

# Nucleotide Bias Observed with a Short SELEX RNA Aptamer Library

William H. Thiel,<sup>1</sup> Thomas Bair,<sup>2</sup> Kristina Wyatt Thiel,<sup>1</sup> Justin P. Dassie,<sup>1,3</sup> William M. Rockey,<sup>4</sup>  
Craig A. Howell,<sup>1</sup> Xiuying Y. Liu,<sup>1</sup> Adam J. Dupuy,<sup>5</sup> Lingyan Huang,<sup>6</sup> Richard Owczarzy,<sup>6</sup>  
Mark A. Behlke,<sup>6</sup> James O. McNamara II,<sup>1</sup> and Paloma H. Giangrande<sup>1,3,4</sup>

Systematic evolution of ligands by exponential enrichment (SELEX) is a powerful *in vitro* selection process used for over 2 decades to identify oligonucleotide sequences (aptamers) with desired properties (usually high affinity for a protein target) from randomized nucleic acid libraries. In the case of RNA aptamers, several highly complex RNA libraries have been described with RNA sequences ranging from 71 to 81 nucleotides (nt) in length. In this study, we used high-throughput sequencing combined with bioinformatics analysis to thoroughly examine the nucleotide composition of the sequence pools derived from several selections that employed an RNA library (*Sel2N20*) with an abbreviated variable region. The *Sel2N20* yields RNAs 51 nt in length, which unlike longer RNAs, are more amenable to large-scale chemical synthesis for therapeutic development. Our analysis revealed a consistent and early bias against inclusion of adenine, resulting in aptamers with lower predicted minimum free energies ( $\Delta G$ ) (higher structural stability). This bias was also observed in control, "nontargeted" selections in which the partition step (against the target) was omitted, suggesting that the bias occurred in 1 or more of the amplification and propagation steps of the SELEX process.

## Introduction

**R**NA APTAMERS ARE single-stranded RNA oligonucleotides that bind with high affinity and specificity to target molecules. An *in vitro* process termed systematic evolution of ligands by exponential enrichment (SELEX) is used to select for RNA aptamers with desired properties (Ellington and Szostak, 1990; Tuerk and Gold, 1990). The SELEX process involves iterative rounds of positive (against a target) and negative (preclearing against a nontarget) selection starting with a highly complex RNA sequence library. To facilitate the use of RNA aptamers for *in vivo* applications, a mutant T7 RNA polymerase that can incorporate 2'-fluoro (2'-F) modified ribonucleotides (rNTPs) as well as 2'-hydroxyl (2'-OH) rNTPs is often used during the *in vitro* transcription step (Huang et al., 1997). RNA aptamers containing these modified rNTPs are resistant to nuclease digestion (Padilla and Sousa, 1999) and are thus more suitable for use in *in vivo* applications. Over the past 2 decades, SELEX technology has been used to generate high-affinity RNAs to a large number of proteins (Nimjee et al., 2004; Thiel and Giangrande, 2009). We have successfully employed aptamer technology to target several proteins such as tran-

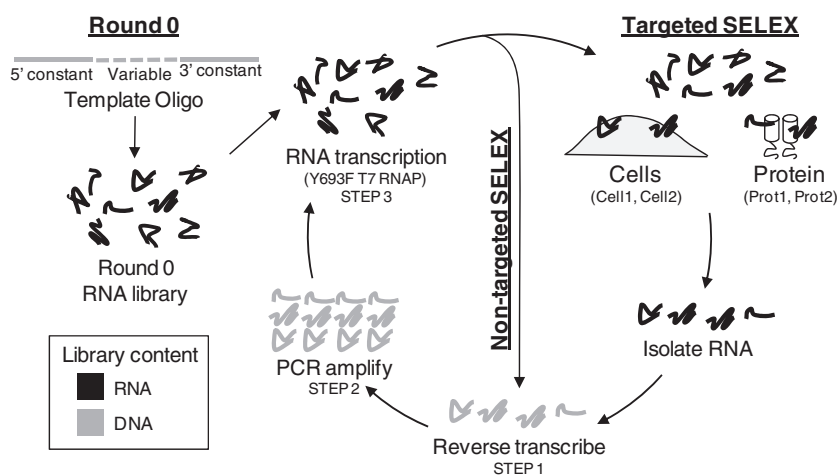
scription factors (Giangrande et al., 2007; Mi et al., 2009) and cell surface receptors (McNamara et al., 2008).

A typical RNA aptamer library starts with a DNA template oligo (Fig. 1) that contains 2 constant sequence regions (5' constant region and 3' constant region) flanking a variable region. The DNA oligo is used to generate a duplex DNA sequence library that is subsequently transcribed into RNA (Ellington and Szostak, 1990; Tuerk and Gold, 1990). The length of the variable region determines the maximum complexity of the RNA aptamer library. For example, a variable region of 20 nt will yield  $4^{20}$  sequences ( $10^{12}$  sequence complexity), whereas a variable region of 40 nt will yield  $4^{40}$  sequences ( $10^{24}$  sequence complexity). A complexity of at least  $10^{11}$  is suggested to provide the necessary structural diversity to identify specific aptamers (Sassanfar and Szostak, 1993). The RNA SELEX libraries used to date include variable regions of 30 nt (Ruckman et al., 1998), 40 nt (Ruckman et al., 1998; Lupold et al., 2002; Rusconi et al., 2002; Giangrande et al., 2007; Oney et al., 2007; Dollins et al., 2008; McNamara et al., 2008; Mi et al., 2009) and 50 nt (Zhou et al., 2009). Aptamers derived from these RNA aptamer libraries are anywhere between 61 and 81 nt long.

<sup>1</sup>Department of Internal Medicine, <sup>2</sup>DNA Facility, <sup>3</sup>Molecular & Cellular Biology Program, <sup>4</sup>Department of Radiation Oncology, and <sup>5</sup>Department of Anatomy and Cell Biology, University of Iowa, Iowa City, Iowa.

<sup>6</sup>Integrated DNA Technologies, Coralville, Iowa.

**FIG. 1.** Schematic of the targeted and nontargeted RNA aptamer selections performed with the *Sel2N20* library. Targeted systematic evolution of ligands by exponential enrichment (SELEX): 8 rounds of selection were performed against either recombinant purified mEphA2 extracellular domain protein (Prot1 and Prot2) or breast cancer cell lines N202.1A and MDA-MB-231 (Cell1 and Cell 2). The round 0 library was initially incubated with a respective nontarget protein or cell line (preclear step). The precleared RNA library was subsequently incubated with the target. The target-bound RNA aptamers were recovered, reverse transcribed (step 1), and PCR amplified with primers specific to the *Sel2N20* constant regions (step 2), and the duplex PCR product was *in vitro* transcribed with the T7(Y693F) RNA polymerase in the presence of 2'-fluoro pyrimidines (step 3). Steps 1–3 were repeated 5 more times. Nontarget-SELEX: 6 rounds of selection in the absence of a target were performed with an RNA pool derived from the *Sel2N20* library. The RNAs were reverse transcribe, PCR amplified, and *in vitro* transcribed into the next round of selection as described above.



While longer variable regions are desirable during the SELEX process since they translate into RNA libraries with higher sequence complexity, RNA aptamers derived from these libraries are often truncated to shorter (<40 nt) functional sequences after selection (Biesecker et al., 1999; Rusconi et al., 2002; Dassie et al., 2009), a process that can be both time consuming and arduous. Truncation of these RNA aptamers is required to enable efficient, large-scale chemical synthesis of the RNAs for subsequent evaluation in animal models of disease or clinical trials (Ruckman et al., 1998; Gragoudas et al., 2004). RNA oligonucleotides >60 nt in length have a high rate of chemical synthesis failure, which is further compounded by the introduction of modified nucleotides (eg, 2'-F pyrimidines or 2'-O-methyl rNTPs) during the synthesis step. This creates a significant hurdle for developing many of these long RNA aptamers into viable human therapies (Reese, 2005). Therefore, ways to expedite the identification of shorter (<60 nt), functional RNA aptamers sequences are highly desirable.

