J|A|C|S ARTICLES

Published on Web 06/21/2010

Enhanced Functional Potential of Nucleic Acid Aptamer Libraries Patterned to Increase Secondary Structure

Karen M. Ruff, Thomas M. Snyder, and David R. Liu*

Howard Hughes Medical Institute and the Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138

Received April 16, 2010; E-mail: drliu@fas.harvard.edu

Abstract: The *in vitro* selection of nucleic acid libraries has driven the discovery of RNA and DNA receptors (aptamers) and catalysts with tailor-made functional properties. Functional nucleic acids emerging from selections have been observed to possess an unusually high degree of secondary structure. In this study, we experimentally examined the relationship between the degree of secondary structure in a nucleic acid library and its ability to yield aptamers that bind protein targets. We designed a patterned nucleic acid library (denoted R*Y*) to enhance the formation of stem-loop structures without imposing any specific sequence or secondary structural requirement. This patterned library was predicted computationally to contain a significantly higher average folding energy compared to a standard, unpatterned N_{60} library of the same length. We performed three different iterated selections for protein binding using patterned and unpatterned libraries competing in the same solution. In all three cases, the patterned R*Y* library was enriched relative to the unpatterned library over the course of the 9- to 10-round selection. Characterization of individual aptamer clones emerging from the three selections revealed that the highest affinity aptamer assayed arose from the patterned library for two protein targets, while in the third case, the highest affinity aptamers from the patterned and random libraries exhibited comparable affinity. We identified the binding motif requirements for the most active aptamers generated against two of the targets. The two binding motifs are 3.4- and 27-fold more likely to occur in the R*Y* library than in the N_{60} library. Collectively, our findings suggest that researchers performing selections for nucleic acid aptamers and catalysts should consider patterned libraries rather than commonly used N_m libraries to increase both the likelihood of isolating functional molecules and the potential activities of the resulting molecules.

Introduction

The known functional scope of nucleic acids in living systems and in the laboratory has expanded dramatically over the past 25 years. Biological RNA is now known to catalyze a variety of essential reactions including peptide bond formation, RNA splicing, and target-specific RNA cleavage.[1] In parallel with these discoveries, scientists have evolved ribozymes *in vitro* that mediate a wide variety of transformations including phosphodiester formation and cleavage,[2] carbon−carbon bond formation,[3,4] and oxidation or reduction.[5] In addition to accelerating chemical reactions, RNA evolved either in nature[6] or in the laboratory[7] has also demonstrated the ability to bind potently and selectively to a broad range of small molecules and macromolecules. Some of these aptamers exhibit therapeutically relevant properties and are in clinical trials.[8]

The abilities of nucleic acids to serve as catalysts and receptors are not limited to RNA, as DNA sequences with these capabilities have also been generated by laboratory evolution.[2,7,9,10] DNA aptamers for small organic dyes exhibit comparable affinity as RNA aptamers selected to bind the same targets.[11,12] DNA aptamers that bind with high affinity to proteins and peptides including thrombin,[13] L-selectin,[14,15] and GnRH[16] have also been evolved. Likewise, DNA has been evolved *in vitro*

(1) Strobel, S. A.; Cochrane, J. C. *Curr. Opin. Chem. Biol.* **2007**, *11*, 636–643.
(2) Joyce, G. F. *Annu. Rev. Biochem.* **2004**, *73*, 791–836.
(3) Seelig, B.; Keiper, S.; Stuhlmann, F.; Jaschke, A. *Angew. Chem., Int. Ed. Engl.* **2000**, *39*, 4576–4579.
(4) Fusz, S.; Eisenfuhr, A.; Srivatsan, S. G.; Heckel, A.; Famulok, M. *Chem. Biol.* **2005**, *12*, 941–950.
(5) Tsukiji, S.; Pattnaik, S. B.; Suga, H. *J. Am. Chem. Soc.* **2004**, *126*, 5044–5045.
(6) Tucker, B. J.; Breaker, R. R. *Curr. Opin. Struct. Biol.* **2005**, *15*, 342–348.
(7) *The Aptamer Handbook: Functional Oligonucleotides and Their Applications*; Klussmann, S., Ed.; WILEY-VCH Verlag GmbH & Co: Weinheim, 2006.
(8) Dausse, E.; Da Rocha Gomes, S.; Toulme, J. J. *Curr. Opin. Pharmacol.* **2009**, *9*, 602–607.
(9) Sen, D.; Geyer, C. R. *Curr. Opin. Chem. Biol.* **1998**, *2*, 680–687.
(10) Baum, D. A.; Silverman, S. K. *Cell. Mol. Life Sci.* **2008**, *65*, 2156–2174.
(11) Ellington, A. D.; Szostak, J. W. *Nature* **1992**, *355*, 850–852.
(12) Wilson, C.; Szostak, J. W. *Chem. Biol.* **1998**, *5*, 609–617.
(13) Bock, L. C.; Griffin, L. C.; Latham, J. A.; Vermaas, E. H.; Toole, J. J. *Nature* **1992**, *355*, 564–566.
(14) Hicke, B. J.; Watson, S. R.; Koenig, A.; Lynott, C. K.; Bargatze, R. F.; Chang, Y. F.; Ringquist, S.; Moon-McDermott, L.; Jennings, S.; Fitzwater, T.; Han, H. L.; Varki, N.; Albinana, I.; Willis, M. C.; Varki, A.; Parma, D. *J. Clin. Invest.* **1996**, *98*, 2688–2692.
(15) O'Connell, D.; Koenig, A.; Jennings, S.; Hicke, B.; Han, H. L.; Fitzwater, T.; Chang, Y. F.; Varki, N.; Parma, D.; Varki, A. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 5883–5887.
(16) Wlotzka, B.; Leva, S.; Eschgfaller, B.; Burmeister, J.; Kleinjung, F.; Kaduk, C.; Muhn, P.; Hess-Stumpp, H.; Klussmann, S. *Proc Natl Acad Sci U S A* **2002**, *99*, 8898–902.

to catalyze a variety of reactions including site-specific phosphodiester cleavage,[9] the Diels−Alder cycloaddition,[17] and light-mediated repair of thymine dimers.[18,19]

Aptamers generated by *in vitro* selection adopt a defined three-dimensional shape that allows them to bind their target with high affinity.[20] These folds contain four main secondary structural motifs: hairpin stem-loops, symmetric and asymmetric bulges flanked by helical regions, pseudoknots, and G-quartets.[21] All of these, with the exception of G-quartets, involve helical regions connected by turns and loops. These secondary structures can be linked together to create complicated globular tertiary structures such as A-minor motifs, ribose zippers, and along-groove packing motifs.[22,23]

The laboratory evolution of nucleic acids with tailor-made catalytic or binding activities begins with the preparation of a library of many different DNA or RNA sequences. In virtually all cases, the nucleic acid diversity in this library is made of a stretch of consecutive random nucleotides comprising an equimolar mixture of A, C, G, and T (or U). Since computational[24−27] and experimental[28] research has correlated increased secondary structure with increased function among evolved nucleic acid aptamers, we hypothesized that the design of a starting nucleic acid library patterned to favor the formation of stem-loops and bulges may increase the frequency or the activity of the resulting evolved nucleic acids. Consistent with this hypothesis, Szostak and co-workers previously observed that incorporating a specific stem-loop into a random RNA library as a structural nucleation point increased the functional potential of the library, as the majority of the unique sequences and total sequences in the final pool came from the stem-loop-containing library.[28]

In this study, we examined more generally the relationship between the degree of secondary structure and the functional potential of a nucleic acid library without imposing any specific sequence or secondary structural element. We designed DNA libraries with varying degrees of predicted average secondary structure by patterning pyrimidine- and purine-rich positions while maintaining the ability of each library to be prepared in a single solid-phase synthesis with no splitting and pooling required. We combined these patterned libraries with standard unpatterned libraries of the same length and followed the relative abundances of patterned versus unpatterned library members over 9 or 10 rounds of *in vitro* selection for binding to each of three different target proteins. During all three selections, the patterned library was enriched relative to the random library. The affinities of the best individual aptamers characterized from the patterned library for their target proteins were higher (in two cases) or as high (in one case) as those from the random library. Our findings suggest that libraries of DNA or RNA patterned to enhance secondary structure may represent more promising starting points for *in vitro* evolution than commonly used unpatterned libraries.

