified from depictions generated using R2R (32).

Aptamer nomenclature

We applied a nomenclature based on clustering, with sequences identified by the population to which they belong, the rank order of the cluster to which they belong and their rank order within the cluster, for example a sequence that is from the pre-enriched population (pre) that belongs to the most abundant cluster and is the second most abundant sequence within that cluster is named 'pre.1.2'.

DNA-dependent DNA polymerization inhibition assays

Primer extension reactions utilized a [³²P]-5' end-labeled 18 nt DNA primer and a 40 nt DNA template corresponding to the viral tRNA primer-binding sequence and U5 segments of HIV-1 strain HXB2. Aptamer-RT complexes were pre-assembled in extension buffer (50 mM TRIS-HCl pH 7.5, 50 mM NaCl, 5 mM MgCl₂) with 100 nM aptamer and 20 nM RT for 5 min. Polymerization reactions were then initiated by adding pre-annealed 10 nM primer and 20 nM template and 100 μM each dNTP (final concentrations). Reactions were allowed to proceed for 20 min at 37°C. The reaction conditions were selected so that strong and persistent inhibition would be clearly differentiated from modest inhibition and so that enzymatic activity would depend on dissociation of the pre-assembled aptamer-RT complex. After 20 min, reactions were quenched by addition of 2 volumes (20 μl) of gel-loading buffer (95% formamide with 0.01% bromophenol blue), heated to 95°C for >3 min and analyzed by high-voltage denaturing PAGE. Electrophoretic migration profiles were generated using

Multigauge software (Fujifilm) and then normalized and plotted in Igor (WaveMetrics).

Single-cycle infectivity assays

Cell culture, virus production and evaluation of viral infectivity were carried out as previously described (18) (Supplementary Text). Single-cycle infectivity assays using aptamer-expressing plasmids, pNL4-3-CMV-GFP and pMD-G, were performed by transfecting 293FT cells with polyethylenimine (PEI) in six-well cell culture dishes. All transfections were performed on cells plated the previous day (~50% confluence) using PEI at 3 μl per microgram of DNA, as previously described (18). Aptamer-expressing plasmids were transfected first (1 μg), followed 4 h later by a mixture of pNL4-3-CMV-GFP (250 ng) and pMD-G (125 ng). The medium was changed between transfections and again 4 h after the second transfection. After collecting virus, transfected cells were trypsinized, collected, fixed in 4% paraformaldehyde, washed with 1X phosphate-buffered saline and analyzed by flow cytometry to determine transfection efficiency by EGFP expression using an Accuri Flow Cytometer (BD Biosciences, San Jose, CA). Supernatant was harvested 48 h post-transfection and syringe filtered through 0.45-μm filters. p24 ELISA was used to quantify virus produced and verify that there was no detectable difference in viral production among samples. Viral supernatant (50 or 100 μl) was added to fresh 293FT cells to determine infectivity. Infected cells were collected 48 h post-infection, fixed with 4% paraformaldehyde and analyzed on an Accuri Flow Cytometer to determine the percentage of infected (EGFP-positive) cells.

RESULTS

Identifying and characterizing individual evolutionary lineages within the SELEX populations

To efficiently extract a structure-based understanding of SELEX populations from HTS data, we developed an analysis pipeline that clusters, aligns and compares sequences based on patterns of conservation and covariance across multiple scales (Figure 1). We analyzed 5.6 million quality-filtered sequences from six populations of aptamers selected to bind HIV-1 RT. One of the populations, referred to here as 'pre-enriched', was previously selected by 11 rounds of capturing the bound complexes on nitrocellulose filters, followed by 3 rounds of capture by native gel electrophoretic mobility shift (2). This population was subjected to LTS and extensive biochemical follow-up of numerous pseudoknot inhibitors of RT (2,8,11,16,18), but repeat sampling of individual sequences was sparse among the 48 published, implying significant untapped diversity. The five additional aptamer populations analyzed here were generated in this work by subjecting the pre-enriched population to one additional round of SELEX on nitrocellulose filters at high ionic strength in the presence of competitor at varying stringency (see below).

Our analysis pipeline begins by grouping sequences of shared evolutionary lineage by identifying clusters of

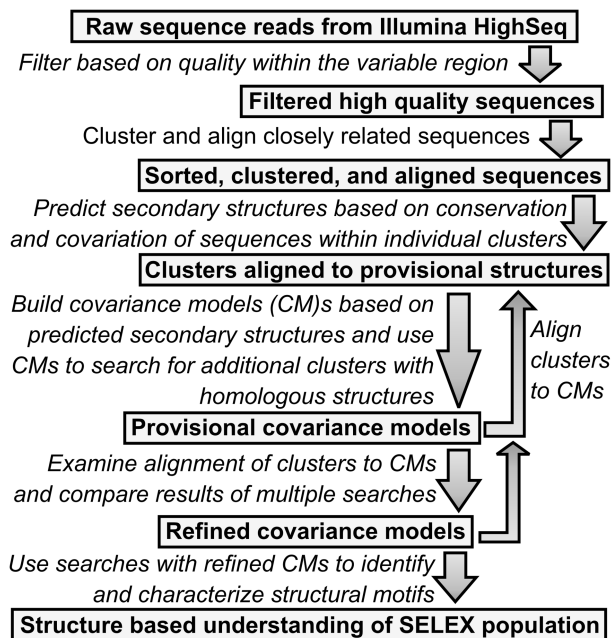


Figure 1. Schematic of the HTS analysis pipeline used in this study.

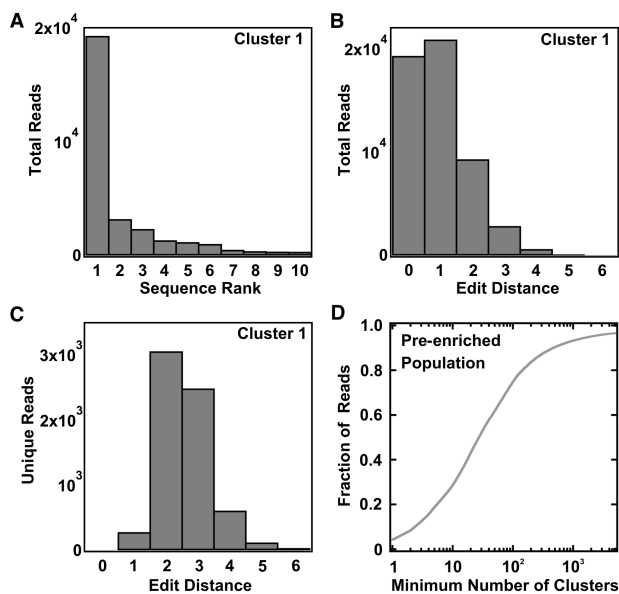


Figure 2. Population structure of pre-enriched population. (A) Distribution of reads among the 10 most abundant sequences (pre.1.1 through pre.1.10) within the most abundant cluster (pre.1) in the pre-enriched population. (B) The number of reads observed as a function of edit distance from the most abundant sequence in the pre-enriched population. (C) The number of unique sequences observed as a function of edit distance from the most abundant sequence in the pre-enriched population. The distributions in panels (A–C) are qualitatively similar to those observed for the majority of clusters within the populations described here. (D) The maximum fraction of the pre-enriched population that can be described by a given number of clusters.