To eliminate the truncation step and facilitate large-scale chemical synthesis of RNA aptamers soon after selection, we have designed and evaluated a short RNA aptamer library, *Sel2N20*. The constant regions of the *Sel2N20* library were derived from the Sel2 SELEX RNA library (Ruckman et al., 1998; Lupold et al., 2002), but the variable region was reduced from 40 to 20 nt in length. This shortened, modified library produces RNAs of 51 nt in length (variable region = 20 nt; 5' constant region = 15 nt; 3' constant region = 16 nt). Despite the shorter variable region, the *Sel2N20* library still contains a complexity ( $10^{12}$ ) greater than needed (theoretically  $10^{11}$ ) to achieve the necessary structural diversity of a starting library (Sassanfar and Szostak, 1993). We have used this library to identify aptamers that bind with high affinity and specificity to their targets in both cell-based and recombinant protein selections (unpublished data) (Fig. 1). The SELEX process may contain an inherent bias that favors the amplification of certain sequences over others (Meyers et al., 2004), and we sought to determine the nucleotide composition of RNA sequences that were enriched during various selections performed with the *Sel2N20* library. High-throughput (454) sequencing (Zimmermann et al., 2010) was performed to

enable the analysis of thousands of sequences at each round of selection for each selection performed. Bioinformatics analysis of the sequences revealed a bias toward pyrimidine-rich sequences for all selections performed with this library. We thus set out to elucidate the nature of the bias.

## Materials and Methods

### Cell culture

All cell lines were cultured at 37°C under 5% CO<sub>2</sub>. The N202.1A and N202.1E cell lines were cultured in DMEM (Gibco, 11965) supplemented with 10% FBS (Hyclone, SH30071.030). The MDA-MB-231 cells were cultured in Leibovitz's L-15 media (Gibco, 11415) with 10% FBS (Atlanta Biologicals, S11550). The MCF10A cells were cultured in MEGM BulletKit media (Lonza, CC-3150) without GA-1000 (gentamycin-amphotericin B mix) and supplemented with 100 ng/mL cholera toxin (Sigma, C8052). All cell lines were split either 1:3 or 1:4 upon reaching confluence by washing with DPBS (Gibco, 14190-144) and detaching cell using either 0.05% Trypsin-EDTA (Gibco, 25300) or 0.25% Trypsin-EDTA (Gibco, 25200) as recommended for each cell line by ATCC.

### Targeted and nontargeted SELEX

The *Sel2N20* DNA template oligo, 5'-TCGGGCGAGTC GTCTG-N20-CCGCATCGTC CTCCC-3' (IDT) was extended using the Sel2 5' primer, 5'-TAATACGACTCAC TATAGGG AGGACGATGCGG-3', to generate the round 0 (Rnd 0) RNA library used for both targeted and nontargeted selections. Choice Taq polymerase (Denville Scientific, Inc., CB40502) was used in the extension reaction to generate the initial duplex DNA library (anneal protocol: 95°C 3 minutes, 25°C 10 minutes; extend protocol: 72°C 30 minutes, 25°C 10 minutes). The duplex DNA library was *in vitro* transcribed (37°C 16 hours) with Y639F mutant T7 RNAP (Sousa and Padilla, 1995; Huang et al., 1997) to enable incorporation of 2'-OH purines (Roche; GTP, 14611221; ATP, 14919320) and 2'-fluoro 2'-F pyrimidines (TriLink BioTechnologies; 2'-fluoro-2'-dCTP, N-1010020509; 2'-fluoro-2'-dUTP, N-1008-013008) in the Rnd 0

RNA library. During the transcription reaction, the final concentration of rNTPs for targeted selections was 4 mM with a 3:1 ratio of 2'-F pyrimidines to 2'-OH purines (3 mM 2'-F pyrimidines and 1 mM 2'-OH purines). For nontargeted selections, both 3:1 and 1:1 ratios of 2'-F pyrimidines to 2'-OH purines were used, with a final rNTP concentration of 4 mM. An optimized rNTP ratio based on the sequences of the constant regions flanking the N20 random region of the *Sel2N20* library was also developed with 2.2 mM 2'-F C's, 1.1 mM 2'-F U's, 1.88 mM 2'-OH A's, and 2.82 mM 2'-OH G's.

Targeted selections were performed against either recombinant purified mouse EphA2 (R&D Systems, 639-A2) or mammary cancer cells (Cell1: HER2-positive N202.1A; Guido Forni) (Cell2: MDA-MB-231, ATCC HTB-26). For the targeted recombinant protein selections, the Rnd0 RNA library and RNA pools from subsequent selection rounds were precleared against affinity beads (Invitrogen, Dynabeads TALON, 101.02D). For the HER2-positive N202.1A mouse mammary carcinoma cells the Rnd0 RNA library and RNA pools from subsequent rounds of selection were precleared against mouse mammary carcinoma cells (N202.1E, Guido Forni) lacking HER2 expression. For the MDA-MB-231 cell-based selection the preclear was done against normal human mammary epithelial cells (MCF10A, ATCC CRL-10317). The targeted selections were carried out for a total of 8 rounds. The target-bound RNA aptamers were recovered by either chloroform extraction (recombinant protein selection) or TRIzol (Invitrogen, 15596-026) extraction. The recovered RNA aptamers were reverse transcribed (Invitrogen; Superscript III, 56575) using the Sel2 3' primer, 5'-TCGGGCGAGTCGCTG-3' [reverse transcriptase (RT) protocol: 55°C 60 minutes, 72°C 15 minutes] and PCR amplified (Choice Taq DNA polymerase; Denville Scientific, Inc., CB4050-2) using the Sel2 5' primer, Sel2 3' primer and 2.5 mM dNTPs (10 mM dNTP mix; Invitrogen, 100004893) [PCR Protocol: 95°C 2 minutes (95°C 30 seconds, 55°C 30 seconds, 72°C 15-30 seconds) × 25 cycles, 72°C 5 minutes]. For the nontargeted selections, RNA aptamers bypassed the preclear/target steps of SELEX by reverse transcribing and PCR amplifying 1 pmol of RNA at each round.

#### 454 deep sequencing

PCR product, or Rnd0-extended template, from selected rounds was PCR amplified using 454 primers with unique barcodes to identify rounds. Each PCR product was run on 3% agarose gel, along with a PCR water control, to confirm size. The precise concentration of the 454 PCR-amplified DNA template was assessed using Quant-iT dsDNA HA Assay (Invitrogen, Q32851). Samples were combined (1 pmol/sample) and submitted for 454 deep sequencing (University of Iowa DNA facility). Sequences obtained from 454 deep sequencing were sorted using bioinformatics analysis based upon barcodes and the variable region identified as the sequence between the 3' and 5' constant regions. Incomplete sequences and sequences containing N's were removed from the analyzed data set. Four hundred fifty-four deep sequencing yielded 190,612 unique sequences (262,539 total). A total of 44,920 unique sequences (50,085 total sequences) were obtained for round 0. About 55,754 unique sequences were identified (117,507 total sequences) from 13 independent rounds of targeted selections (average 9039 sequence/round). About 89,638 unique sequences (94,947 total sequences) were

identified from 9 rounds of nontargeted selections (average 10,549 sequences/round).

#### High-performance liquid chromatography-based quantitation of NTP concentrations

The triphosphate groups from the nucleotides (ribonucleotides and deoxynucleotides) were removed using shrimp alkaline phosphatase (Promega, M9910). The nucleosides were then separated by high-performance liquid chromatography (HPLC) (Agilent 1100 with Diode Array Detector), using the Clarity column (Phenomenex) and 100 mM TEAA pH7.0 with 5% (v/v) acetonitrile. Peak identification was achieved by comparing the absorption profile of each peak to that of known nucleosides. The absorption spectrum of 2'-F uridines and 2'-F cytosines is similar to that of 2'-OH uridines and 2'-OH cytosines, respectively. The molar ratio of the 2'-F and 2'-OH mixes was calculated from nucleoside extinction coefficients and areas below HPLC peaks (see Table 1). The extinction coefficient values for uridine and cytidine were used in place of 2'-F uridine and 2'-F cytidine. The ratios were normalized by setting either deoxyadenosine or adenosine to the expected value.

#### Bioinformatics

The variable region sequences found for each round, sequence between the 5' and 3' constant regions, were analyzed for nucleoside content by counting (NoteTab Light, Count Occurrences function) the total number of each nucleoside (A's, G's, U's, and C's). These data were entered into Microsoft Excel for further data analysis. For every round of selection data, the percent of total bases was then determined for each nucleoside. Pyrimidines (U's and C's) and purines (A's and G's) were totaled and the pyrimidine value was divided by the purine value to yield the pyrimidine to purine ratio (Pyr:Pur).