Results

Library Design. First, we designed nucleic acid libraries with higher average secondary structure that do not require any specific sequence or secondary structural element. Inspired by the alternation of hydrophobic and hydrophilic amino acids to form β-strands in peptides and proteins,[29] we sought a pattern of alternating nucleobases that would increase the propensity of a nucleic acid to form regions rich in base pairing. One possible pattern to increase pairing alternates 1:1 mixtures of C/G (S) with A/T (W) so that potential Watson−Crick partners can be aligned when two such regions interact. In this case, any S:S or W:W pair would have a 0.5 probability of forming a base pair. Therefore, the probability of forming a 6-bp stem between two SWSWSW regions would be $(0.5)^6 = 0.016$ if the strands are in alignment, corresponding to an overall probability of 0.008. This likelihood is higher than the probability of forming a 6-bp stem between two NNNNNN regions (where N is a equimolar mixture of A, C, G, and T), which is $(0.375)^6 = 0.0028$ if one considers A−T, C−G, and G−T as viable base pairs. A 6-bp stem is thus ~3-fold more likely to form in a double-stranded SW patterned region than in a standard unpatterned N region.

A second possible pattern alternates purines (R) and pyrimidines (Y). In this case, any R:Y pair would have a $3 \times (0.5 \times 0.5) = 0.75$ probability of forming a Waston−Crick or wobble base pair. Therefore, the probability of forming a 6-bp stem between two RYRYRY regions would be $(0.75)^6 = 0.18$ if the strands are in alignment, corresponding to an overall probability of 0.09. A 6-bp stem in a double-stranded RY-patterned region would therefore be ~32-fold more likely to form than in a standard N library, and ~11-fold more likely than in a SW patterned region. On the basis of these and other related calculations (see the Supporting Information), a RY-based pattern was chosen to increase the likelihood of secondary structure formation in our patterned library. We chose a 60-base variable region based on work by Knight, Yarus, and co-workers addressing optimal library length for small molecule-binding aptamers[30] and other considerations (see the Supporting Information for a full discussion).

While a library containing strictly alternating purines and pyrimidines is predicted to have large average folding energies and thus a high degree of average secondary structure, we speculated that a library entirely of RY repeats would not be ideal for aptamer or catalyst function because it would not easily accommodate loops or more complex structural motifs. Therefore, we designed two additional features to increase the frequency and diversity of loops and other nonstem structures within the patterned library. First, we incorporated short intervening stretches of three or four consecutive N nucleotides between RY regions. We envisioned that these N regions could participate in the formation of a wide variety of target-binding loops, bulges, or other structures and would be stabilized by flanking RY base pairs that have an enhanced tendency to form a stem. The resulting library, designated the "RY library", has

(17) Chandra, M.; Silverman, S. K. *J. Am. Chem. Soc.* **2008**, *130*, 2936–2937.
(18) Chinnapen, D. J.; Sen, D. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 65–69.
(19) Thorne, R. E.; Chinnapen, D. J.; Sekhon, G. S.; Sen, D. *J. Mol. Biol.* **2009**, *388*, 21–29.
(20) Mayer, G. *Angew. Chem., Int. Ed.* **2009**, *48*, 2672–2689.
(21) Eaton, B. E.; Gold, L.; Zichi, D. A. *Chem. Biol.* **1995**, *2*, 633–638.
(22) Noller, H. F. *Science* **2005**, *309*, 1508–1514.
(23) Gagnon, M. G.; Steinberg, S. V. *RNA* **2010**, *16*, 375–381.
(24) Carothers, J. M.; Oestreich, S. C.; Davis, J. H.; Szostak, J. W. *J. Am. Chem. Soc.* **2004**, *126*, 5130–5137.
(25) Carothers, J. M.; Oestreich, S. C.; Szostak, J. W. *J. Am. Chem. Soc.* **2006**, *128*, 7929–7937.
(26) Kim, N.; Gan, H. H.; Schlick, T. *RNA* **2007**, *13*, 478–492.
(27) Chushak, Y.; Stone, M. O. *Nucleic Acids Res.* **2009**, *37*, e87.
(28) Davis, J. H.; Szostak, J. W. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 11616–11621.
(29) Xiong, H.; Buckwalter, B. L.; Shieh, H. M.; Hecht, M. H. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 6349–6353.
(30) Legiewicz, M.; Lozupone, C.; Knight, R.; Yarus, M. *RNA* **2005**, *11*, 1701–1709.

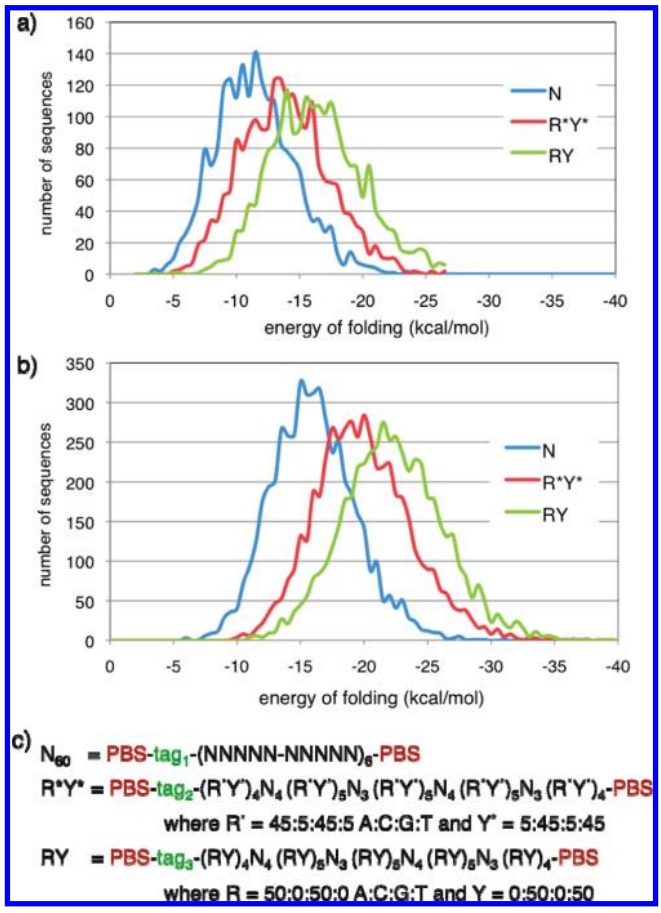

Figure 1. Distributions of predicted folding energies for two patterned libraries and a standard N_{60} library (a) without primers, and (b) with primers used in the IgE and VEGF selections. (c) The DNA sequences of each of the three libraries is shown.

a 60-base variable region of the form $(RY)_4N_4(RY)_5N_3$-$(RY)_5N_4(RY)_5N_3(RY)_4$, where R = 50:50 A/G and Y = 50:50 C/T.

Second, certain positions in a nucleic acid receptor or catalyst might require a specific base not available in a purely R or Y patterned position. We therefore envisioned that adding a small fraction of the other bases into the R or Y positions would help balance the benefits of enhanced secondary structure with the drawbacks of excessively constrained base composition. This library, designated the "R*Y* library", has the same form as the RY library, but uses an R* and Y* nucleotide mixture of 45:5:45:5 A/C/G/T and 5:45:5:45 A/C/G/T, respectively. The final form of the 60-base variable region of R*Y* is $(R^*Y^*)_4$-$N_4(R^*Y^*)_5N_3(R^*Y^*)_5N_4(R^*Y^*)_5N_3(R^*Y^*)_4$, where R* = 45:5:45:5 A/C/G/T and Y* = 5:45:5:45 A/C/G/T.

We characterized computationally the average predicted folding energy of 5000 randomly chosen members of the N_{60}, RY, and R*Y* libraries using the Oligonucleotide Modeling Platform (OMP, DNA Software) (Figure 1a). The average predicted folding energy of sequences in the RY library is almost two standard deviations higher in magnitude than that of N_{60} ($\Delta\Delta G = -4.9$ kcal/mol). R*Y* library members exhibited an intermediate degree of predicted structure ($\Delta\Delta G = -2.4$ kcal/mol relative to N_{60}), consistent with the expected tradeoff between the degree of secondary structure enhancement and motif flexibility. These results suggest that the use of patterns containing alternating purine-rich and pyrimidine-rich positions can significantly increase the average secondary structure of a nucleic acid library relative to that of a standard unpatterned library of the same length.