sequences within a small edit distance of the most abundant sequence within the cluster. The edit distance is defined as the minimum number of substitutions, insertions and deletions required to convert one sequence into

the other. Most clusters are dominated by a single sequence (Figure 2A), and the majority of the remaining reads belong to single point mutants of the dominant sequences, with the number of reads falling to zero with increasing edit distance (Figure 2B). On the basis of the rapid fall-off in the number of mutants observed as edit distance increases, we defined a cluster as the set of sequences that are fewer than seven edits from the dominant sequence in that cluster. Sequences within a given cluster are very likely related by descent given the sparse sampling of the $>10^{42}$ potential sequences (70 fully randomized positions = 4^{70} possible combinations) in the original starting library of $\sim 10^{14}$ unique sequences. Most of the intra-cluster diversity is found among the double and triple mutants (Figure 2C). This diversity represents mutations accumulated through multiple rounds of amplification and acted on by selection, in addition to a contribution from randomly distributed variation due to sequencing errors and mutations that occurred during the final cycle of selection. For our analysis, we evaluated the 5000 most abundant clusters in each of the populations, encompassing $>96\%$ of all quality-filtered sequence reads within the populations (Figure 2D).

Covariance analysis of HTS data identifies structural motifs within the SELEX populations

Variation that evolved within clusters (divergence) and shared features between clusters (convergence) were both exploited to infer functional structural motifs. Sequences within clusters were aligned using *Mafft* (29), and then *RNAalifold* (30) was used to generate provisional secondary structure predictions based on conservation and covariation within the ~ 70 nucleotide variable region of individual clusters. The software *Infernal* (31) was used to generate initial covariance models (CMs) based on the provisional secondary structures and alignments. *Infernal* was then used to search all 5000 clusters to identify clusters that share significant structural and sequence features with those CMs. Those clusters were then aligned to the original CMs and used to generate new, refined CMs, which were again searched against the entire population of clusters. The process of searching and refinement was repeated for 10 rounds, at which point the number of clusters identified in successive rounds had plateaued (Supplementary Figure S1 and Supplementary Text). Searches with our final refined CMs revealed convergence on shared structural motifs, with large groups of clusters satisfying individual CMs.

Our analysis identified five distinct CM-based definitions of structural motifs within the populations—three classes of pseudoknots and two variations on a newly identified structural element—that account for approximately 97% of the 1.2×10^6 sequences within the top 5000 clusters in the pre-enriched HTS data set. Although pseudoknots are not explicitly predicted by the comparative sequence analysis tools used here, their presence is revealed by stem-specific CMs based on each of the two component stems that make up the pseudoknots (Supplementary Figure S1). The Family 1 Pseudoknot (F1Pk) motif identified here is essentially identical to the

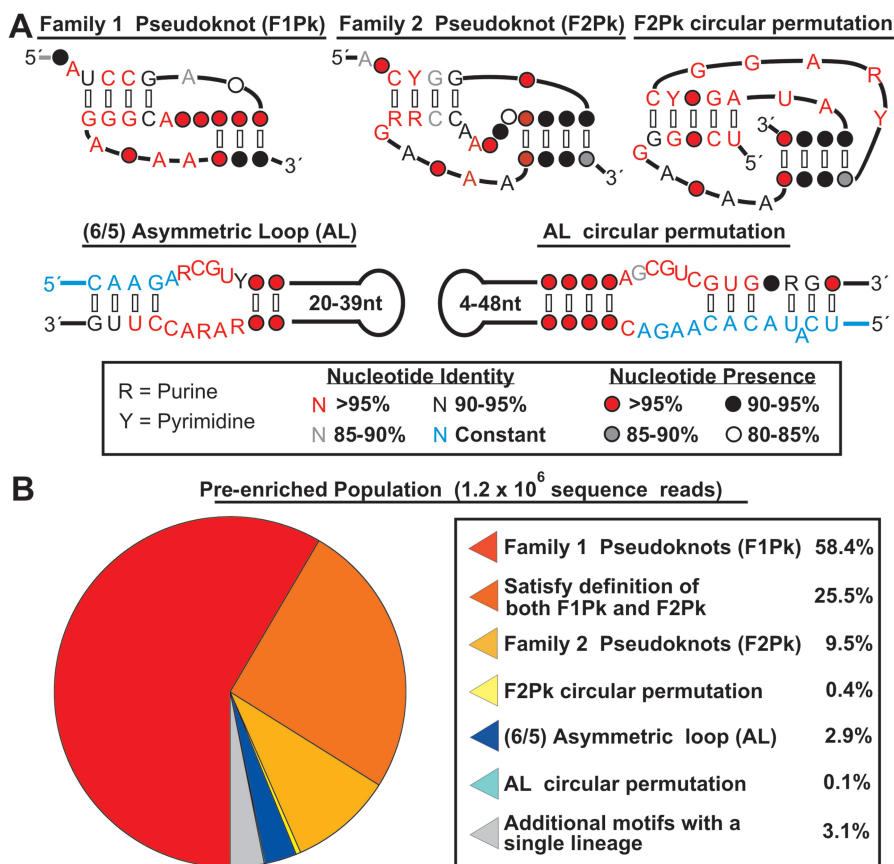


Figure 3. Consensus descriptions and distribution of the converged motifs identified through covariance analysis. (A) The consensus features of five converged motifs are shown. Sequence conservation within the context of the specified structures is color-coded as indicated in the figure. Three of the conserved motifs are structurally similar pseudoknots; note the shared ANAA loop feature. The remaining two motifs are circular permutations of a novel asymmetric loop structure. (B) The abundances of the five motifs within the pre-enriched population are shown. The overlapping fraction of sequences that simultaneously satisfy the closely related F1Pk and F2Pk motif definitions is indicated in orange.