The distribution of the dinucleoside pairs was determined by counting the number of times a dinucleoside pair was

TABLE 1. HIGH-PERFORMANCE LIQUID CHROMATOGRAPHY-BASED QUANTITATION OF NTP CONCENTRATIONS IN NTP MIXES

Sample	NTP	A260 Area	e260 std	Molar ratio	
				Calculated	Expected
dNTP mix	C	58.4	7400	1.1	1.0
	G	86.9	11500	1.1	1.0
	T	64.2	8700	1.0	1.0
	A	109.1	15400	1.0 <sup>a</sup>	1.0
rNTP mix (1:1)	fC	55.5	7400	1.3	1.0
	rG	74.9	11500	1.1	1.0
	fU	65.1	9900	1.2	1.0
	rA	88	15400	1.0 <sup>a</sup>	1.0
rNTP mix (3:1)	fC	161.9	7400	3.1	3.0
	rG	87.8	11500	1.1	1.0
	fU	203.4	9900	2.9	3.0
	rA	107.9	15400	1.0 <sup>a</sup>	1.0
rNTP mix Opti	fC	129.6	7400	12.9	11.0
	rG	247.5	11500	15.8	14.1
	fU	75.9	9900	5.6	5.5
	rA	197.1	15400	9.4 <sup>a</sup>	9.4

<sup>a</sup>Ratios are normalized by setting base A at expected value.

found within each variable region using the NSTACK macro function (IDT) within Microsoft Excel. For each dinucleoside pair, the percent of total pairs found was determined and then the percent change from round 0 calculated.

The minimum free energy was determined by first adding the 5' and 3' constant region sequence to each unique variable region sequence. The full aptamer sequence was then analyzed by RNAfold (Vienna RNA Package 1.8.2) (Hofacker et al., 1994) to determine the minimum free energy ( $\Delta G$ , kcal/mol) under default conditions (37°C, 1 M Na<sup>+</sup> solution).

#### Transcription efficiency measurements

**Y639F T7 RNAP efficiency.** ssDNA template oligos (IDT) were designed to have either a high adenosine content (75%) or a low adenosine content (25%). Two oligos for each condition were designed. The oligos were extended to generate dsDNA templates using the Sel2 5' primer. After the extension reaction the duplex DNA templates were desalted and separated from the excess Sel2 5' primer using a miniprep column (QIAprep Spin Miniprep Kit No. 27106, Qiagen Sciences). The size and purity of each dsDNA template was confirmed on a 3% agarose gel. The dsDNA templates were transcribed using the mutant T7 RNA polymerase enzyme as described above. After *in vitro* transcription step, the 2 templates with low adenosine content yielded RNA oligos low in adenosines and with a predicted  $\Delta G$  of -15 kcal/mol: Template 1 (low A): 5'-TCGGGCGAGTCGTCTGGACAGAAACAGGTCTGGAGGCCGCATCGTCTCCCC-3', Template 2 (low A): 5'-TCGGGCGAGTCGTCTGCATGAATCAAGG GGATAAAGCCGCATCGTCTCCCC-3' The 2 templates with high adenosine content yielded RNA oligos high in adenosines with a predicted  $\Delta G$  of -15 kcal/mol: Template 3 (high A): 5'-TCGGGCGAGTCGTCGTCCCTTATGTTTCGTCATTGCCGCATCGTCTCCCC-3' Template 4 (high A): 5'-TCGGGCGAGTCGTCTGCCTTTGCATCCTTAGTACTCCCGCATCGTCTCCCC-3'. *In vitro* transcription assays were done under the same conditions described for the SELEX process using the 3:1 ratio of 2'-F-pyrimidines to 2'-OH-purines at different times; 0, 15, and 960 minutes (16 hours). Each *in vitro* transcription assay was initiated by the addition of the Y639F T7 RNAP and terminated by heat inactivation (75°C for 10 minutes) except time 0 minute where Y639F T7 RNAP was heat inactivated and then added to the reaction. After heat inactivation, DNase I (No. 04-716-728-001; Roche) was added to the transcription reactions and incubated at 37°C for 30 minutes to digest away any dsDNA template. To determine Y639F T7 RNAP efficiency, total *in vitro* transcribed RNA was measured using RiboGreen (Invitrogen, Molecular Probes, R-11490) and a plate reader (Analyst AD&HT 96-384; LJI BioSystems, Inc.). Fluorescence data were normalized using fluorescence measured at time 0.

**Superscript III RT efficiency.** The same 4 DNA templates used for the Y639F T7 RNAP efficiency experiment where *in vitro* transcribed and gel purified for use in testing the efficiency of Superscript III RT (56575, Invitrogen). The Sel2 3' primer was labeled with <sup>32</sup>P  $\gamma$ -ATP using T4 PNK (M0201, NEB) at 37°C for 30 minutes followed by heat inactivation at 65°C for 20 minutes. Free <sup>32</sup>P  $\gamma$ -ATP was separated from the labeled Sel2 3' primer by centrifugation through G25 sephadex beads (17-0572-02; GE Healthcare Life Sciences). RT reactions were carried out as described above for the SELEX

process using 100 nM template RNA and 2  $\mu$ M labeled Sel2 3' primer. The template RNA and primer were allowed to anneal at 65°C followed by 20 minutes at RT. After annealing, the RT reaction was initiated by the addition of Superscript III (100 U/25  $\mu$ L reaction) with dNTPs (200  $\mu$ M final concentration) and terminated after 60 minutes by the addition of EDTA (~7.5  $\mu$ M final concentration) and 2 $\times$ formamide gel loading dye. Time 0 had EDTA added and 2 $\times$ formamide gel loading dye added before the addition of Superscript III and dNTPs. Each sample was run on a 10% acrylamide gel with urea at 200 V for ~20 minutes using prewarmed 0.5 $\times$ TBE (~65°C–70°C). The gel was dried and exposed using a phosphorimager (FujiFilm LifeScience, FLA 7000). Exposed bands of non-extended primer and extended product were quantified using the phosphorimager software (FujiFilm LifeScience, FLA-7000 v1.1).

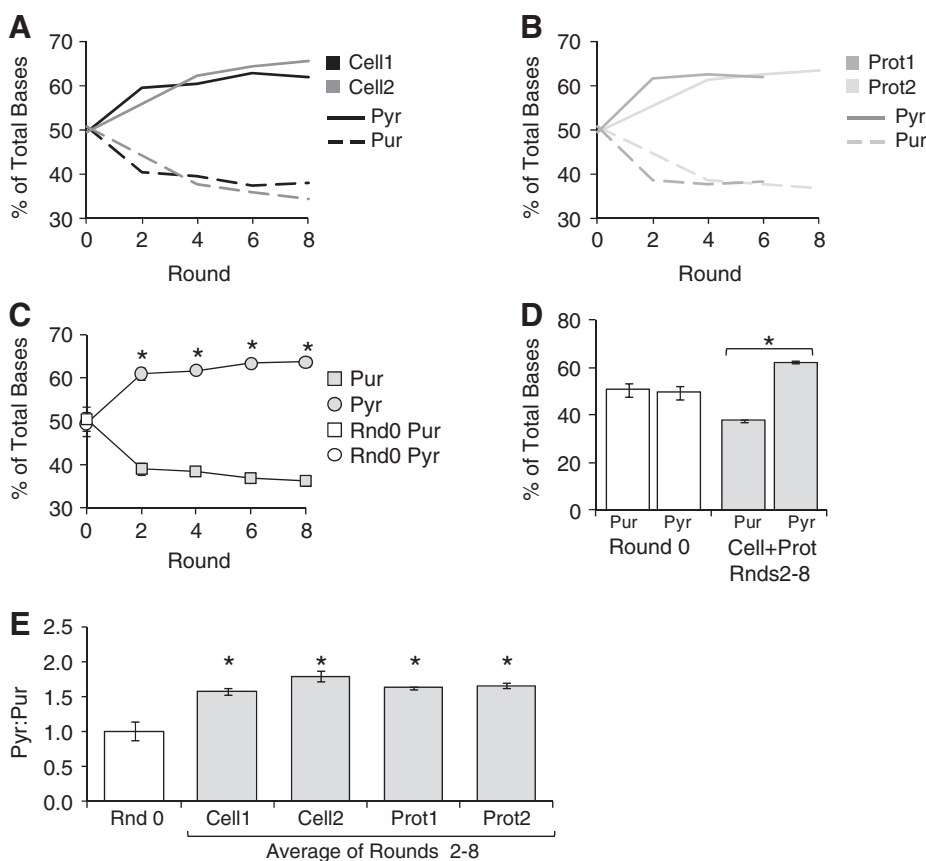
#### Data analysis and statistics

Round 0 was divided into 5 sets of ~10,000 sequences and randomized, and the mean with standard error of the mean (SEM) calculated. Each round of a targeted selection or nontargeted selection was compiled using means and error was calculated using SEM. Multiple rounds for each targeted and nontargeted selection were averaged with SEM. Overall significance ( $P < 0.05$ ) for 2 groups of data over 2 variables were determined using a 2-way analysis of the variance (ANOVA). Significance ( $P < 0.05$ ) between 2 groups was tested using a Student's *t*-test. Programs used to handle data and statistics included Microsoft Excel and SigmaStat.