Primer and Tag Design. To rigorously compare the fitness of the libraries during *in vitro* selection, we combined patterned and unpatterned libraries into a single solution prior to the start of the selection process to avoid any differences in the way the libraries were treated during selections. This method of competitive selection[28] allows the direct comparison of libraries by following their relative abundances over rounds of selection. Because a competitive selection requires that the libraries be amplifiable in a single solution, they must share common PCR primer-binding sites. In addition, in order to ascertain the origin of individual clones surviving selection, we included a short 6-nucleotide library-specific tagging sequence in each library that was recognized by a different restriction endonuclease. Since these constant regions can interact with the variable region of a library member and affect its structure, we examined the effects of prospective constant sequences on the predicted folding energies of all three libraries using OMP (see the Supporting Information for details). Primer-binding site and tag sequences were chosen such that the average folding energy among N_{60} and RY sequences remained separated by ~2 standard deviations ($\Delta\Delta G = -6.3$ kcal/mol), and N_{60} and R*Y* remained separated by ~1 standard deviation ($\Delta\Delta G = -3.9$ kcal/mol) (Figure 1b). The final primer-binding and tag sequences are provided in Materials and Methods.

Library Synthesis and Characterization. We created mixtures of phosphoramidites in the target ratios described above for N, R, Y, R*, and Y* and measured the resulting ratio of nucleotides using a previously described HPLC assay (see the Supporting Information for details).[31] Consistent with previous reports describing differences in coupling efficiency among different phosphoramidites,[32] we observed product ratios that differed from the expected ratios and adjusted phosphoramidite stoichiometries accordingly. For example, G coupled at a significantly higher rate than the other bases in the initial N mixture for the N_{60} library, such that a 1:1:1:1 phosphoramidite mixture resulted in 36% G (rather than 25%) by HPLC assay. After adjusting phosphoramidite stoichiometries to 30:28:21:21 A/C/G/T, we achieved coupling of the four bases approximately equally (25 ± 2%). Similar empirical optimization led to phosphoramidite stoichiometries for R (42:58 A/G), Y (52:48 C/T), R* (55:3:40:2 A/C/G/T), and Y* (4:52:2:41 A/C/G/T). Note that day-to-day differences in humidity and reagent quality led to differences in product ratios from phosphoramidite mixtures prepared to contain identical molar ratios.

Using these phosphoramidite mixtures, we synthesized the N_{60}, RY, and R*Y* DNA libraries with the variable and constant regions described above, in the form 5′-(primer binding site 1)-tag-(60-base variable region)-(primer binding site 2)-3′. The libraries were each 100 nucleotides in total length. They were purified by reverse-phase cartridge followed by preparative denaturing PAGE, quantified by UV spectroscopy, and combined into a single solution. The ratio of the three libraries in the mixture was evaluated by digestion of the PCR-amplified mixture with library tag-specific restriction endonucleases, and the mix was adjusted such that ~30% of the pool was digested by each restriction enzyme. In addition, 89 individual clones from the starting mixture of N_{60}, RY, and R*Y* libraries were sequenced. The ratio of N_{60}/RY/R*Y* was 31:30:28 among these 89 clones.

(31) Ward, B.; Juehne, T. *Nucleic Acids Res.* **1998**, *26*, 879–886.
(32) Chapple, K. E.; Bartel, D. P.; Unrau, P. J. *RNA* **2003**, *9*, 1208–1220.

Table 1. Observed Nucleotide Ratios for IgE and VEGF Libraries by DNA Sequencing (Average ± Standard Error)

library	mix	A	C	G	T
N_{60}	N	27.2 ± 1.0%	23.4 ± 1.0%	26.6 ± 1.0%	22.8 ± 1.1%
R*Y*	N	30.9 ± 2.3%	23.0 ± 2.1%	18.1 ± 2.5%	28.1 ± 2.8%
	R*	45.8 ± 2.1%	3.5 ± 0.7%	47.9 ± 2.2%	2.8 ± 0.6%
	Y*	2.7 ± 0.5%	48.8 ± 2.1%	3.0 ± 0.8%	45.6 ± 1.9%
RY	N	27.5 ± 2.3%	18.5 ± 1.6%	28.2 ± 2.1%	25.8 ± 1.8%
	R	55.8 ± 1.5%	0 ± 0.0%	44.0 ± 1.5%	0.1 ± 0.0%
	Y	0.2 ± 0.0%	44.5 ± 2.0%	0.3 ± 0.0%	55.0 ± 2.0%

The observed nucleotide abundances of 58 clones from the RY or R*Y* libraries were also used to evaluate the ratios of the bases at each position in the patterned libraries (Table 1). The folding energy of 3000 randomly chosen sequences with these experimentally observed ratios were calculated (see Supporting Information Table S2). The average predicted folding energies for the N_{60} and R*Y* libraries were very close to the designed values, and remained separated by one standard deviation. In contrast, the RY library with experimentally observed nucleotide abundances had a predicted folding energy with a significantly smaller magnitude than the designed RY library ($\Delta\Delta G = 2$ kcal/mol), such that its average folding energy was the same as that of the R*Y* library. Because the experimental RY library is predicted to have the same average energy of folding and standard deviation as the R*Y* library, these libraries can be more directly compared to determine the importance of incorporating a small fraction of off-pattern bases at each position in the pattern.

Finally, we measured the efficiency with which each library is amplified by PCR. We performed quantitative PCR (QPCR) on varying amounts of each library (1 amol to 1 fmol) using their common primers. For each starting concentration tested, all three libraries exhibited similar C_T values (Supporting Information Figure S3a). Next, we explicitly tested if the uniform ability of the libraries to be amplified during PCR would apply in a single-solution mixture containing all three libraries. We carried an equimolar mixture of all three libraries through a mock selection consisting of 10 iterated rounds of 128-fold dilution followed by PCR amplification. The ratio of N_{60}, RY, and R*Y* libraries did not significantly change during this process (Supporting Information Figure S3b). Taken together, these experiments suggest that any enrichment observed during selections results from differences in the fitness of library members, rather than from differences in their amplification efficiencies during PCR.

We note that the careful measures described here to ascertain library composition and behavior were implemented to maximize the rigorousness and validity of the library comparisons in this study. For routine nucleic acid library preparation and selection experiments, we anticipate that using DNA synthesizer-programmed mixtures or commercially available premixed phosphoramidites and choosing constant sequences with only routine attention to primer design will suffice to create patterned nucleic acid libraries with the properties described in this work.

Three *In Vitro* Selections for Target Protein Binding. Mixtures of patterned and N_{60} libraries were used in three parallel selections for target protein affinity to compare the functional potential of the N_{60}, RY, and R*Y* libraries. The first selection compared N_{60} and R*Y* over 10 rounds of selection for binding to streptavidin. For each round of selection, we mixed the surviving DNA from the previous round (or the starting library for the first round) with streptavidin-linked agarose beads,

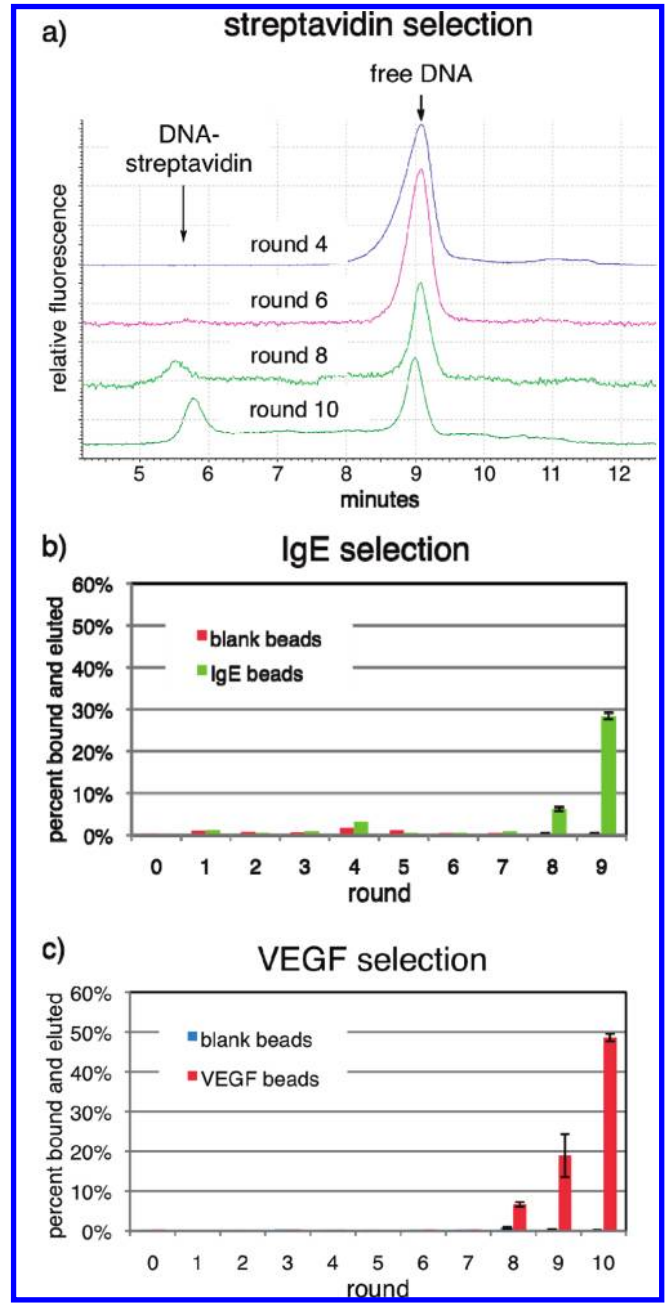

Figure 2. Binding activity of pools over 9 to 10 rounds of selections for binding (a) streptavidin, (b) IgE, and (c) VEGF.

washed with buffer, and eluted with excess free streptavidin. The eluted library members were amplified by PCR and the 5′-phosphorylated complementary (noncoding) strand was digested with λ-exonuclease to regenerate the single-stranded pool. A fraction of the pool was analyzed for streptavidin binding by capillary electrophoresis after every two rounds. Activity was detected after eight rounds and ~50% of the pool bound streptavidin after 10 rounds (Figure 2a).