dominant pseudoknot motif identified previously by LTS analysis (1,2). Its recurrence here served as a convenient control for validating our analysis. A second converged pseudoknot motif (F2Pk) was also identified by the CM searches, along with a circular permutation of F2Pk (F2Pkcp). An additional conserved base pair is present in one of the stems of F2Pkcp, potentially compensating for a loss of stabilization due to the altered connectivity of the stems in the permuted form (Figure 3A). Although previous LTS analysis noted the presence of potential pseudoknots that did not conform to the F1Pk consensus (2), the ‘Family 2’ classification in that work was based on an exclusive definition (i.e. ‘not F1Pk’), in contrast to the new inclusive definition described here (i.e. ‘satisfies a specific CM’). The F1Pk and F2Pk motifs are structurally very similar, and many clusters simultaneously satisfy the definitions of both motifs (Figure 3 and Supplementary Figure S2). Collectively, the pseudoknot motifs represent 93.8% of the pre-enriched population (Figure 3B). In addition to the pseudoknots, a previously unidentified asymmetric internal loop structure ((6/5)AL) that represents only 2.9% of the pre-enriched population was revealed by this analysis. A circular permutation of this structural element, the (6/5)ALcp motif, comprises 0.2%

of the population and provides additional evidence for the functional significance of the predicted internal loop. Additional clusters that do not show clear evidence of converging on shared structures collectively comprise the remaining 3.1% of the pre-enriched population (Figure 3B), with the most abundant cluster among them representing 0.4% of the population.

Population dynamics under increasingly stringent selection pressure reveals relative potencies of aptamer inhibitors of RT

The relative fitnesses of sequences within the pre-enriched population were probed by carrying out an additional round of SELEX under conditions of increasing stringency. Before incubation with RT, a subtractive step was included to limit the prevalence of aptamers with affinity for nitrocellulose. The RNA population was then equilibrated with HIV-1 RT in a high ionic strength buffer (150 mM KCl, 10 mM MgCl₂) to limit non-specific binding to the highly basic RT, and a fraction of the equilibrated sample was partitioned on a nitrocellulose filter to establish a pre-competition baseline of sequences bound to RT (the 0s population). A large molar excess of the high-affinity DNA aptamer R1T (14)

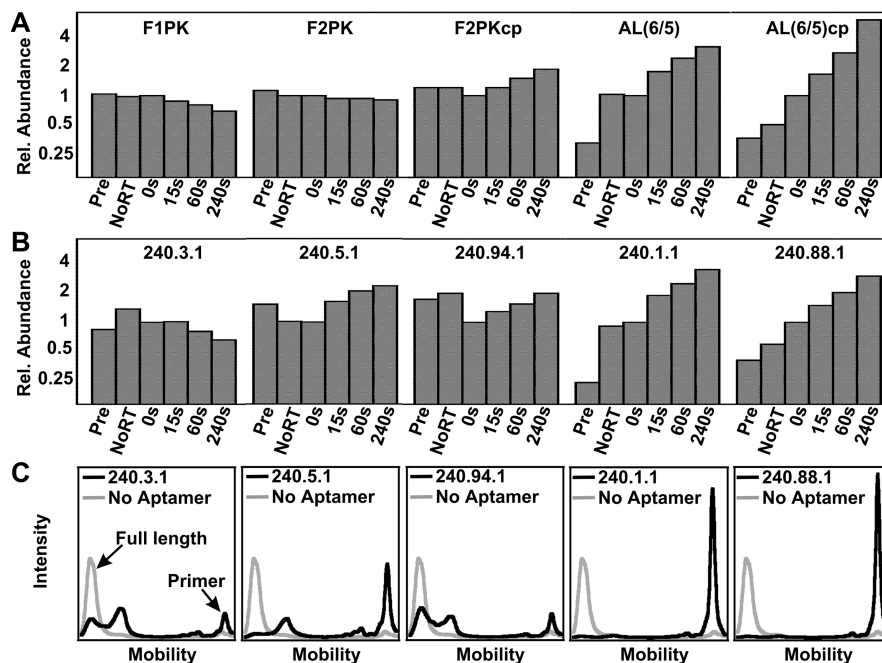


Figure 4. Relative abundances of motifs and specific sequences within different populations. (A) The fractional abundances of the five converged motifs in all six populations are shown relative to their abundances within the population of aptamers bound to RT immediately before the addition of competitor (0s). (B) The relative fractional abundances of the most abundant sequences from each of the five converged motifs are shown. (C) Inhibition of primer extension by RT in the presence of the same five representative sequences shown in (B). Electrophoretic mobility profiles of ^{32}P -labeled primer and extended products are shown. RT-catalyzed primer extension results in slower mobility. The relative mobilities of the unextended and fully extended primer are indicated in the first panel, and the profile from the uninhibited reaction is shown in each panel as a reference.

was then added to the remaining sample to block the substrate binding cleft (15) and compete for RT binding with any RNA aptamers that dissociate from RT. Aptamers that remained bound were captured by partitioning the sample on nitrocellulose filters at 15, 60 and 240 s following addition of the DNA trap. Addition of the trap is expected to increase the fraction of the population represented by aptamers that dissociate slowly. In addition, the resulting reduction in overall protein-dependent retention on the nitrocellulose filter (Supplementary Figure S3) is also expected to favor those sequences that exhibit the greatest propensity for non-specific retention on nitrocellulose. To explicitly account for the differential RT-independent retention of the aptamers on the nitrocellulose filter, a fraction of the sample was partitioned on a nitrocellulose filter before the addition of any protein (NoRT) and then recovered and sequenced.

The F1Pk and F2Pk pseudoknot motifs that dominate the pre-enriched population decreased as a fraction of the population following the addition of trap, while the (6/5)AL and (6/5)ALcp motifs increased (Figure 4A). The most abundant representatives of these motifs (only one of which was present in the LTS data set) follow the same trend as their corresponding motifs with the exception of the F2Pk representative (Figure 4B). Due to the high number of reads afforded by HTS, these observed changes in the number of sequence reads exceed the variation expected due to simple resampling and therefore likely reflect functional differences between the aptamers

within the population (Supplementary Figure S4). To evaluate the functional significance of these changes in abundance, we tested the ability of these specific representative to inhibit RT. Consistent with their responses to trap, representatives of the two topologies of the (6/5)AL motif exhibited the strongest inhibition of DNA-dependent DNA polymerization (DDDP) by RT (Figure 4C). The F2Pkcp motif's most abundant representative also increased in relative abundance following the addition of trap, in spite of the observation that the F2Pkcp representative exhibited only limited RT inhibition in DDDP assays (Figure 4C). This sequence's relatively high abundance within the NoRT sample reveals that this increase reflects a higher level of protein-independent background binding to the filter and not a higher affinity for RT.