## Results

### Enrichment for pyrimidine-rich RNA sequences observed with the Sel2N20 library

Targeted selections using the Sel2N20 library were performed against 2 different mammary carcinoma cell lines (Cell1 and Cell2): one of mouse origin (N202.1A) and one of human origin (MDA-MB231) (Fig. 1). In addition to cell-based selections, we also carried out 2 selections against a recombinant cell surface receptor (mouse EphA2) by varying the SELEX conditions (eg, salt concentration in the binding buffer, the protein:RNA ratio, and preclear conditions) (Prot1 and Prot2). A total of 8 rounds were performed for each selection. Several rounds (rounds 2, 4, 6, and 8) from each of the targeted selections were submitted for 454 sequencing. Bioinformatics analysis of the thousands of selected RNA sequences from rounds 2, 4, 6, and 8 of the selections revealed an inherent bias toward pyrimidine-rich sequences (Fig. 2A, B). The extent of the bias was consistent from selection to selection favoring the incorporation of 2'-F pyrimidines over 2'-OH purines in the selected RNA sequences (compare rounds 2 through 8 vs. round 0;  $P < 0.001$ ) (Fig. 2C). On average for all of the targeted selections, the percent nucleotide composition of the selected RNA sequences was as follows: 62.3% 2'-F pyrimidines vs. only 37.7% 2'-OH purines (Fig. 2C). In contrast, the starting library (round 0) contained an even distribution of rNTPs that was not found to be significantly different ( $P = 0.78$ ) with a percent nucleotide composition of 49.4% 2'-F pyrimidines and 50.6% 2'-OH purines for all the sequences analyzed (Fig. 2C, D). The shift toward 2'-F pyrimidines-rich sequences was observed as early as round 2 for all of the selections (Fig. 2C–



**FIG. 2.** Nucleotide bias observed in the targeted selections. A bias for pyrimidines (Pyr) over purines (Pur) is observed in both the cell-based (**A**) as well as the recombinant protein (**B**) targeted selections. (**C**) The number of pyrimidines (filled circles) and purines (filled squares) were averaged at each round of selection for all selections analyzed. A 2-way ANOVA ( $P < 0.001$ ) and Student's *t*-test ( $*P < 0.001$ ) were performed at each round of selection. The number of pyrimidines (empty circles) and purines (empty squares) for the round 0 RNA library is also reported. (**D**) Average of all rounds from the various selections indicates a significant difference in the number of pyrimidines vs. purines incorporated ( $*P < 0.001$ ). No significant difference was observed in round 0 between pyrimidines and purines. (**E**) The pyrimidine to purine ratio (Pyr:Pur) was significantly increased ( $*P < 0.05$ ) for each selection as compared to round 0 (Rnd 0). Between selections no significant difference was observed in the Pyr:Pur ratio.

E), suggesting that the bias might be due to some component of the SELEX process (eg, rNTP/dNTP mixes, DNA and RNA polymerases, or the RT), and not due to the nature of the target.

#### The pyrimidine bias is not target dependent

To rule out the possibility that our targets (N202.1A and MDA-MB-231 cells and mEphA2 recombinant protein) were somehow contributing to the bias, we performed a nontargeted selection. The nontargeted selection involves iterative rounds of reverse transcription, followed by PCR amplification and RNA transcription without exposing the RNA library to any particular target (Fig. 1). A total of 6 rounds of nontargeted selection were performed. Rounds 2, 4, and 6 of the nontargeted selection were sequenced with the 454 platform.

Bioinformatics analysis of the nucleotide composition of the selected sequences from rounds 2, 6, and 8 of the nontargeted selection also revealed an inherent bias for pyrimidines over purines, which was similar to what was observed for the targeted selections (Fig. 3A) ( $P = 0.11$ ). These data suggest that the SELEX target is not responsible for the enrichment of pyrimidine-rich sequences observed with the *Sel2N20* library.

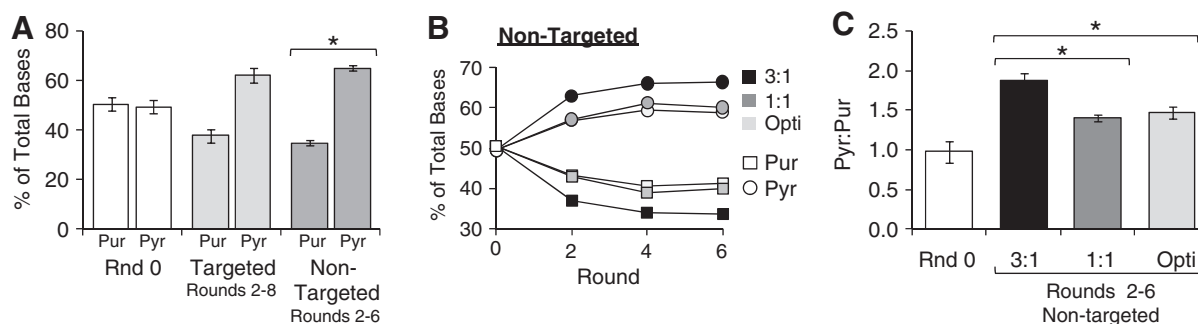
#### Altering the pyrimidine:purine rNTP ratio during *in vitro* transcription only minimally affects the nucleotide bias

As discussed above, aptamers intended for *in vivo* applications are usually selected from RNA libraries that incor-

porate modified rNTPs (eg, 2'-F pyrimidines) within the RNA sequences to provide resistance against nucleases (Padilla and Sousa, 1999). Due to the inefficiency of incorporation of 2'-F modified rNTPs into RNA by the T7 RNA polymerase enzyme, a mutant T7 RNA polymerase (Y639F mutant T7 RNAP) is used to increase the efficiency of incorporation of 2'-F rNTPs during *in vitro* transcription (Padilla and Sousa, 1999). A ratio of 3:1 (modified pyrimidines:nonmodified purines) was selected for the incorporation of 2'-F pyrimidines into RNA aptamers due to the inefficiency of 2'-F incorporation (Huang et al., 1997). Many RNA aptamer selections, with an *in vivo* end goal, have been performed using ratios of 2'-F pyrimidines to 2'-OH purines of 3:1 (Ruckman et al., 1998; Lupold et al., 2002; Rusconi et al., 2002; Giangrande et al., 2007; Oney et al., 2007; Dollins et al., 2008; McNamara et al., 2008; Mi et al., 2009; Zhou et al., 2009).

To determine whether the 3:1 ratio of pyrimidines to purines was contributing to the observed pyrimidine bias, we altered the ratio of 2'-F pyrimidines to 2'-OH purines in the transcription reaction to either reflect an even distribution of 2'-F pyrimidines to 2'-OH purines (1:1 rNTP mix) or to reflect the pyrimidine:purine ratio composition of the 5' and 3' constant regions of the *Sel2N20* RNA library (Opti mix; 07:1 rNTP mix). We verified the composition of each rNTP in these mixes using HPLC analysis (Table 1). In addition to the rNTP mixes, we also verified that the dNTP mix used for the reverse transcription and the PCR amplification step was not biased against any one NTP (Table 1).

Nontargeted SELEX using the 1:1 and Opti rNTP mixes was performed as described above for the 3:1 rNTP mix.