We performed two additional competitive selections for binding to immunoglobulin E (IgE) and vascular endothelial growth factor (VEGF). For both selections, a ~1:1:1 starting mixture of N_{60}/RY/R*Y* was subjected to 9 or 10 rounds of binding to agarose beads conjugated to IgE or VEGF, followed by elution of bound library members with 0.1 M NaOH and 10 mM EDTA. Following round 3, round 5, and all subsequent rounds, a negative selection to remove any library members with

Table 2. Library Enrichment after Each Selection

	N_{60}:R*Y* ratio from DNA sequencing			N_{60}:R*Y* ratio from restriction digestion		
selection	input	final	enrichment	input	final	enrichment
streptavidin	1.8:1 ($n = 36$)	1:3.2 ($n = 75$)	6-fold	1:1.4	1:2.5[a]	2-fold[a]
IgE	1.1:1 ($n = 59$)	1:8.8 ($n = 78$)	10-fold	1:1.2	1:29	25-fold
VEGF	2.0:1 ($n = 67$)	1.6:1 ($n = 90$)	1.3-fold	2.1:1	-[b]	-[b]

[a] The round 10 pool for the streptavidin selection contained a significant fraction (~25%) of DNA not cleaved by treatment with restriction endonucleases. [b] -: Because of a mutation in the R*Y*-specific tag sequence among a significant fraction of the pools surviving round 9 and round 10, restriction enzyme digestion could not be used to assess the final N_{60}:R*Y* ratio for the VEGF selection.

bead-binding, rather than target protein-binding, activity was performed by incubation with agarose beads lacking any target protein. A small fraction of the pool was radiolabeled by 5′ phosphorylation with γ-^{32}P-ATP and assayed for bead binding after each round. For both the IgE and VEGF selections, binding activity was detectable after eight rounds, and significant after nine or 10 rounds (Figure 2b,c).

Selection Results. The relative abundances of patterned and unpatterned libraries were evaluated by DNA sequencing of 75−90 clones surviving the final round of each selection. Restriction digestion of library-specific tag sequences was also performed after each round (Supporting Information Figure S5), but for both the streptavidin and VEGF selections, the presence of significant uncut DNA especially in later rounds (representing up to ~30% of either selection's pool) obscured interpretation. We attribute the uncut DNA to mutations in tag sequences, as was explicitly observed in the case of R*Y*-derived clones in the VEGF selection, or to inefficient digestion of particularly well-folded DNA sequences. The initial and final ratios of N_{60} to R*Y* for each selection based on DNA sequencing and on restriction enzyme digestion are summarized in Table 2.

For all three selections, the patterned R*Y* library was enriched over the course of the selection relative to the unpatterned N_{60} library (Table 2), suggesting that the R*Y* library contained a higher fraction of active sequences, or sequences with greater target binding potency, than the standard unpatterned library. For the streptavidin selection, out of the 75 clones sequenced from the round 10 library material, 57 clones (29 unique) came from R*Y* and 18 clones (eight unique) arose from N_{60}, corresponding to a final N_{60}:R*Y* ratio of 1:3.2. Given the starting ratio of 1.8:1 favoring N_{60}, this outcome represents a 6-fold enrichment of the patterned R*Y* library members relative to the unpatterned sequences in the final pool.

During the selection for IgE binding, R*Y* library members were also enriched significantly. DNA sequencing revealed an N_{60}/R*Y*/RY starting ratio of 31:28:30, while the ratio in the round 9 surviving DNA was 8:70:4 N_{60}/R*Y*/RY, representing a ~10-fold enrichment of the patterned R*Y* library over the random, unpatterned library.

Members of both the N_{60} and R*Y* libraries were effectively maintained through round 10 of the VEGF binding selection. Among the 90 sequences obtained from DNA surviving round 10, the ratio of N_{60}/R*Y* was 55:35. Given a starting ratio of N_{60}/R*Y*/RY of 45:22:26 based on DNA sequencing, R*Y* was maintained at least as well as N_{60} (1.3-fold enrichment of R*Y* relative to N_{60}) over the course of the VEGF selection.

In both cases in which the RY library competed with the N_{60} and R*Y* libraries (the IgE and VEGF selections), the RY library decreased in abundance dramatically by round 8, the first round in which either selection pool exhibited significant target protein-binding activity. For the IgE selection, DNA sequencing of 82 clones from the round 9 surviving pool revealed only four sequences (5%) from the RY library. For the VEGF selection, none of the 62 sequenced clones surviving round 9 and none of the 90 clones surviving round 10 were RY library members. These results indicate that the RY library competed poorly with the N_{60} or R*Y* libraries, and are consistent with our hypothesis that a pure RY patterned library does not provide sufficient sequence flexibility to support common secondary structural motifs in aptamers.

Taken together, these findings indicate that, for the three independent competitive selections performed here, the R*Y* patterned library substantially outperformed the unpatterned library in two cases and performed at least as well as the unpatterned library in the third case as measured by library enrichment over the course of these selections.

Analysis of Streptavidin Aptamers. The predicted secondary structures of the 37 unique sequences from the pool surviving round 10 of the streptavidin selection were analyzed computationally using OMP (see the Supporting Information for sequences of individual clones from all three selections). The predicted structures were inspected to identify common motifs. A fixed six-base GNYGCA loop connected by a four-base variable stem to a 5′ CGC bulge, all flanked by a longer variable stem-forming region, were found among 62% of the unique clones and 75% of the total clones of the 75 streptavidin sequences obtained from round 10. Among the 37 unique sequences, 13 contained this set of secondary structure motifs in the lowest energy fold predicted by OMP, and 10 more sequences are also predicted to access this motif, although not in the fold predicted to be lowest in energy.

Additional sequences emerging from the streptavidin selection were found to be related to the above set of consensus binding motifs. Three sequences from the N_{60} library contained the same motif but with a TGC instead of CGC bulge. One sequence from the R*Y* library, observed twice, contained a GTCACA loop instead of a GNYGCA loop but otherwise fit the consensus binding motifs. Two more sequences, one each from the N_{60} and R*Y* libraries, share the consensus hexanucleotide loop but contain an expansion of the secondary bulge to TYGCW. A third sequence, from the R*Y* library, shares the same hexanucleotide loop and expanded bulge, and also a lengthened five-base pair internal stem. Finally, another variant shortens the internal stem by one position, while also expanding the loop to seven positions. In summary, a substantial majority of surviving streptavidin selection clones are predicted to adopt or at least access this common set of structural motifs. Among the clones from this family, 24 (77%) of the unique sequences and 52 (76%) of the total sequences were from the R*Y* library. This family of streptavidin-binding motifs were also independently identified by Bing and co-workers based on an analysis of previously published streptavidin aptamers.[33]

(33) Bing, T.; Yang, X.; Mei, H.; Cao, Z.; Shangguan, D. *Bioorg. Med. Chem.* **2010**, *18*, 1798–1805.