Identifying the functional core of the (6/5)AL motif

Both the (6/5)AL and (6/5)ALcp motifs utilize a portion of the 5' constant region to form the structures identified by the CM searches. Stem 2 of the (6/5)AL motif is entirely contained within the variable region, whereas stem 1 usually includes pairing with the constant region, and this pattern is reversed for the (6/5)ALcp motif (Figure 3A). The HTS data reveal that stem 2 of (6/5)AL can be as short as 24 nt with as few as 8 bp and that stem 1 of (6/5)ALcp can be as short as 12 nt with as few as 3 bp. To determine whether these patterns reflect physical constraints, DDDP inhibition assays were carried

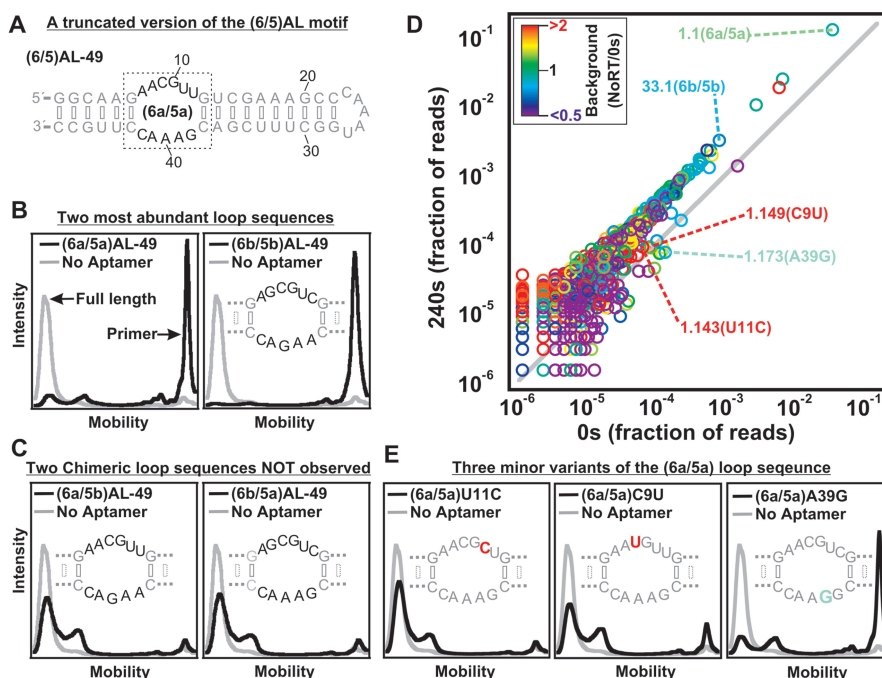


Figure 5. Experimental validation of HTS inferences for the (6/5)AL motif. (A) The 49nt truncated (6/5)AL aptamer based on the analysis of sequences within the two circular permutations of the motif. The most common (6a/5a) loop combination is shown in bold (B) Inhibition of primer extension by RT in the presence of truncated aptamers containing the (6a/5a) (left) and (6b/5b) (right) loop sequences. Profiles of the electrophoretic mobility of ³²P-labeled primer are shown as in Figure 4D. The (6b/5b) loop combination is shown as inset. (C) Inhibition of primer extension by RT in the presence of truncated aptamers containing loop sequences that are conspicuously absent from the population. (D) Correlation plot of the abundances of specific (6/5)AL aptamer sequences between populations. The fractional abundances of sequences within the 0s population (X-axis, 886 756 total reads) are plotted against the fractional abundances of those sequences within the 240s population (Y-axis, 711 100 total reads). Data points are color coded according to the ratio of their fractional abundances in the 0s population relative to their fractional abundances in the (NoRT) population (high background binding is red and low is violet). (E) Inhibition of primer extension by RT in the presence of truncated point mutants of the most abundant (6/5)AL aptamer.

out using a series of increasingly aggressive truncations of the (6/5)AL sequence 240.1.1 (see Materials and Methods for sequence nomenclature). Truncations in which stem 1 was shortened to 6 bp retained strong RT inhibition, as did truncations in which stem 2 was truncated to 10 bp, but truncating stem 2 to only 6 bp resulted in the loss of RT inhibition by the aptamer (Figure 5A and Supplementary Figure S5). These results verify that the (6/5)AL motif represents the functionally relevant structure within the full-length aptamers.

HTS provides the sequencing depth necessary to define details of motifs that extend beyond a simple consensus

A simple consensus description of the (6/5)AL motif includes internal loops of the sequence ARCGUY/RARAC. Although 16 sequence combinations fit this description, only a small subset of those 16 are observed (Table 1). The most frequently observed loop is designated (6a/5a)AL, in which the 6nt element ‘6a’ is AACGUU and the 5nt element ‘5a’ is GAAAC. The second most abundant loop (6b/5b)AL uses the combination AGCGUC and AAGAC to form the loop, and this is the dominant combination in the (6/5)ALcp. Together these two combinations account for 98.6% of all (6/5)AL reads in the 240s population. The 6a element is observed in 50 of the 129 (6/5)AL clusters and the 5b element in 65 of

Table 1. Relative abundance of (6/5)AL variants

	%AL 240s	aACGUU “6a”	aGCGUC “6b”	aACGUC “6c”	aACGUG “6d”	aACGUA “6e”
%AL 0s						
%AL NoRU						
GAAAC “5a”	95.9 95.5 97.1		0	0	0.076 0.064 0.074	0
AAGAC “5b”	0	2.73 2.47 1.52	0.082 0.104 0.075	0	0	0
AAAAC “5c”	0.017 0.043 0.071	0	0.804 1.28 0.918	0	0	0
UAAAC “5d”	0.37 0.56 0.20	0	0	0	0	0
GACAC “5e”	0	0	0	0.006 0.019 0		
CAAAC “5f”	0	0	0	0	0	0.008 0.011 0.014

Combinations of 6 and 5 nt sequences observed in the (6/5)AL motif. The top row lists the sequences observed on the 6 nt side of the loop and the first column lists the sequences observed on the 5 nt side of the loop. The percentage of (6/5)AL reads belonging to clusters that have a given 6 and 5 nt combination are indicated in the table. The percentages reported are from the 240s (top value), 0s (middle value) and NoRT (bottom value) populations. Only those combinations that appear in the dominant sequence of clusters that are represented by >10 reads are shown.

the clusters. Therefore, the absence of any cluster in which 6a is found in combination with 5b is meaningful. In fact, both of the chimeras (6a/5b)AL and (6b/5a)AL are conspicuously absent from the sequenced populations (Table 1). To determine the biochemical significance of this observation, we generated a 49 nt versions of the (6/5)AL with these four loops inserted modularly between two perfectly base-paired stems. The (6a/5a) and (6b/5b) combinations potentially inhibited RT (Figure 5B), whereas the chimeric (6a/5b) and (6b/5a) combinations exhibited dramatically less inhibition (Figure 5C) consistent with their absence from the sequencing data. Thus, HTS analysis provides a means of inferring details of the functional requirements of the motifs present in a complex aptamer population that extend beyond identifying simple consensus features.