**FIG. 3.** The pyrimidine bias is partially influenced by the NTP ratio during RNA transcription. **(A)** A similar bias for pyrimidines over purines ( $*P < 0.001$ ) was observed for the nontargeted selections. Data are presented as the average of selected rounds (rounds  $\geq 2$ ) for the targeted and nontargeted selections. The ratio of pyrimidines:purines in the rNTP mix used for the *in vitro* transcription step for both the cell and protein targeted and nontargeted selections was 3:1. **(B)** The pyrimidine bias observed was only moderately decreased by adjusting the ratios of pyrimidines:purines in the rNTP mix. Compare the 3:1 ratio (black) to the 1:1 (gray) and optimized (Opti) rNTP (0.7:1) (white). Pyrimidines (open squares), purines (open circles). **(C)** Average of pyrimidine:purine (Pyr:Pur) ratio for rounds 2–6 of the nontargeted selection compared to that of Rnd 0. The average was calculated for each selection performed using different rNTP ratios (3:1, 1:1, and Opti) ( $*P < 0.05$ ).

Rounds 2, 4, and 6 of the 1:1 and Opti mix selections were submitted for 454 sequencing. Bioinformatics analysis of the sequences from these selections revealed a reduction in pyrimidine-rich sequences compared to the 3:1 nontargeted selection (Fig. 3B, C) (3:1 vs. 1:1  $P = 0.019$ ; 3:1 vs. Opti  $P = 0.026$ ). However, both the 1:1 and Opti mix nontargeted selections still exhibited a significant pyrimidine bias compared to round 0 (Fig. 3C) (1:1 vs. round 0  $P = 0.040$ ; Opti vs. round 0  $P = 0.024$ ). Interestingly, the Opti rNTP mix with a pyrimidine to purine ratio of 0.7:1 displays a similar bias to the 1:1 rNTP mix ( $p = 0.49$ ), despite the lower concentration of pyrimidines to purines in this mix. Together, these data suggest that the ratio of 2'-F pyrimidines to 2'-OH purines in the transcription reaction may affect the degree of the pyrimidine bias but not eliminate it.

#### Reduction in the incorporation of adenine in the selected RNA sequences

To better understand the nature of the 2'-F pyrimidine bias observed with the *Sel2N20* RNA library, we examined the abundance of each ribonucleotide in the selected RNA sequences over the course of the targeted and nontargeted selections (Fig. 4). We observed a loss of adenine (purine) in both the targeted and nontargeted selections (Fig. 4A). The extent of decrease in the incorporation of adenine was comparable between the targeted or nontargeted selections ( $P > 0.1$ ) (Fig. 4A, top/middle vs. bottom panel). In contrast, we observed an increase in guanines (purine) as well as cytosines and uracil (pyrimidines) in the targeted selections (Fig. 4A, top and middle panels). The increase in these ribonucleotides was observed in all of the targeted selections but to varying degrees (Fig. 4A, top and middle panels).

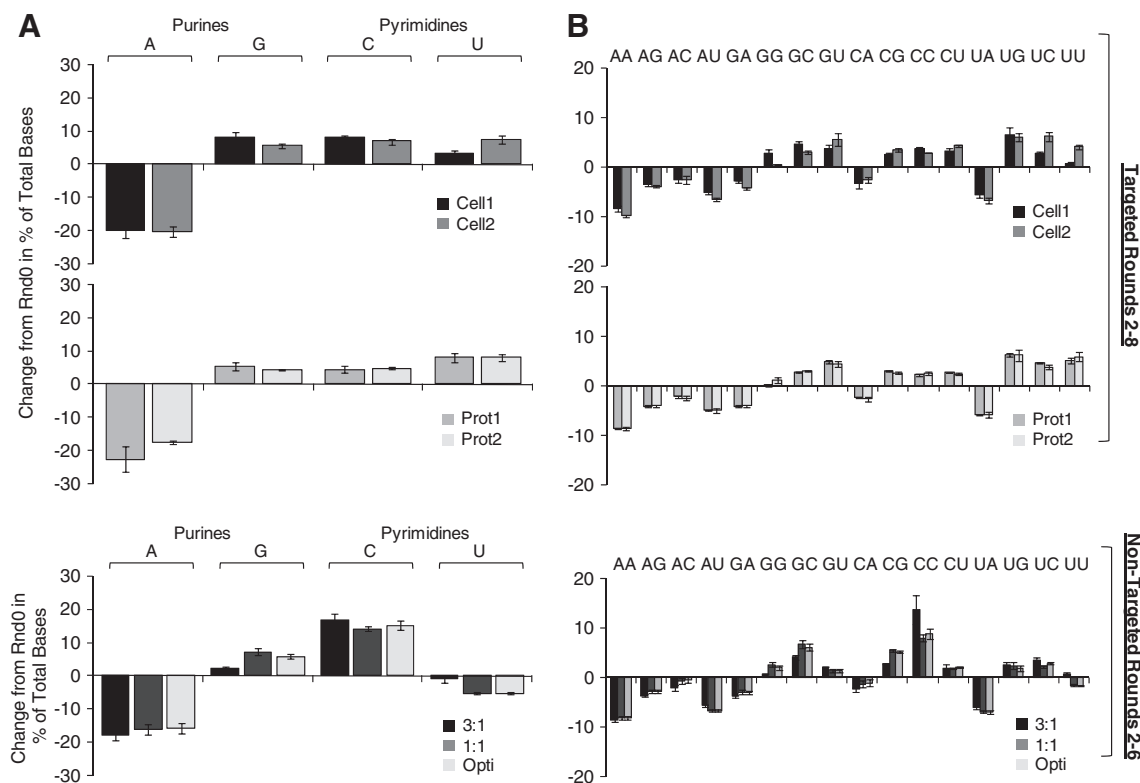
An increase in guanines and cytosines was also observed in the selected RNA sequences from the nontargeted selections performed under the different rNTP ratios (Fig. 4A, bottom panel). Interestingly, in contrast to the targeted selections, the selected RNA sequences from the nontargeted selections displayed a slight loss of uracil (pyrimidine) (Fig. 4A, bottom panel). Further, the reduced pyrimidine bias observed with the 1:1 and Opti nontargeted selections (Fig. 3B, C) appears to be

due to a gain in guanines and a loss of uracil as compared to the 3:1 rNTP mix nontargeted selection (Fig. 4A, bottom panel). In conclusion, while a loss of adenines was observed for all selections (targeted vs. nontargeted), the abundance of other ribonucleotides in the selected RNA sequences was not consistent between the targeted and nontargeted selections. These data suggest that the loss of adenines is independent of the targeting steps and is not sensitive to the ratio of rNTPs used during RNA transcription. In contrast, the abundance of other ribonucleotides (guanines, cytosines, and uracil) in the selected RNA sequences may be affected by the targeting steps (compare levels of uracil in the targeted vs. nontargeted selections) (Fig. 4A, top/middle panels vs. bottom panel) and the ratio of rNTPs used during the RNA transcription step (compare levels of guanines and uracil in nontargeted selections performed under varying rNTP mixes) (Fig. 4A, bottom panel).

We expanded the ribonucleotide analysis to include an analysis of the frequency of diribonucleotide pairings within the selected RNA sequences (Fig. 4B). The adjacent base to each ribonucleotide is thought to influence the stability of the RNA structure (Borer et al., 1974; Freier et al., 1986; Mathews et al., 1999). Di-ribonucleotide pairs that include adenine consist of higher (more positive) free energy than the remaining di-ribonucleotide pairs (Borer et al., 1974; Freier et al., 1986; Mathews et al., 1999). Not surprisingly, a reduction in adenine-containing diribonucleotide pairs is observed for the selected RNA sequences over the course of the selection process. The reduction of adenine is observed in all adenine-containing diribonucleotide pairs (AA, AU, AC, AG, UA, CA, GA) for both targeted selections (Fig. 4B, top and middle panels) and nontargeted selections (Fig. 4B, bottom panel). Together, these data suggest that the overall minimum free energy of the RNA aptamer library decreases over the course of the selection process resulting in RNA sequences with higher predicted structural stability.

#### Enrichment for RNA sequences with lower free-energy with the *Sel2N20* RNA aptamer library

The ribonucleotide analysis illustrated in Fig. 4 revealed a significant loss of adenines and an increase in guanines and



**FIG. 4.** The pyrimidine bias results from a loss of A's. **(A)** The change in % of total bases for each rNTP for each cell-based (N202.1A and MDA-MB-231) (*top panel*) and recombinant protein (mEphA2) (*middle panel*) targeted selections and nontargeted selections with different rNTP ratios, 3:1, 1:1 and optimized (Opti) (*bottom panel*) used during the *in vitro* transcription step. **(B)** The change in % of total bases for each diribonucleotide pair for each targeted selection, cell-based (*top panel*) and recombinant (*middle panel*), and nontargeted selection (*bottom panel*).