Table 3. Binding Activities of Streptavidin Aptamers[a]

sequence	percent of round 10 pool	library	motif	percent bound at 1 μM streptavidin
S10-101	5%	R*Y*	standard	>70%
S10-103	11%	N_{60}	standard	48%
S10-104	3%	R*Y*	standard	53%
S10-113	4%	N_{60}	heptaloop	50%
S10-115	1%	N_{60}	5-base bulge	60%
NRR			heptaloop	33%

[a] NRR refers to the streptavidin-binding aptamer previously evolved by nonhomologous random recombination.[34]

After identifying the probable binding motif, we synthesized minimized forms of five sequences from among the round 10 clones, including several with the standard bulge/loop sizes and also one each with an expanded bulge and expanded loop. Each minimized version contained the major bulge-stem-loop as well as the flanking stem (see the Supporting Information for sequences). For comparison, we also synthesized a previously described DNA aptamer to streptavidin evolved by nonhomologous random recombination[34] (K_d = 105 nM for the minimized 40-mer by nitrocellulose filter-binding assay) to provide a reference point for analysis. We compared the relative streptavidin binding activities of these aptamers by capillary electrophoresis.[35] All of the new sequences bound to streptavidin to a greater extent than the known aptamer (Table 3). While many of the clones bound with similar affinities, the most potent binder was clone S10-101 from the R*Y* library. In addition, the majority of the round 10 pool survivors and the majority of distinct binding sequences were from the R*Y* library. These observations support a model in which the patterned library both contained a greater fraction of active binders than the unpatterned library and also gave rise to the most potent aptamers.

Analysis of IgE Aptamers. The IgE selection resulted in a single predominant clone (I9-102) which arose from the R*Y* patterned library. In the DNA sequences acquired following round 9, clone I9-102 appeared 60 times while the other 22 sequences were unique and distributed in a 8:10:4 ratio of N_{60}/R*Y*/RY libraries (see the Supporting Information for sequences). Clone I9-102 was amplified by PCR with a 5′ biotinylated noncoding strand primer. After immobilization of the PCR product with streptavidin-linked magnetic beads, the coding strand was selectively eluted with 0.1 M NaOH, conditions that prevent DNA hybridization but preserve streptavidin−biotin complex formation,[13] and assayed for IgE binding. Clone I9-102 bound immobilized IgE to a similar extent as the round 9 pool (~40% bound and eluted). IgE affinity was also assayed by nitrocellulose filter binding at varying concentrations of IgE, resulting in an apparent post-round 9 pool affinity of K_d = 30 nM, and a clone I9-102 affinity of K_d = 26 nM.

The 22 other sequences isolated in the selection were similarly analyzed for IgE binding (Table 4 and Figure S6). Of these, only three showed ≥1% binding to IgE-linked beads under selection conditions. The best of these three secondary clones, clone I9-202 (from the unpatterned library), bound immobilized IgE at 2% and was further evaluated by filter binding. At the highest accessible IgE concentration of 2 μM, only ~5% of clone I9-202 bound. Therefore, clone I9-102 binds IgE ≥100-fold better than clone I9-202 or any other isolated IgE aptamer. Neither clone I9-102 nor clone I9-202 exhibited binding affinity for BSA (Table 4), suggesting that these aptamers are not nonspecific protein binders. These results indicate that the patterned R*Y* library accounted for a much greater total percentage of the final pool and yielded more potent IgE aptamers than the unpatterned N_{60} library.

Analysis of VEGF Aptamers. The DNA surviving the VEGF binding selection after rounds 9 and 10 was sequenced (n = 62 clones for round 9 and n = 90 clones for round 10) and found to contain many different sequences, several of which were observed in high frequency (see the Supporting Information for sequences). All of the sequences observed more than once in round 9, plus one unique sequence from round 10, were amplified by PCR, strand separated as described above, and radiolabeled by 5′ phosphorylation for analysis. All of the assayed clones bound significantly better to immobilized VEGF than the unselected starting library (Table 5 and Figure S8). The four clones that exhibited the strongest immobilized VEGF binding activities, which included all of the most common sequences, were further characterized by nitrocellulose filter binding assay using solution-phase VEGF (Table 5). The average predicted folding energy for these eight sequences was −19.5 kcal/mol.

The three most common clones from rounds 9 and 10 were clone V9-103 (from R*Y*, representing 40% and 38% of the pool after rounds 9 and 10, respectively), clone V9-105 (from N_{60}, representing 16% and 13% of the pool after rounds 9 and 10), and clone V9-101 (from N_{60}, representing 13% and 28% of the pool after rounds 9 and 10). These aptamers were assayed by filter binding and all three exhibited comparable VEGF binding activity (K_d = ~30−50 nM). The average predicted folding energy of these three sequences was −21.2 kcal/mol, which is greater than all but a small fraction (8%) of the N_{60} library members but which is accessible by a substantial fraction (36%) of the R*Y* library members (see Discussion below). Similar to the case of the IgE aptamers, neither clone V9-103 nor clone V9-105 exhibited binding affinity for BSA (Table 5), suggesting that these aptamers do not bind proteins nonspecifically.

In summary, the most common clone as revealed by DNA sequencing was from the R*Y* library, and the R*Y* library was modestly enriched over rounds of selection relative to the N_{60} library. In contrast with the two other selections in this work, the most active characterized aptamers from the patterned R*Y* library and the unpatterned library bound VEGF with similar affinity.

Probability of Occurrence of Aptamers in the Patterned and Unpatterned Libraries. Finally, we sought to determine the probability of occurrence in the patterned and unpatterned libraries of the most active characterized aptamers. The calculation of these probabilities first required elucidating a consensus binding motif for each aptamer of interest.

For the streptavidin aptamer, we created a consensus 29-base sequence based on 27 sequences related to clone S10-101 (Figure 3). This consensus 29-mer bound streptavidin at a 52% level as assayed by CE. To dissect the requirements for binding, 19 site-directed mutants of this consensus sequence were tested for streptavidin binding (Figure 3). On the basis of the results, the binding motif was inferred as XXXXXX-YGC-XXXX-GNYGCA-XXXX-XXXXXX, where X_4 and X_6 represent nucleotides that participate in base pairing to form two stems (with either Watson−Crick or G:T wobble pairing). We calculated the probability of this motif occurring in either the patterned R*Y* library or in the unpatterned N_{60} library in any

(34) Bittker, J. A.; Le, B. V.; Liu, D. R. *Nat. Biotechnol.* **2002**, *20*, 1024–1029.
(35) Mosing, R. K.; Bowser, M. T. *Methods Mol. Biol.* **2009**, *535*, 33–43.

Table 4. Binding Activities of IgE Aptamers against Target (IgE) and Nonspecific Binding Control (BSA)

sequence	percent of round 9	library	% bound to immobilized IgE (± standard error)	K_d from filter binding (± standard deviation)	% bound to immobilized BSA	predicted ΔG of folding (kcal/mol)
I9-102	73%	R*Y*	39.9 ± 3.4%	26 ± 5nM	0.1%	−18.4
I9-202	1%	N_{60}	2.0 ± 0.1%	>2 uM	0.1%	−21.4
I9-216	1%	R*Y*	1.0 ± 0.1%	n.d.[a]	n.d.	−17.8
I9-110	1%	RY	0.9 ± 0.2%	n.d.	n.d.	−22.4
starting library		mix	0.4 ± 0.1%	n.d.	0.2%	

[a] n.d. = not determined.

Table 5. Binding Activities of VEGF Aptamers against Target (VEGF) and Nonspecific Binding Control (BSA)

clone	% of pool after round 9	% of pool after round 10	library	% bound to immobilized VEGF	K_d from filter binding (± st. deviation)	% bound to immobilized BSA	predicted ΔG of folding (kcal/mol)
V9-101	13%	28%	N_{60}	27.2%	52 ± 15 nM	n.d.[a]	−22.8
V9-103	40%	38%	R*Y*	23.3%	45 ± 13nM	0.1%	−19.5
V9-104	3%	0%	N_{60}	15.0%	n.d.	n.d.	−22.8
V9-105	16%	13%	N_{60}	56.2%	28 ± 7 nM	0.1%	−21.4
V9-114	2%	1%	R*Y*	10.9%	n.d.	n.d.	−20.8
V9-204	5%	4%	N_{60}	29.7%	150 ± 56 nM	n.d.	−13.6
V9-215	6%	1%	N_{60}	13.8%	n.d.	n.d.	−16.0
V10-122	0%	1%	N_{60}	13.9%	n.d.	n.d.	−18.8
starting library			mix	0.3%	n.d.	0.2%	

[a] n.d. = not determined.

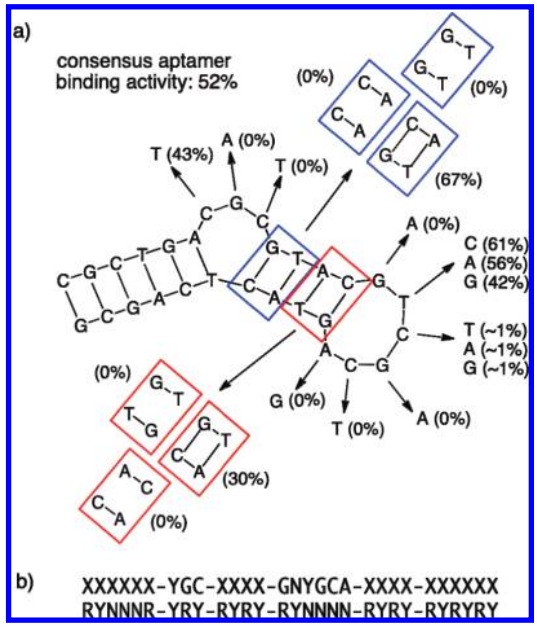

Figure 3. Streptavidin binding motif analysis. (a) The binding activities of point mutants (listed in parentheses) reveal a consensus set of binding motifs. (b) The most likely frame for this consensus sequence in the R*Y* library is shown.