Intracluster dynamics across multiple populations represents a level of resolution that is essentially invisible to LTS data sets and that has been strikingly underexploited for HTS data sets. For example, within cluster 240.1, specific sequences are less enriched than the majority of sequences within that cluster in response to the addition of trap. These sequences display variation either within or immediately adjacent to the conserved loop, indicating that these variations are at a selective disadvantage. In contrast, sequences that exhibit average levels of enrichment generally vary from the dominant sequence far outside the conserved loop (Figure 5D and E and Supplementary Figure S6), thus reinforcing the functional significance of this loop.

Critical insight into aptamer function from explicitly addressing target-independent contributions to survival

Further insight into the functional requirements of the (6/5)AL and (6/5)ALcp motifs was revealed by examining the relative abundances of individual clusters that make up the motif across multiple populations. Variation within the (6/5)AL motif was examined by plotting the fractional abundance of sequences within the 0s population (X-axis) against their fractional abundance in the 240s populations (Y-axis) such that sequences above the diagonal are enriched relative to the population as a whole following the addition of trap. Importantly, the contribution of background nitrocellulose binding was also examined by using a color code to display on the graph the sequence abundance in the NoRT (target independent) population relative to the sequence abundance in the 0s population (Figure 5D). Clusters 240.1 and 240.33 are the most abundant representatives of the (6a/5a) and (6b/5b) loop combinations in the 240s population, respectively. Both clusters increase in abundance to approximately the same extent in response to trap. However, the target-independent background binding of cluster 240.33 in the NoRT populations is less than that of cluster 240.1 as exemplified by the most abundant sequences within each cluster (240.33.1 and 240.1.1, respectively) (Figure 5D). Therefore, the trap-dependent enrichment for cluster 240.33 appears to have a smaller background component than that of cluster 240.1, with a larger component derived

from association with RT. Thus, the slightly better RT inhibition observed for sequences with the (6b/5b) combination in DDDP assays (Figures 4C and 5B) and the significantly better inhibition in cell-based assays (see following paragraph) reflect the larger contribution of RT binding to their fitness as predicted by the inter-cluster comparison of abundance. As an additional example of the importance of explicitly monitoring target-independent enrichment, when three point mutants that exhibit limited or no enrichment within cluster 240.1 (C9U, U11C and A39G) were tested for inhibition in the context of the 49 nt versions of the (6/5)AL, they were less inhibitory in DDDP assays than the dominant loop sequence (Figure 5E), but the mutant A39G exhibited only a slight loss of inhibition despite being the least enriched of the three mutants. Examining the NoRT population revealed that the protein-independent, background binding of the C9U and U11C mutants was significantly higher than that of the A39G mutant and the parental sequence, indicating that survival of the C9U and U11C point mutants was based primarily on nitrocellulose binding and not affinity for RT, even though they differ by only one mutation from an otherwise high-affinity (6a/5a)AL aptamer. Therefore, direct sequencing of a target-independent population is critical to interpreting enrichment in the face of increasing selection pressure and ultimately in predicting the functionality of specific sequence variants.

Population dynamics and biochemical inhibition of RT predict relative potency of HIV-1 suppression

We recently developed an aptamer expression cassette that reliably delivers bioactive aptamers to mammalian cells for inhibition of single-cycle virus pseudotyped with glycoprotein from vesicular stomatitis virus via transient transfection (18). The viral genome carries the gene for EGFP, which is expressed in infected cells whenever the viral genome is copied into dsDNA. Inhibition of reverse transcription during infection can therefore be measured by a loss of fluorescence in the target cells. We used this assay to evaluate viral inhibition by the most abundant aptamers for each of the five converged motifs analyzed in the DDDP assays (Figure 6A). Importantly, the two circular permutations of the new (6/5)AL motif [a (6a/5a)AL sequence (240.1.1) and a (6b/5b)ALcp sequence (240.88.1)] show the strongest inhibition in both biochemical and cell-based assays (although the inhibition of 240.1.1 is within error of 240.5.1 in the cell-based assays). This overall trend is consistent with the motif's positive response to increased selection pressure. Further, comparison across the populations reveals that collectively the (6b/5b) combination is the only combination observed within the (6/5)AL that both increases in fractional abundance in response to addition of trap and is less abundant in the NoRT population than in the 0s population (Table 1). Therefore, the (6b/5b) combination exhibits the most unambiguously positive signature of sequence abundance across populations with respect to the desired phenotype. Consistent with this observation, the 240.88.1 sequence that contains the (6b/5b)

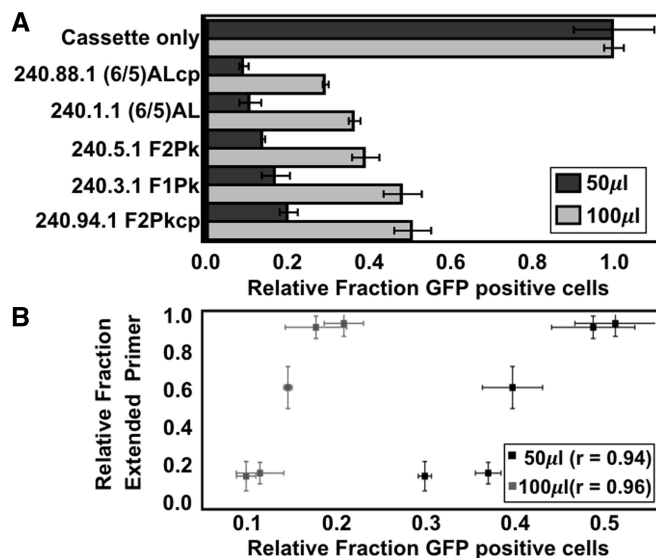


Figure 6. Correlation between viral inhibition of pseudotyped HIV and inhibition of RT enzymatic activity by specific RNA aptamers. (A) Infectivity of virus produced in 293FT cells that were transfected with aptamer expressing constructs or control (empty cassette). A total of 50 or 100 µl of virus-containing supernatant is plotted as a function of 293FT cells. After 48 h GFP-expressing cells were counted by flow cytometry as a measure of infectivity. Error bars represent the standard deviation for triplicate transfection. (B) Viral infectivity of pseudo-typed HIV from experiments using 50 or 100 µl of virus-containing supernatant is plotted as a function of DDDP activity of purified RT in the presence of the same aptamers. Pearson's correlation coefficients (r) are shown in the legend.

combination is unambiguously better at inhibiting viral replication in cell culture than all other aptamers tested in this study (Figure 6) and previous work (18) (data not shown). Overall, we observe a statistically significant correlation between the potency of viral inhibition observed in the cell-based assay (both 50 and 100 µl viral supernatant) and the inhibition of enzymatic activity observed in the biochemical DDDP assays ($r = 0.94$ and 0.96 ; for five points, $r = 0.934$ is statistically significant at the 98% level) (Figure 6B). The functional inferences derived from comparative sequence analysis of the SELEX populations are therefore relevant in both biochemical and cellular contexts.