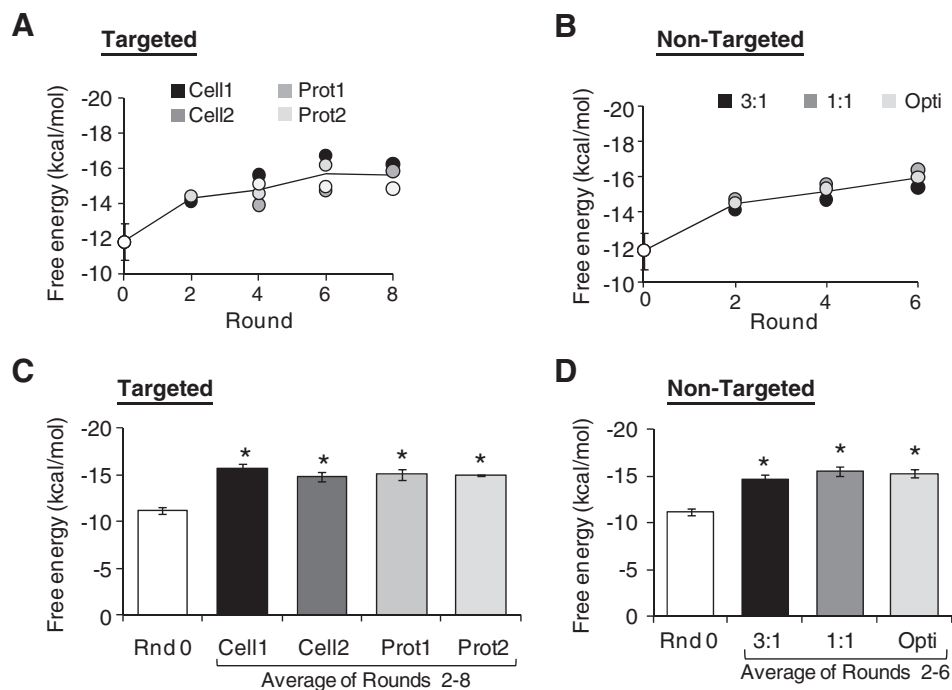
cytosines in the selected RNA sequences for all selections (targeted and nontargeted) performed with the *Sel2N20* library. Further, this loss of adenines included all di-ribonucleotide pairs with adenine. These data suggest that we are enriching for RNA sequences with an overall higher structural stability (Borer et al., 1974; Freier et al., 1986; Mathews et al., 1999). To assess the degree of structural stability of the selected RNAs sequences, we used an RNA structure algorithm (Vienna Package, RNAfold). This algorithm was used to predict the minimum free energy ( $\Delta G$ ) of the full length RNA aptamer sequences (5' constant region + variable region + 3' constant region). The average minimum free energy for each round of the targeted selections (Fig. 5A) and nontargeted selections (Fig. 5B) was determined. These data were compared to the average minimum free energy obtained for the round 0 RNA sequences (Fig. 5A, B). For both the targeted (Fig. 5A, C) and nontargeted (Fig. 5B, D) selections, the predicted minimum free energy ( $\Delta G$ ) of the selected RNAs (rounds 2–8 for the targeted selections; rounds 2–6 for the nontargeted selections) was found to be significantly less than round 0 (vs. round 0: N202.1A selection  $P = 0.017$ ; MDA-MB-231 selection  $P = 0.045$ , mEphA2 (Prot1) selection  $P = 0.036$ , mEphA2 (Prot2) selection  $P = 0.038$ , nontargeted selections at various 2'-F-pyrimidine to 2'-OH-purine ratios: 3:1 pyr:pur ratio  $P = 0.046$ , 1:1 pyr:pur  $P = 0.046$ , Opti pyr:pur  $P = 0.026$ ). The decrease in minimum free energy was consistent for all selections performed; no significant difference was found between any 2 selections ( $P > 0.3$ ) (Fig. 5C, D). These data

suggest that the ribonucleotide bias observed in both targeted and nontargeted selections (see Figs. 4 and 5) results in increased structural stability as indicated by decreased minimum free energy of the RNA aptamers.

#### *Transcription efficiencies of the SELEX enzymes do not account for the observed nucleotide bias*

Overall, our data suggest that factors intrinsic to the selection process (eg, enzymes and selection conditions) may be responsible for the observed nucleotide bias. To determine whether the mutant T7 RNA polymerase or the RT enzymes were influencing the sequence composition of the selected RNAs, we used synthetic oligonucleotide templates to directly measure the transcription efficiencies during mutant T7 RNA transcription (Fig. 6A) and reverse transcription (Fig. 6B). The variable regions of the synthetic templates were designed to have either a low A content (25% A) or a high A content (75% A). Interestingly, we observed that the mutant T7 RNA polymerase favors the transcription of templates with high adenosine (rA) content at early time points (15 minutes into the transcription reaction) ( $P < 0.001$ ) (Fig. 6A). This is probably due to the higher rate of incorporation of 2'-OH adenosines compared to 2'-F pyrimidines (Huang et al., 1997). However, at 16 hours (960 minutes) the transcription yields from templates with high A content and low A content are equivalent ( $P = 0.403$ ) (Fig. 6A). Because all the mutant T7 RNA transcriptions described in the article were performed

**FIG. 5.** SELEX of the *Sel2N20* library enriches for stable RNA structures. **(A)** Targeted selections: the predicted minimum free energy of round 0 (white circles) and rounds 2 through 8 of the targeted selections [N202.1A (Cell1), rounds 2–8; MDA-MB-231 (Cell2), rounds 4–8; Prot1, rounds 2–6; Prot2 rounds 4–8]. Line represents average predicted minimum free energy for all targeted selections. **(B)** Nontargeted selections: the minimum free energy of round 0 (white circles) and rounds 2 through 6 of the nontargeted selections. Line represents average minimum free energy for all nontargeted selections (performed at the indicated rNTP ratios). **(C)** The predicted minimum free energy was significantly ( $*P < 0.001$ ) lower for each targeted selection and **(D)** nontargeted selections as compared to round 0. Data are presented as average of rounds  $\geq 2$  for each selection.



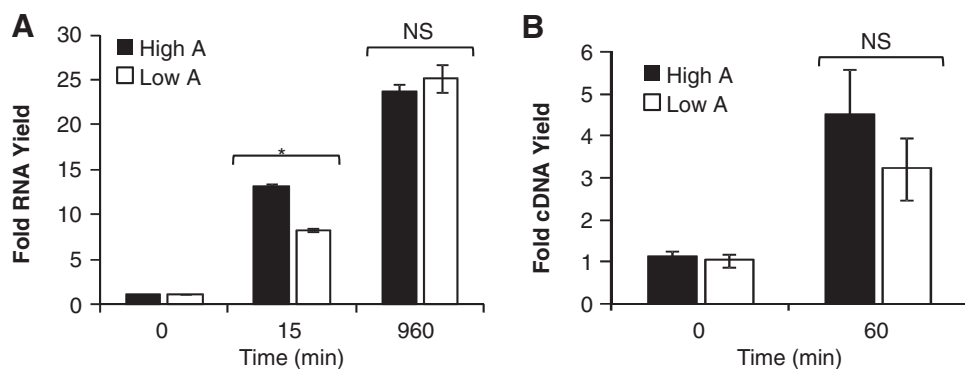
for a total of 16 hours, we do not believe that the reason behind the observed nucleotide bias is differences in the transcription rates/efficiencies of DNA templates with a high A content vs. templates with a low A content. No statistically significant difference in the transcription efficiency of similar RNA templates was observed during reverse transcription with Superscript III RT ( $P = 0.336$ ) (Fig. 6B). Together, these data suggest that the transcription efficiencies of templates

with varying adenosine content are not responsible for the observed nucleotide bias.