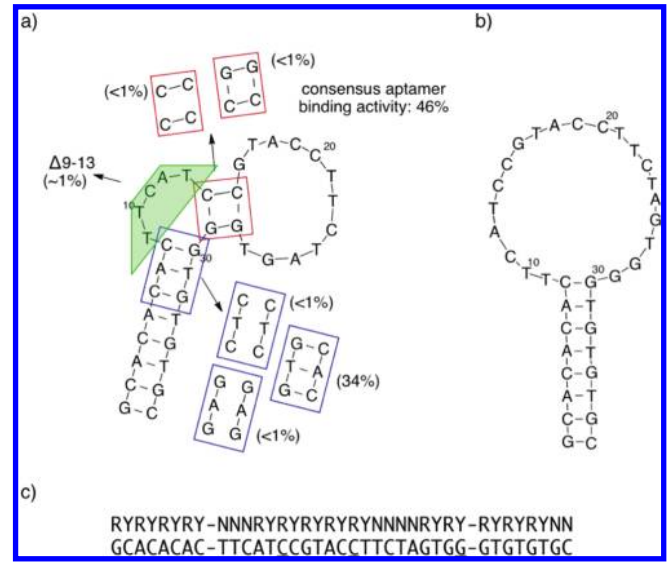

Figure 4. IgE binding motif analysis. (a) An initial predicted fold and mutational analysis of IgE clone I9−102 is shown; binding activities appear in parentheses. (b) A revised secondary structure is shown, reflecting the lack of apparent base pairing between the nucleotides boxed in red in panel a. (c) The most likely frame for the consensus sequence in the R*Y* library is shown; underlined bases indicate off-pattern positions.

of each library's 32 frames that are capable of containing a 29-base motif. These calculations reveal that the consensus streptavidin binding motif is 27-fold more likely to occur in the R*Y* library than in the unpatterned library (see Supporting Information for the complete analysis). The increased occurrence of this binding motif in the R*Y* library arises from the stems being 15- to 120-fold more likely and the loop being up to 10-fold more likely to occur in the patterned library than the unpatterned library, counterbalanced by the motif being possible in any of the 32 frames of the N_{60} library but likely in only five frames of the R*Y* library. Importantly, the consensus streptavidin-binding motif does not occur in a purely patterned RY library because every frame in the pattern requires at least one off-pattern bulge/loop position or R−R pairing position.

Similarly, we generated a series of truncation mutants of IgE aptamer clone I9-102 to identify a 55-base binding region containing a long stem plus a large loop (see the Supporting Information and Figure S9). All activity was lost when the 21-base loop was truncated to a 5-base loop (see the Supporting Information). In addition, a minimized version with an 8-base stem (no bulges) and the intact loop bound comparably well to immobilized IgE as the full-length clone (Figure 4a). We generated and assayed a series of site-directed mutants to further probe the proposed secondary structure model. The predicted stem was confirmed by observing a loss of binding upon mutation of bases 6−8 that could be rescued by compensatory mutations in bases 30−32 to restore base pairing (Figure 4a). Consistent with the design principles of the R*Y* library, the majority of this required stem region falls within the R*Y*

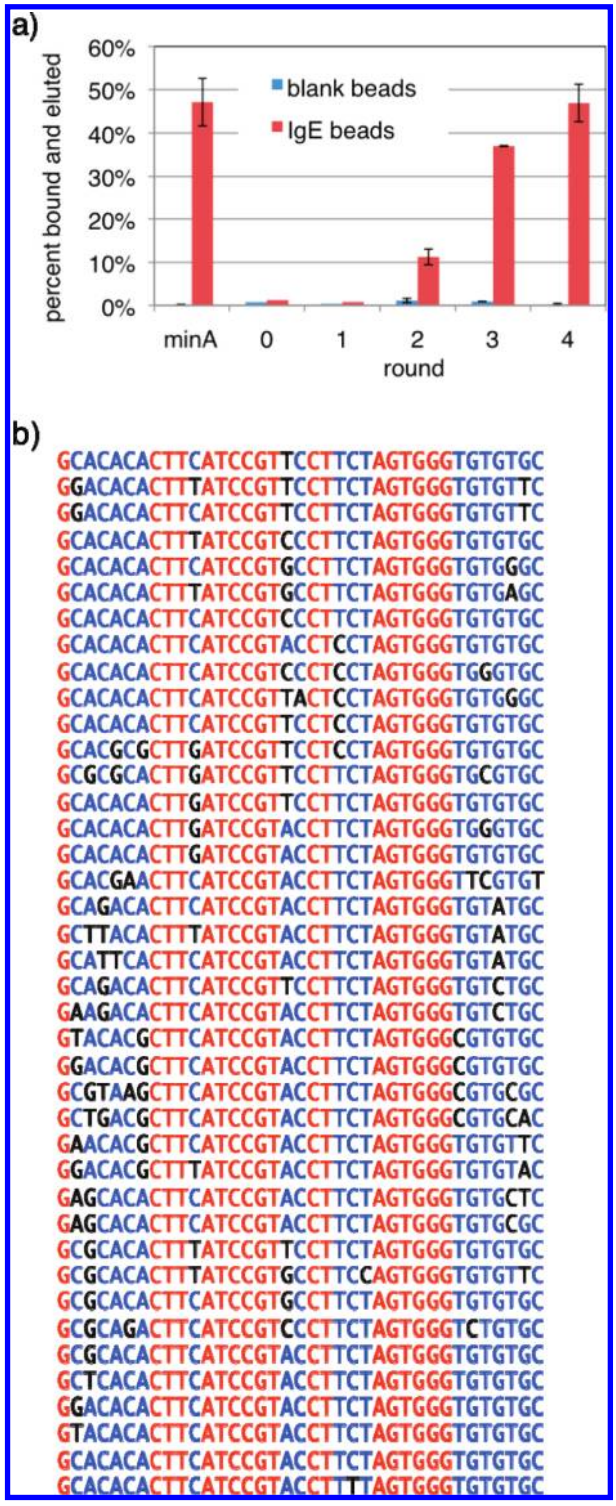

Figure 5. Reselection of a library based on the I9−102 IgE aptamer. (a) Progression of the selection as monitored by bead-binding activity. (b) Unique sequences surviving round 4; red bases are totally conserved, blue bases are partially conserved, and black bases are mutations relative to 102 min.

patterned region of the library, while the loop spans an entire $N_3 + (R^*Y^*)_5 + N_4$ region (Figure 4c).

While this analysis identified the location of the binding motif and confirmed the presence of the stem within the IgE-binding motif, the absence of related sequences to guide our analysis precluded the determination by inspection of which positions in the binding loop were required to be a specific nucleotide and which could accommodate other bases. Our attempts to truncate or mutate the 21-base loop resulted in complete loss of activity (Figure 4a), leading to a revision of the predicted structure (Figure 4b). To further analyze the binding motif, we created a library based on the minimized I9-102 in which each position in the library was a 4:1 mixture of the base observed in I9-102/the other three bases (in equal stoichiometry).[36,37] Only ~1% of this secondary library of minimized I9-102 variants bound immobilized IgE. We performed four rounds of IgE binding selection on this library as described above. After four rounds of selection, the surviving secondary library subpopulation exhibited activity equivalent to that of the minimized I9-102 clone (Figure 5a).

The binding motif in sequences surviving this secondary selection was highly conserved (Figure 5b). The putative 8-base pair stem exhibited covariation at all 13 positions in which mutations were observed. Of the 42 unique clones, all eight base pairs in this stem were preserved in 25 sequences, and seven of the eight base pairs were preserved in another 16 clones. These results strongly suggest that the hypothesized stem is required for IgE binding activity (see the Supporting Information for a detailed analysis).