DISCUSSION

By comparing HTS data across multiple scales (within clusters, among clusters, and between populations), we identified and characterized the (6/5)AL motif as a new, highly potent aptamer inhibitor that blocks primer extension by the RT of HIV-1 *in vitro* and that blocks viral replication of HIV-1 in cell culture. This element inhibits RT more potently than the previously characterized pseudoknots and inhibits HIV-1 replication more effectively when expressed in virus-producing cells. The HTS analysis identified an especially potent and very rare (0.4% of the pre-enriched population) variant of the (6/5)AL motif [the (6b/5b)AL] (Figure 6) and provided a

detailed description of this motif. Importantly, this description included a statistically meaningful absence of non-functional sequence combinations that otherwise satisfy a more simplistic consensus description (Table 1). The HTS analysis also enabled an inclusive definition of the F2Pk motif to describe pseudoknots that deviate from the previously identified F1Pk definition (1, 2). Thus, the high-resolution perspective afforded by HTS data sets reveals features of selected aptamer populations that can be otherwise overlooked, and accelerates the discovery, refinement and prioritization of aptamers for gene therapy and molecular virology applications.

Inference of functional structures through comparative sequence analysis is a powerful tool that has been applied since the earliest structural studies of RNA (33). The strength of the inferences increases with the diversity and population size of the functionally related sequences under evaluation. Mutations that accumulate during SELEX result in divergent evolution that reflects selection pressure acting on a collection of closely related sequences. At the same time, shared selection pressures experienced by a SELEX population frequently result in convergence among sequences from independent lineages on a shared structure. Both of these features can be exploited to infer functional structures through comparative sequence analysis; however, these features have been strikingly underutilized in prior HTS analysis of aptamer populations. In this work, initial structural inferences derived from individual clusters were used to generate preliminary secondary structure predictions, and subsequent comparisons across multiple clusters using CMs derived from these preliminary predictions ultimately identified the functional secondary structures present in the populations.

Higher-order structural features can be inferred from the HTS data

For the pseudoknots, a loop containing the sequence ANAA is a consensus feature shared by the three pseudoknot motifs (Figure 3A). In the crystal structure of a complex between RT and a pseudoknot aptamer conforming to the F1Pk motif definition, the ANAA loop forms a triple-stranded interaction in the minor groove of the conserved UCCG:CGGG stem (9). By extension, a homologous intramolecular minor groove interaction is predicted to be a feature of the F1Pk, F2Pk and F2Pkcp motifs (Supplementary Figure S7). The presence of a shared minor groove interaction would imply that these pseudoknots bind RT in the same global orientation, with functional differences arising from the differences in stem 2. This is consistent with the prior observations of functional differences between pseudoknot aptamers that conform to the F1Pk motif and those that do not (11). F1Pk aptamers are much less potent inhibitors of RT from HIV-1 subtype A than pseudoknot aptamers that do not conform to the F1Pk motif, and sensitivity to the amino acid identity at this position was shown to be a consequence of the R277K polymorphism (11). Position 277 contacts stem 2 in the crystal structure of the F1Pk (9). For the circularly permuted pseudoknots, the

position of the break in connectivity is also consistent with a common orientation for all the pseudoknots. We do not observe a circular permutation that introduces a break in the conserved ANAA sequence nor in the single-stranded region that contacts RT in this model, but rather we observe a permutation that introduces a break in a highly variable region that does not contact RT in the corresponding FIPk-RT co-crystal structure.

For the newly identified (6/5)AL motif, a structural bioinformatics search using the software *JAR3D* (34) identified multiple asymmetric loops of the same size and related sequence composition as the (6/5)AL internal loop within functionally unrelated reference RNAs. Within those reference structures, the positions equivalent to G10, U11 and A39 in the truncated (6/5)AL construct form a base triple (Figure 7). The dramatic loss of activity observed for the U11C mutant (Figure 5E) is consistent with the formation of an analogous base triple in the (6/5)AL motif.

Progressively increasing the competition for target binding is a common strategy in SELEX experiments that is intended to favor survival of the highest affinity aptamers within the population; however, this approach also runs the risk of magnifying target-independent contributions to survival. Although alternating the partitioning method and including target-independent subtractive steps can reduce this risk, aptamer sequences that bind to the target will nevertheless display varying degrees of target-independent retention that will contribute to their relative abundances in the final selected populations, as we observe here. This variation can be explicitly accounted for by directly sequencing a target-independent population (NoRT). This type of sequencing-based accounting is almost never undertaken for analysis of SELEX populations and before the emergence of HTS techniques would have likely been of limited utility given the limited statistical strength associated with LTS data sets (Supplementary Figure S4). When the stringency of the RT-binding selection was increased by raising the ionic strength and by the addition of a non-amplifiable molecular trap, the population became enriched in aptamers that form a highly stable complex with the molecular target and/or exhibit above average target-independent retention. With respect to the desired phenotype, this combination of factors favored both the best (e.g. 240.88.1) and the worst (e.g. 240.94.1) aptamers in the population at the expense of those aptamers of intermediate quality (e.g. 240.3.1). This same trend was also observed among point mutants within individual clusters. For example, the strong inhibitor 240.1.1 and the very weak inhibitors 240.1.143 and 240.1.149 all out-compete the intermediate inhibitor 240.1.173 (Figure 5). Importantly, this intracluster comparison reveals that enrichment of sequences exhibiting target-independent binding required only a single point mutation of an otherwise desirable aptamer sequence. Therefore, the potential for target-independent contributions to survival should be considered a persistent danger during a selection, which can readily re-emerge following even the most effective subtractive steps through a single point mutation of a desirable aptamer sequence. Here, target-independent