## Discussion

Long RNA aptamers (60–100 nucleotides long) are routinely synthesized using solid-phase phosphoramidite chemistry in an automated process used for small-scale syn-

**FIG. 6.** Transcription efficiencies do not account for the nucleotide bias. **(A)** Mutant T7 RNA polymerase (Y639F T7 RNAP) efficiency. Yield of RNA transcribed from dsDNA templates with either high adenosine (black; High A) or low adenosine (white; Low A) content at 15 and 960 minutes (16 hours) after the addition of the mutant T7 RNA polymerase enzyme. RNA yield was measured using RiboGreen (Invitrogen) after DNaseI treatment. The total yield of RNA transcribed from dsDNA templates with high adenosine content (black; High A) was significantly greater ( $*P < 0.001$ ) at 15 minutes than RNA transcribed from dsDNA templates with low adenosine content (white; Low A). In contrast, at 16 hours (960 minutes) the total RNA yield from templates with high adenosine and low adenosine content was not significantly different (NS,  $P = 0.40$ ). **(B)** Superscript III reverse transcriptase (RT) efficiency. Total yield of cDNA reverse transcribed from RNA templates with either low adenosine (white; Low A) or high adenosine (black; High A) content using Superscript III RT. The total cDNA yield from templates with high adenosine and low adenosine content was not significantly different (NS,  $P = 0.34$ ).





thesis of oligonucleotides (Reese, 2005). However, therapeutic uses require large-scale, high-quality, cGMP-grade synthesis. RNA aptamers of long length are still significantly difficult to synthesize under these conditions (Reese, 2005). One solution to this problem has been extensive truncation of RNA aptamer sequences down to minimal functional sequences (<60 nt) after selection (Biesecker et al., 1999; Rusconi et al., 2002; Dassie et al., 2009). As discussed above, this process is often time-consuming, arduous, and does not work for all aptamers. Another potential solution to this problem is the identification of shorter RNA aptamer sequences through the use of short RNA SELEX libraries.

In this study, we characterized the nucleotide composition of RNA aptamers derived from several independent selections carried out with a 51-nt-long RNA SELEX library (*Sel2N20*). The RNA aptamers isolated from this library are, on average, 20–30 nucleotides shorter than RNA aptamers generated from conventional SELEX libraries. High-throughput 454 sequencing combined with bioinformatics analysis was used to determine the nucleotide composition of the sequence pools from the various aptamer selections. This analysis revealed a bias toward pyrimidine-rich sequences (Figs. 2 and 3) as a result of loss of adenines in the RNA pools when compared with the starting library (Fig. 4). This loss was already observed after 2 selection rounds (Fig. 2) and was independent of a partitioning step (against target) since the loss also occurred when the partition step was omitted (nontargeted selection) (Fig. 3). The bias was only partially due to the 3:1 ratio of 2'-F pyrimidines to 2'-OH purines used in the transcription step as a selection with a 1:1 or 0.7:1 ratio of these nucleotides also exhibited a bias (Fig. 2B, C). The selected pools also exhibited a greater predicted structural stability (lower minimum free energy) (Fig. 5), likely a direct consequence of adenine depletion.

In accordance with our observations, several groups have postulated that functional RNAs (both artificially selected and naturally occurring RNAs) have a more stable secondary structure than random RNA sequences (Le et al., 1989; Chen et al., 1990; Clote et al., 2005). The reason for this is thought to be that functional RNAs depend on a defined secondary structure for function. Indeed, current *in silico* prediction algorithms use structural stability as evidence of a functional RNA (Bejerano et al., 2004a, 2004b; Bonnet et al., 2004; Washietl et al., 2005). Like naturally occurring functional RNAs (eg, tRNAs, rRNAs, hammerhead ribozymes and miRNAs), RNA sequences from a SELEX experiment can be considered functional RNAs and thus are predicted to exhibit higher structural stability. Hermann and colleagues have further suggested that, in general, artificially selected RNA aptamers have significantly higher predicted structural stability compared with natural occurring nucleic acids (functional RNAs) (Hermann and Patel, 2000). The reason for this difference was thought to be that selection of natural RNAs depends on multiple factors such as biological function as well as ligand binding, whereas, artificial RNA aptamers depend solely on ligand binding. RNAs with greater structural stability are postulated to be better at interlocking with their cognate target ligands. Therefore, because SELEX experiments are usually designed to isolate RNA aptamers with high affinity for their target, the selection pressure is thought to favor the isolation of RNAs with greater structural stability.

While binding to a target ligand may play a role in selecting for RNAs with overall higher structural stability it is unlikely to be the only contributing factor. Indeed, we observed that selected sequences from the nontarget selections also display more structural stability compared to the random *Sel2N20* RNA sequence library (Fig. 5B, D). Thus it is likely that factors intrinsic to the selection process (eg, mutant T7 RNA polymerase, RT-PCR, nature of the SELEX library) may also influence the structural composition of the selected RNAs. We have investigated the transcription efficiencies of synthetic oligonucleotide templates with either a low adenosine (A) content or a high adenosine (A) content (Fig. 6). Interestingly, our data suggest that the template's adenosine content does not influence the transcription efficiency of either the mutant T7 RNA polymerase or the RT enzyme and, therefore, does not dictate the observed nucleotide bias. Alternative explanations for the nucleotide bias that we have not ruled out include biases in nucleotide misincorporations by these enzymes. For example, the rate of base mispair insertion of various RT (eg, HIV RT, AMV RT, and M-MLV RT) may favor the insertion of one base over another resulting in a nucleotide bias. Indeed, the fidelity-rates of RT have been implicated as the possible reason behind the genetic variability and convergence observed with retroviral genomes (Roberts et al., 1988; Yu and Goodman, 1992).

Another possible explanation for the observed nucleotide bias is the nature of the RNA library (eg, influence of the fixed regions or an unforeseen consequence of a short variable region). To this end, Zimmermann et al. (2010) observed a shift toward less stable predicted secondary structures when performing both targeted and nontargeted SELEX with an RNA library derived from the *E. coli* genome (genome-SELEX). Unlike the *Sel2N20* library, the complexity of the genome-SELEX library is constrained by the *E. coli* genome and the length of the library is highly variable (>60 nt). The authors of the study also reasoned that factors intrinsic to the selection process (amplification steps or nature of SELEX library) may favor the propagation of certain RNA sequences over others.

While the SELEX process with the *Sel2N20* RNA library yields RNAs with overall predicted increased structural stability and ease of chemical-synthesis, a potential concern is that a nonspecific bias is influencing the selections, potentially reducing the likelihood that the optimal sequences are ultimately identified. In general, the nonspecific propagation of sequences in rounds reduces the SELEX efficiency. However, we have identified several useful aptamers with the *Sel2N20* library using similar protocols (our unpublished data), suggesting that the bias is not so severe as to result in the necessary failure of aptamer selections. Indeed, similar bias is likely to be present in successful selections described by others, but unnoticed because of the limited number of sequences that were determined. It is expected that the identification of such nonspecific influences will provide the understanding with which the SELEX process can be optimized into a more robust methodology.

## Acknowledgments

We thank Dr. Rui Sousa (University of Texas, San Antonio) for his generous gift of the mutant (Y639F) T7 RNAP. We thank Drs. Guido Forni (University of Torino, Italy) and Eli Gilboa (University of Miami) for providing cell lines for the

targeted selections. We also thank the DNA core facility (University of Iowa) for 454 sequencing and bioinformatics protocols/advice. This work was supported by funding from the NIH (1R01 CA138503-01, 1R21DE019953-01) and the Mary Kay Ash Foundation (MKACF 001-09) to P.H.G.

### Author Disclosure Statement

No competing financial interests exist.