In addition, of the 21 bases in the large loop, only three exhibited any variation among the sequences surviving the secondary selection. The other 18 bases showed little or no tolerance for mutation. These nonmutable positions included three 'off-pattern' (non-R or non-Y) positions (underlined in Figure 4c); therefore, this binding motif is not found in the pure RY library, consistent with the gradual loss of the RY library in favor of R*Y* and N_{60} clones. Notably, this structure−function dissection of the IgE-binding motif found in clone I9-102 reveals that this clone is substantially similar to a previously described IgE aptamer reported by Tasset and co-workers[38] and analyzed by Woodbury and co-workers.[39]

With the IgE binding motif elucidated, we calculated the probability that it occurred in the patterned and unpatterned libraries using a computer algorithm to determine the loop probability for each frame (see the Supporting Information for the complete analysis). The likelihood that this motif would occur in the N_{60} library is 2.0×10^{-13}, while its likelihood is 6.7×10^{-13} in the R*Y* library. This 3.4-fold increase in overall probability for the patterned library is a product of two opposing factors: a ~19-fold increase in the likelihood of forming a seven base pair stem in the R*Y* library, counterbalanced by a ~12-fold increase in number of registers in the N_{60} library (because the motif only fits two frames in the R*Y* library).

Collectively, these calculations demonstrate that both the streptavidin-binding consensus motif and the IgE-binding motif occur significantly more frequently in the R*Y* library than in the unpatterned library. Alignment of the variable regions of the 16 unique sequences from round 10 of the VEGF selection revealed that 12 share a common sequence (GTCCGGAATGG-$N_{(0-4)}$-GTGC). In contrast with the streptavidin and IgE cases, however, this consensus sequence was not predicted by OMP to occur in a common context, and variations from the consensus within the 16 unique clones do not correlate with changes in VEGF binding affinity. The motif was therefore considered not sufficiently conserved to enable rigorous probability calculations.

Materials and Methods

General. Analytical and preparative PAGE was performed on 15% polyacrylamide, TBE/urea Criterion precast gels (BioRad). Primers were synthesized by IDT in desalted form (for unmodified primers) or HPLC purified form (for biotinylated primers). Primers

and libraries were quantified by UV absorbance at 260 nm on a NanoDrop1000 spectrophotometer.

Library Computational Analysis. A random sample of sequences from each library was generated computationally. These sequences were modeled using OMP under binding buffer and temperature conditions (0.15 M monovalant cation, 0.001 M $MgCl_2$, 22 °C). The predicted energy of the most probable predicted fold for each sequence was used for analysis.

IgE and VEGF Library Synthesis. The libraries were synthesized on a PerSeptive Biosystems Expedite 8090 DNA synthesizer using standard protocols with DNA phosphoramidites and reagents from Glen Research. Phosphoramidite mixtures were made starting from the four individual bases suspended in acetonitrile under anhydrous conditions. HPLC analysis of X−C dimers[31] using a mobile phase of aqueous triethylammonium acetate and acetonitrile (2−8% acetonitrile) was used to assess product nucleotide ratios. The mixtures were adjusted empirically to give desired product nucleotide ratios. The library sequences are as follows, including primers and library tags.

N_{60}: 5′TGTCGCTGCGTCGCCTGGGATCC6666666666666666666666666666666666666666666666666666666666666666CACCGGAAGACGCACGC, where 6 is a mixture that produces 1:1:1:1 A/C/G/T.

R*Y*: 5′TGTCGCTGCGTCGCCTGCAGCTG787878786666787878787866678787878786667878787878666787878787866678787878CACCGGAAGACGCACGC, where 6 is a mixture that produces 1:1:1:1 A/C/G/T, 7 is a mixture that produces 45:5:45:5 A/C/G/T, and 8 is a mixture that produces 5:45:5:45 A/C/G/T.

RY: 5′TGTCGCTGCGTCGCCTGGCTAGC90909090666690909090666690909090666690909090666690909090CACCGGAAGACGCACGC, where 6 is a mixture that produces 1:1:1:1 A/C/G/T, 9 is a mixture that produces 1:1 A/G, and 0 is a mixture that produces 1:1 C/T.

A 1 μmol-scale synthesis of each library was purified first by OPC cartridge (Applied Biosystems), then by denaturing PAGE purification.

IgE and VEGF Library PCR. PCR and QPCR were performed with the AmpliTaq Gold system from Roche. The standard buffer was supplemented with 20% DMSO and forward and reverse primers at 800 nM. For QPCR, the mix was further supplemented with 0.5× SYBR Green I (Invitrogen). The enzyme was activated by 15 min at 95 °C, followed by amplification cycles of (95 °C for 30 s, 60 °C for 30 s, and 72 °C for 30 s), with a final extension at 72 °C for 5 min. QPCR was performed on a CFX96 Real Time System (Bio-Rad), and PCR was performed on a PTC-200 thermocycler (MJ Research). Common primers were used to amplify all library members. Forward primer, 5′TGTCGCTGCGTCGCCTG; reverse primer, 5′GCGTGCGTCTTCCGGTG.

Amplification Uniformity Tests. On the basis of the bead binding assays, ∼1% of the starting library mix survived each round of selection. The test for amplification uniformity mimicked this rate of survival through 10 rounds of dilution followed by PCR. In each round, ∼160 fmol of library mix DNA was diluted 128-fold, amplified by 7−9 cycles of PCR, purified using a Qiagen MinElute PCR purification kit, and quantified by fluorescence spectroscopy with PicoGreen (Invitrogen Molecular Probes) on a Spectramax Plus 384 plate reader (excitation at 480 nm, emission at 520 nm) by comparison to a standard curve of λ DNA in TE buffer. After 10 rounds of dilution and PCR, the output was amplified and analyzed by tag digestion as described below.

Library Tag Digestion. For the IgE and VEGF selection material, the pool to be analyzed by digestion was amplified by PCR to the highest cycle that remained in the exponential amplification phase based on QPCR. Typically, 1.13 mL of PCR resulted in ∼4.5 pmol total dsDNA, or ∼300 fmol dsDNA for each digestion reaction, performed in triplicate. Following PCR, the DNA was purified with a Min-Elute PCR Purification Kit (Qiagen) and dissolved in 1× NEBuffer 2 + 1× BSA (NEB). To each 20 μL sample containing 300 fmol DNA, 10 units *BamH*I, 2.5 units *Pvu*II, or 2.5 units *Nhe*I was added (restriction enzymes were from New England Biolabs). The DNA was digested for 1 h at 37 °C, then analyzed on a 3% agarose gel. The digestion products were visualized and quantified by ethidium bromide staining and densitometry on an AlphaImager HP using AlphaEase FC software.

Tag digestion for the streptavidin selection was performed as above, except ∼400 fmol of dsDNA was purified after PCR and dissolved in 1× NEBuffer 3 + 1× BSA (NEB). To each 20 μL sample containing ∼100 fmol DNA, 20 units *BamH*I or 10 units *Pvu*II was added. The digestions were analyzed as described above.

IgE and VEGF Selections. NHS-activated Sepharose 4 Fast Flow beads (240 μL, GE Healthcare) were washed three times with PBS. Human IgE, produced in human myeloma plasma (Athens Research), recombinant human $VEGF_{165}$ produced in Sf9 cells (Invitrogen/Gibco), or bovine serum albumin (BSA, Sigma-Aldrich) was prepared as a 2.5 μM solution in 200 μL PBS, pH 7.4, and mixed with the beads by slow end-over-end rotation overnight at 4 °C in PBS. The beads were washed three times and any remaining NHS ester groups were quenched by incubation with 0.1 M Tris, pH 8, for 3 h of slow end-over-end rotation at 4 °C. Finally, the beads were washed three times and suspended in 300 μL of total volume by the addition of 120 μL of PBSM (138 mM NaCl, 2.7 mM KCl, 8.1 mM Na_2HPO_4, 1.1 mM KH_2PO_4, and 1 mM $MgCl_2$, pH 7.4).

In a typical selection round, ∼20 μL of beads containing ∼50 pmol immobilized protein was suspended in PBSM. DNA from the previous round (∼6 pmol) was suspended in PBSM and heated to 95 °C for 5 min, then cooled rapidly on ice for 5 min to anneal the ssDNA into intramolecular secondary structures. The DNA and beads were combined in 200 μL total volume (∼250 nM protein; ∼25 nM DNA) and shaken on a vortexer at 800 rpm for 1 h at 25 °C. The beads were harvested by brief centrifugation and washed one to two times with 200 μL of PBSM. Bound DNA was eluted in two sequential 30-s treatments with 200 μL of 0.1 M NaOH + 10 mM EDTA, with centrifugation after each treatment. The supernatants were combined and precipitated with ethanol.