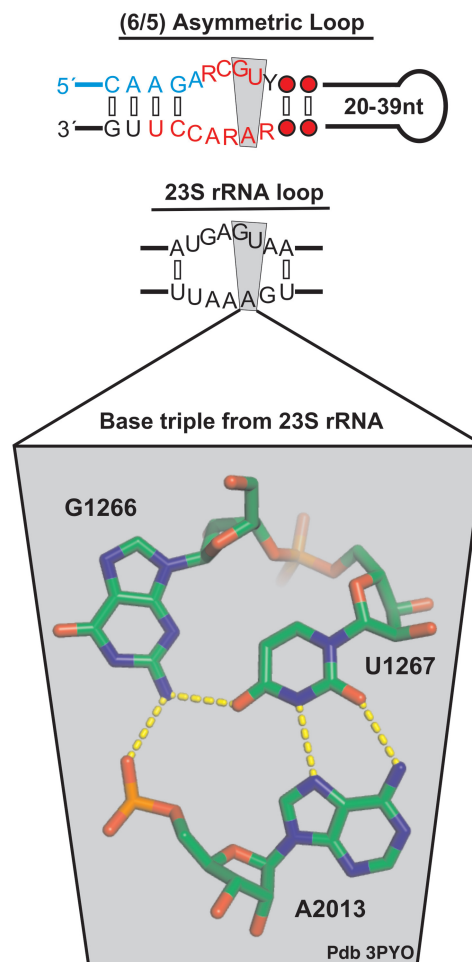


Figure 7. Inference of higher-order structural features of the (6/5)AL motif. The consensus description of the (6/5)AL motif is shown along with the sequence of a similar loop from the 23S rRNA and a base triple interaction observed in the crystal structure of the 23S rRNA. The position of a potential homologous base triple in the (6/5)AL motif is shaded in gray.

contributions persisted in spite of the fact that the selection strategy included switching partitioning methods from filter binding in rounds 1–11 to native PAGE mobility shift in rounds 12–14 and included a subtractive step to remove filter binding sequences in the final 15th round of selection.

It is clear that sheer abundance in the final population of a SELEX experiment does not adequately correlate with desired phenotype, which can be instead more accurately predicted based on population dynamics observed in response to increased competition, coupled with a critically important explicit assessment of target-independent survival. Here, this approach more accurately predicted the patterns of inhibition observed in both DDDP assays of enzymatic activity and cell-based assays of viral replication among the structural motifs that bind HIV RT. Target-independent contributions to survival discussed above, the high occurrence of simple motifs relative to more complex ones in random sequence space (35,36) and preferential amplification of specific sequences

all contribute to the poor correlation between desired phenotype and abundance. The dominant F1Pk represents a simple motif with the relatively limited sequence constraints of four highly conserved bp, 3–4 generic bp, and loops with few conserved positions. In contrast, the less abundant but more potent (6/5)AL is more constrained, requiring approximately 10–12 largely generic bp and one of a limited combination of sequences within the 11 nt asymmetric internal loop. Therefore, RNA strands carrying the simple pseudoknots are expected to have been 10^3 - to 10^6 -fold more abundant in the starting random population than those carrying either topology of the (6/5)AL motif, consistent with the dominance of the F1Pk motif in these populations.

The high-resolution view of population dynamics described here is readily achieved through comparative analysis of HTS data sets and is largely impossible with LTS data. Overall, this work demonstrates that HTS comparative sequence analysis across closely related populations enhances the capacity to infer aptamer structure and function, which can be extended to a cellular environment, and ultimately increase the efficiency and efficacy of the post-SELEX development of aptamers.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Text and Supplementary Figures 1–6.

ACKNOWLEDGEMENTS

We thank A. Petrov, C. Zirbel and N. Leontis for early access to JAR3D.

FUNDING

National Institute of Allergy and Infectious Disease [R01AI074389 to D.H.B. and F32AI085627 to M.J.L.]. Funding for open access charge: National Institutes of Health [R01 AI074389].

Conflict of interest statement. None declared.