### References

- BEJERANO, G., HAUSSLER, D., and BLANCHETTE, M. (2004a). Into the heart of darkness: large-scale clustering of human non-coding DNA. *Bioinformatics* **20**, i40–i48.
- BEJERANO, G., PHEASANT, M., MAKUNIN, I., STEPHEN, S., KENT, W.J., MATTICK, J.S., and HAUSSLER, D. (2004b). Ultraconserved elements in the human genome. *Science* **304**, 1321–1325.
- BIESECKER, G., DIHEL, L., ENNEY, K., and BENDELE, R.A. (1999). Derivation of RNA aptamer inhibitors of human complement C5. *Immunopharmacology* **42**, 219–230.
- BONNET, E., WUYTS, J., ROUZÉ P., and VAN DE PEER, Y. (2004). Evidence that microRNA precursors, unlike other non-coding RNAs, have lower folding free energies than random sequences. *Bioinformatics* **20**, 2911–2917.
- BORER, P.N., DENGLER, B., TINOCO, I., and UHLENBECK, O.C. (1974). Stability of ribonucleic acid double-stranded helices. *J. Mol. Biol.* **86**, 843–853.
- CHEN, J.H., LE, S.Y., SHAPIRO, B., CURREY, K.M., and MAIZEL, J.V. (1990). A computational procedure for assessing the significance of RNA secondary structure. *Comput. Appl. Biosci.* **6**, 7–18.
- CLOTE, P., FERRE, F., KRANAKIS E., and KRIZANC, D. (2005). Structural RNA has lower folding energy than random RNA of the same dinucleotide frequency. *RNA* **11**, 578–591.
- DASSIE, J.P., LIU, X.Y., THOMAS, G.S., WHITAKER, R.M., THIEL, K.W., STOCKDALE, K.R., MEYERHOLZ, D.K., McCAFFREY, A.P., McNAMARA, J.O., and GIANGRANDE, P.H. (2009). Systemic administration of optimized aptamer-siRNA chimeras promotes regression of PSMA-expressing tumors. *Nat. Biotech.* **27**, 839–846.
- DOLLINS, C.M., NAIR, S., BOCZKOWSKI, D., LEE, J., LAYZER, J.M., GILBOA, E., and SULLENGER, B.A. (2008). Assembling OX40 aptamers on a molecular scaffold to create a receptor-activating aptamer. *Chem. Biol.* **15**, 675–682.
- ELLINGTON, A.D., and SZOSTAK, J.W. (1990). *In vitro* selection of RNA molecules that bind specific ligands. *Nature* **346**, 818–822.
- FREIER, S.M., KIERZEK, R., JAEGER, J.A., SUGIMOTO, N., CARUTHERS, M.H., NEILSON, T., and TURNER, D.H. (1986). Improved free-energy parameters for predictions of RNA duplex stability. *Proc. Natl. Acad. Sci. U. S. A* **83**, 9373–9377.
- GIANGRANDE, P.H., ZHANG, J., TANNER, A., ECKHART, A.D., REMPEL, R.E., ANDRECHEK, E.R., LAYZER, J.M., KEYS, J.R., HAGEN, P.O., NEVINS, J.R., KOCH, W.J., and SULLENGER, B.A. (2007). Distinct roles of E2F proteins in vascular smooth muscle cell proliferation and intimal hyperplasia. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 12988–12993.
- GRAGOUDAS, E.S., ADAMIS, A.P., CUNNINGHAM, E.T., FEINSOD, M., and GUYER, D.R. (2004). Pegaptanib for neovascular age-related macular degeneration. *N. Engl. J. Med.* **351**, 2805–2816.
- HERMANN T, PATEL DJ. (2000). Adaptive recognition by nucleic acid aptamers. *Science* **287**, 820–825.
- HOFACKER, I.L., FONTANA, W., STADLER, P.F., BONHOEFFER, L.S., TACKER, M., and SCHUSTER, P. (1994). Fast folding and comparison of RNA secondary structures. *Monatsh Chem Chem Mon* **125**, 167–188.
- HUANG, Y., ECKSTEIN, F., PADILLA, R., and SOUSA, R. (1997). Mechanism of ribose 2'-group discrimination by an RNA polymerase. *Biochemistry* **36**, 8231–8242.
- LE, S.Y., CHEN, J.H., and MAIZEL, J.V. (1989). Thermodynamic stability and statistical significance of potential stem-loop structures situated at the frameshift sites of retroviruses. *Nucl. Acids Res.* **17**, 6143–6152.
- LUPOLD, S.E., HICKE, B.J., LIN, Y., and COFFEY, D.S. (2002). Identification and Characterization of nuclease-stabilized rna molecules that bind human prostate cancer cells via the prostate-specific membrane antigen. *Cancer Res.* **62**, 4029–4033.
- MATHEWS, D.H., SABINA, J., ZUKER, M., and TURNER, D.H. (1999). Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.* **288**, 911–940.
- McNAMARA, J.O., KOLONIAS, D., PASTOR, F., MITTLER, R.S., CHEN, L., GIANGRANDE, P.H., SULLENGER, B., and GILBOA, E. (2008). Multivalent 4-1BB binding aptamers costimulate CD8+ T cells and inhibit tumor growth in mice. *J. Clin. Invest.* **118**, 376–386.
- MEYERS, L.A., LEE, J.F., COWPERTHWAIT, M., and ELLINGTON, A.D. (2004). The robustness of naturally and artificially selected nucleic acid secondary structures. *J. Mol. Evol.* **58**, 681–691.
- MI, Z., GUO, H., RUSSELL, M.B., LIU, Y., SULLENGER, B.A., and KUO, P.C. (2009). RNA aptamer blockade of osteopontin inhibits growth and metastasis of MDA-MB231 breast cancer cells. *Mol. Ther.* **17**, 153–161.
- NIMJEE, S.M., RUSCONI, C.P., and SULLENGER, B.A. (2004). Aptamers: an emerging class of therapeutics. *Annu. Rev. Med.* **56**, 555–583.
- ONEY, S., NIMJEE, S.M., LAYZER, J., QUE-GEWIRTH, N., GINSBURG, D., BECKER, R.C., AREPALLY, G., and SULLENGER, B.A. (2007). Antidote-controlled platelet inhibition targeting von Willebrand factor with aptamers. *Oligonucleotides* **17**, 265–274.
- PADILLA, R., and SOUSA, R. (1999). Efficient synthesis of nucleic acids heavily modified with non-canonical ribose 2'-groups using a mutantT7 RNA polymerase (RNAP). *Nucleic Acids Res.* **27**, 1561–1563.
- REESE, C.B. (2005). Oligo- and poly-nucleotides: 50 years of chemical synthesis. *Org. Biomol. Chem.* **3**, 3851–3868.
- ROBERTS, J.D., BEBENEK, K., and KUNKEL, T.A. (1988). The accuracy of reverse transcriptase from HIV-1. *Science* **242**, 1171–1173.
- RUCKMAN, J., GREEN, L.S., BEESON, J., WAUGH, S., GILLETTE, W.L., HENNINGER, D.D., CLAESSEON WELSH, L., and JANJIC, N. (1998). 2'-Fluoropyrimidine RNA-based aptamers to the 165-amino acid form of vascular endothelial growth factor (VEGF165). Inhibition of receptor binding and VEGF-induced vascular permeability through interactions requiring the exon 7-encoded domain. *J. Biol. Chem.* **273**, 20556–20567.
- RUSCONI, C.P., SCARDINO, E., LAYZER, J., PITOC, G.A., ORTEL, T.L., MONROE, D., and SULLENGER, B.A. (2002). RNA aptamers as reversible antagonists of coagulation factor IXa. *Nature* **419**, 90–94.
- SASSANFAR, M., and SZOSTAK, J.W. (1993). An RNA motif that binds ATP. *Nature* **364**, 550–553.

- SOUSA, R., and PADILLA, R. (1995). A mutant T7 RNA polymerase as a DNA polymerase. *EMBO J.* **14**, 4609–4621.
- THIEL, K.W., and GIANGRANDE, P.H. (2009). Therapeutic applications of DNA and RNA aptamers. *Oligonucleotides* **19**, 209–222.
- TUERK, C., and GOLD, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505–510.
- WASHIETL, S., HOFACKER, I.L., LUKASSER, M., HUTTENHOFER, A., and STADLER, P.F. (2005). Mapping of conserved RNA secondary structures predicts thousands of functional noncoding RNAs in the human genome. *Nat. Biotech.* **23**, 1383–1390.
- YU, H., and GOODMAN, M.F. (1992). Comparison of HIV-1 and avian myeloblastosis virus reverse transcriptase fidelity on RNA and DNA templates. *J. Biol. Chem.* **267**, 10888–10896.
- ZHOU, J., SWIDERSKI, P., LI, H., ZHANG, J., NEFF, C.P., AKKINA, R., and ROSSI, J.J. (2009). Selection, characterization and application of new RNA HIV gp 120 aptamers for facile delivery of Dicer substrate siRNAs into HIV infected cells. *Nucleic Acids Res.* **37**, 3094–3109.
- ZIMMERMANN, B., GESELL, T., CHEN, D., LORENZ, C., and SCHROEDER, R.E. (2010). Monitoring genomic sequences during SELEX using high-throughput sequencing: neutral SELEX. *PLoS ONE* **5**, e9169.

Address correspondence to:  
*Prof. Paloma H. Giangrande*  
*Department of Internal Medicine*  
*University of Iowa*  
*285 Newton Road*  
*5202 MERF*  
*Iowa City, IA 52242*

*E-mail:* paloma-giangrande@uiowa.edu

Received for publication February 14, 2011; accepted after revision April 26, 2011.