REFERENCES

- Tuerk,C., MacDougal,S. and Gold,L. (1992) RNA pseudoknots that inhibit human immunodeficiency virus type 1 reverse transcriptase. *Proc. Natl Acad. Sci. USA*, **89**, 6988–6992.
- Burke,D.H., Scates,L., Andrews,K. and Gold,L. (1996) Bent pseudoknots and novel RNA inhibitors of type 1 human immunodeficiency virus (HIV-1) reverse transcriptase. *J Mol. Biol.*, **264**, 650–666.
- Li,N., Wang,Y., Pothukuchy,A., Syrett,A., Husain,N., Gopalakrishna,S., Kosaraju,P. and Ellington,A.D. (2008) Aptamers that recognize drug-resistant HIV-1 reverse transcriptase. *Nucleic Acids Res.*, **36**, 6739–6751.
- Andreola,M.L., Pileur,F., Calmels,C., Ventura,M., Tarrago-Litvak,L., Toulme,J.J. and Litvak,S. (2001) DNA aptamers selected against the HIV-1 RNase H display in vitro antiviral activity. *Biochemistry*, **40**, 10087–10094.
- Schneider,D.J., Feigon,J., Hostomsky,Z. and Gold,L. (1995) High-affinity ssDNA inhibitors of the reverse transcriptase of type 1 human immunodeficiency virus. *Biochemistry*, **34**, 9599–9610.
- Somasunderam,A., Ferguson,M.R., Rojo,D.R., Thivyanathan,V., Li,X., O'Brien,W.A. and Gorenstein,D.G. (2005) Combinatorial selection, inhibition, and antiviral activity of DNA thioaptamers targeting the RNase H domain of HIV-1 reverse transcriptase. *Biochemistry*, **44**, 10388–10395.
- Mosing,R.K., Mendonsa,S.D. and Bowser,M.T. (2005) Capillary electrophoresis-SELEX selection of aptamers with affinity for HIV-1 reverse transcriptase. *Anal. Chem.*, **77**, 6107–6112.
- Held,D.M., Kissel,J.D., Saran,D., Michalowski,D. and Burke,D.H. (2006) Differential susceptibility of HIV-1 reverse transcriptase to inhibition by RNA aptamers in enzymatic reactions monitoring specific steps during genome replication. *J. Biol. Chem.*, **281**, 25712–25722.
- Jaeger,J., Restle,T. and Steitz,T.A. (1998) The structure of HIV-1 reverse transcriptase complexed with an RNA pseudoknot inhibitor. *EMBO J.*, **17**, 4535–4542.
- Fisher,T.S., Joshi,P. and Prasad,V.R. (2002) Mutations that confer resistance to template-analog inhibitors of human immunodeficiency virus (HIV) type 1 reverse transcriptase lead to severe defects in HIV replication. *J. Virol.*, **76**, 4068–4072.
- Held,D.M., Kissel,J.D., Thacker,S.J., Michalowski,D., Saran,D., Ji,J., Hardy,R.W., Rossi,J.J. and Burke,D.H. (2007) Cross-clade inhibition of recombinant human immunodeficiency virus type 1 (HIV-1), HIV-2, and simian immunodeficiency virus SIVcpz reverse transcriptases by RNA pseudoknot aptamers. *J. Virol.*, **81**, 5375–5384.
- Kissel,J.D., Held,D.M., Hardy,R.W. and Burke,D.H. (2007) Active site binding and sequence requirements for inhibition of HIV-1 reverse transcriptase by the RT1 family of single-stranded DNA aptamers. *Nucleic Acids Res.*, **35**, 5039–5050.
- Kissel,J.D., Held,D.M., Hardy,R.W. and Burke,D.H. (2007) Single-stranded DNA aptamer RT1t49 inhibits RT polymerase and RNase H functions of HIV type 1, HIV type 2, and SIVCPZ RTs. *AIDS Res. Hum. Retroviruses*, **23**, 699–708.
- Michalowski,D., Chitima-Matsiga,R., Held,D.M. and Burke,D.H. (2008) Novel bimodular DNA aptamers with guanosine quadruplexes inhibit phylogenetically diverse HIV-1 reverse transcriptases. *Nucleic Acids Res.*, **36**, 7124–7135.
- Ditzler,M.A., Bose,D., Shkriabai,N., Marchand,B., Sarafianos,S.G., Kvaratskhelia,M. and Burke,D.H. (2011) Broad-spectrum aptamer inhibitors of HIV reverse transcriptase closely mimic natural substrates. *Nucleic Acids Res.*, **39**, 8237–8247.
- Joshi,P. and Prasad,V.R. (2002) Potent inhibition of human immunodeficiency virus type 1 replication by template analog reverse transcriptase inhibitors derived by SELEX (systematic evolution of ligands by exponential enrichment). *J. Virol.*, **76**, 6545–6557.
- Chaloin,L., Lehmann,M.J., Sczakiel,G. and Restle,T. (2002) Endogenous expression of a high-affinity pseudoknot RNA aptamer suppresses replication of HIV-1. *Nucleic Acids Res.*, **30**, 4001–4008.
- Lange,M.J., Sharma,T.K., Whatley,A.S., Landon,L.A., Tempesta,M.A., Johnson,M.C. and Burke,D.H. (2012) Robust suppression of HIV replication by intracellularly expressed reverse transcriptase aptamers is independent of ribozyme processing. *Mol. Ther.*, **20**, 2304–2314.
- Bock,L.C., Griffin,L.C., Latham,J.A., Vermaas,E.H. and Toole,J.J. (1992) Selection of single-stranded DNA molecules that bind and inhibit human thrombin. *Nature*, **355**, 564–566.
- Green,L.S., Jellinek,D., Jenison,R., Ostman,A., Heldin,C.H. and Janjic,N. (1996) Inhibitory DNA ligands to platelet-derived growth factor B-chain. *Biochemistry*, **35**, 14413–14424.
- Burke,D.H. and Hoffman,D.C. (1998) A novel acidophilic RNA motif that recognizes coenzyme A. *Biochemistry*, **37**, 4653–4663.
- Yu,H., Jiang,B. and Chaput,J.C. (2011) Aptamers can discriminate alkaline proteins with high specificity. *ChemBiochem*, **12**, 2659–2666.
- Gold,L., Janjic,N., Jarvis,T., Schneider,D., Walker,J.J., Wilcox,S.K. and Zichi,D. (2011) Aptamers and the RNA world, past and present. *Cold Spring Harb. Perspect. Biol.*, **4**, pii: a003582.

24. Cho, M., Xiao, Y., Nie, J., Stewart, R., Csordas, A.T., Oh, S.S., Thomson, J.A. and Soh, H.T. (2010) Quantitative selection of DNA aptamers through microfluidic selection and high-throughput sequencing. *Proc. Natl Acad. Sci. USA*, **107**, 15373–15378.
25. Hoon, S., Zhou, B., Janda, K.D., Brenner, S. and Scolnick, J. (2011) Aptamer selection by high-throughput sequencing and informatic analysis. *Biotechniques*, **51**, 413–416.
26. Kupakuwana, G.V., Crill, J.E. II, McPike, M.P. and Borer, P.N. (2011) Acyclic identification of aptamers for human alpha-thrombin using over-represented libraries and deep sequencing. *PLoS One*, **6**, e19395.
27. Bayrac, A.T., Sefah, K., Parekh, P., Bayrac, C., Gulbakan, B., Oktem, H.A. and Tan, W. (2011) In vitro Selection of DNA Aptamers to Glioblastoma Multiforme. *ACS Chem. Neurosci.*, **2**, 175–181.
28. Pitt, J.N., Rajapakse, I. and Ferre-D'Amare, A.R. (2010) SEWAL: an open-source platform for next-generation sequence analysis and visualization. *Nucleic Acids Res.*, **38**, 7908–7915.
29. Katoh, K., Kuma, K., Toh, H. and Miyata, T. (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.*, **33**, 511–518.
30. Hofacker, I.L., Fekete, M. and Stadler, P.F. (2002) Secondary structure prediction for aligned RNA sequences. *J. Mol. Biol.*, **319**, 1059–1066.
31. Nawrocki, E.P., Kolbe, D.L. and Eddy, S.R. (2009) Infernal 1.0: inference of RNA alignments. *Bioinformatics*, **25**, 1335–1337.
32. Weinberg, Z. and Breaker, R.R. (2011) R2R—software to speed the depiction of aesthetic consensus RNA secondary structures. *BMC Bioinformatics*, **12**, 3.
33. Madison, J.T., Everett, G.A. and Kung, H. (1966) Nucleotide sequence of a yeast tyrosine transfer RNA. *Science*, **153**, 531–534.
34. Rahrig, R.R., Leontis, N.B. and Zirbel, C.L. (2010) R3D Align: global pairwise alignment of RNA 3D structures using local superpositions. *Bioinformatics*, **26**, 2689–2697.
35. Ellington, A.D. (1994) The RNA world. Empirical explorations of sequence space: Host-guest chemistry in the RNA world. *Berichte der Bunsengesellschaft für physikalische Chemie*, **98**, 1115–1121.
36. Joyce, G.F. (2004) Directed evolution of nucleic acid enzymes. *Annu. Rev. Biochem.*, **73**, 791–836.