For the first round of the IgE selection, the protocol was as described above except the selection was performed in 400 μL of total volume with 100 pmol DNA and ∼200 pmol IgE on beads (250 nM DNA, 500 nM IgE). For the first round of the VEGF selection, the selection was performed in 200 μL of total volume with 200 pmol DNA and 100 pmol VEGF (1 μM DNA, 500 nM VEGF). For the second round of the VEGF selection, the protocol was performed in 200 μL of total volume, with ∼12 pmol DNA and ∼67 pmol VEGF (∼60 nM DNA, ∼330 nM protein). The protocol for the negative selections to remove DNAs that bound beads rather than target proteins was the same as above, except that no target protein was used (beads were simply blocked with 0.1 M Tris, pH 8, as described above), and the supernatant, rather than the pelleted beads, was saved following the incubation of the library pool with the blank beads. The beads were washed with 200 μL of PBSM, which was also saved and combined with the original supernatant. The recovered DNA was precipitated with ethanol and resuspended in PBSM for the subsequent round of selection.

The DNA output from each round was dissolved in 0.1 M NaCl. To minimize the possibility of dynamic compression,[40] 1/130 of the output was analyzed by QPCR to determine the end of the exponential phase (typically 8−12 cycles). Once the correct number of cycles had been determined, half of the output mix was amplified

(36) Ekland, E. H.; Bartel, D. P. *Nucleic Acids Res.* **1995**, *23*, 3231–3238.
(37) Knight, R.; Yarus, M. *Nucleic Acids Res.* **2003**, *31*, e30.
(38) Wiegand, T. W.; Williams, P. B.; Dreskin, S. C.; Jouvin, M. H.; Kinet, J. P.; Tasset, D. *J. Immunol.* **1996**, *157*, 221–230.
(39) Katilius, E.; Flores, C.; Woodbury, N. W. *Nucleic Acids Res.* **2007**, *35*, 7626–7635.
(40) VanGuilder, H. D.; Vrana, K. E.; Freeman, W. M. *BioTechniques* **2008**, *44*, 619–626.

to the end of the exponential phase. The other half was saved for recovery and analysis. PCR amplification used the exact same protocol as QPCR, but omitted SYBR Green. For a standard round, 1.6 mL of PCR reactions resulted in 5−10 pmol of DNA. During PCR, the reverse primer was biotinylated, while the forward primer was unmodified. The PCR mix was purified to remove unincorporated primers using a Qiagen Min-Elute PCR Purfication Kit. The double-stranded DNA was eluted into a small volume of 0.1 M NaCl.

To effect strand separation, 150 μL of streptavidin magnetic particles (Roche) were washed and suspended in 900 μL of 0.1 M NaCl. The DNA was added and shaken for 30 min at 25 °C. The streptavidin particles were captured with a magnet and washed with 900 μL of 0.1 M NaCl. The forward (desired) strand was eluted with 900 μL of 0.1 M NaOH. The eluted single-stranded DNA was precipitated with ethanol and resuspended in PBSM for the next round of selection.

Bead Binding Analysis. Selection progress was monitored by analyzing a small fraction of each pool for activity in an assay that mimicked the selection protocol. Approximately 1/30 of each round's surviving, amplified, strand-separated DNA was radioactively labeled with γ-^{32}P-ATP and T4 polynucleotide kinase (T4 PNK, NEB). Each labeling reaction contained 10 μL of 1× T4 PNK buffer with 5 mM DTT, 5 μCi γ-^{32}P-ATP, and 10 U T4 PNK. The labeling reaction was incubated at 37 °C for 30 min, then diluted to 200 μL with 0.3 M NaCl. The DNA was purified by extraction with 200 μL of 1:1 phenol/chloroform and precipitated with ethanol. The labeled DNA was diluted to a working concentration of ∼10 nM.

The bead-binding analysis was identical to the selection binding protocol above, except ∼50 fmol radioactively labeled DNA was bound to beads in the presence of an excess (5 pmol) of unlabeled N_{60} library as a blocking agent. A labeled positive control (50 fmol) was also included in every experiment to verify successful binding and elution.[38,41] Each sample was bound in parallel to both blank beads and target-bound beads to test for target-independent binding to beads. The flow through, washes, and elution were saved separately and quantified by PAGE and densitometry. The fractions were separated by 15% denaturing PAGE so that the pool or clone material and positive control could be compared in one pot. The samples were exposed for 2−5 h on a GE Healthcare Phosphor Screen, then imaged on a Typhoon Imager (GE Healthcare) and quantified by densitometry using ImageQuant software.

Filter-Binding Assay. Strand-separated and kinase-radiolabeled DNA (∼3 pmol) was purified by 15% denaturing PAGE. The DNA was eluted into 0.3 M NaCl by rotation for ∼6 h at 25 °C. The purified stock was resuspended in PBSM to a final concentration of ∼10 nM. The double filter method[42] (nitrocellulose on top of nylon) was used to collect both bound and unbound DNA. The filters were Pall Biotrace NT 0.2 μm (nitrocellulose) and BioDyne B 0.45 μm (nylon). Each filter was presoaked for at least 2 h in PBSM. The target protein was suspended in PBSM to make a 5 μM stock, then diluted to the experimental concentrations. The DNA was diluted to ∼2 nM in PBSM, heated at 65 °C for 5 min, then cooled to room temperature (RT). The DNA and protein were combined in 20 μL total, at the following concentrations. For the IgE selection, ∼ 0.05 nM/1 fmol DNA with 0.4−2000 nM IgE; for the VEGF selection, 0.4 nM/4 fmol DNA with 0.4−1,500 nM VEGF. The DNA and protein were incubated 1 h at 25 °C before filtration. The filters were layered into a slot-blot manifold (GE Healthcare PR 648 Slot Blot Manifold, 48-well) and vacuum (∼30 Torr) was applied constantly. The filters were washed with 200 μL of PBSM before the 20 μL samples were applied. Under vacuum, the samples passed through the filters in ∼1 s. Each sample was washed immediately twice with 200 μL of PBSM (∼5 s to filter). The filters were air-dried and exposed 24−52 h on a GE Healthcare Phosphor Screen. The screen was imaged on a Typhoon Imager (GE Healthcare) and quantified by densitometry with ImageQuant software. For each sample, the percent bound was calculated as nitrocellulose counts/(nitrocellulose counts + nylon counts).

IgE and VEGF Selection DNA Sequencing. Pools were amplified with 5′ unmodified forward and reverse primers to the cycle number indicated by QPCR. A total of 60 μL of PCR mix (∼250 fmol DNA) was ligated into the TOPO 4 vector (Invitrogen), and cloned into Mach1 chemically competent *Escherichia coli*. Individual colonies were picked and amplified with the TempliPhi system (GE Healthcare). The inserts were sequenced with the M13Reverse primer by the Dana-Farber/Harvard Cancer Center DNA Resource Core. Inserts were trimmed, assigned, and aligned in Vector NTI (Invitrogen).

Streptavidin Library Synthesis. The libraries for the streptavidin selection were synthesized as above, except with different primer binding sites. The sequences as the follows.

N_{60}: 5′CGGTGCTCCTTGCGGTCGGATCC6666666666666666666666666666666666666666666666666666666666666666666666GCACCAGACCACACGG, where 6 is a mixture that produces 1:1:1:1 A/C/G/T.

R*Y*:5′CGGTGCTCCTTGCGGTCCAGCTG7878787866667878787866678787878786666787878787866678787878 GCACCAGACCACACGG, where 6 is a mixture that produces 1:1:1:1 A/C/G/T, 7 is a mixture that produces 45:5:45:5 A/C/G/T, and 8 is a mixture that produces 5:45:5:45 A/C/G/T.

The primers for the streptavidin selection were synthesized on an Expedite 8090 with Glen Research reagents and phosphoramidites and purified by reverse-phase HPLC. Forward primer, 5′CGGTGCTCCTTGCGGTC; reverse primer, 5′CCGTCGTGGTCTGGTCG. The forward primer was synthesized in both unmodified form (for selection, digestion, and cloning) and with a 5′-fluorescein group (for CE analysis). The reverse primer was synthesized in both unmodified form (for digestion and cloning) and in 5′ phosphorylated form (using CPR II, Glen Research) for exonuclease digestion before selection.

Streptavidin Selections. Round 1: 300 pmol of library mix was denatured in binding buffer (100 mM NaCl, 10 mM $MgCl_2$, 50 mM Tris, pH 7.8) at 95 °C for 5 min, then cooled on ice for 5 min. The DNA was added to 40 μL of streptavidin-agarose resin (Novagen) and incubated with end-over-end rotation for 2 h at 25 °C. The DNA and resin were transferred to a 5.0 μm Amicon Ultrafree-MC spin filter (Millipore) and centrifuged to separate the supernatant from the resin. The resin was washed six times with 200 μL of binding buffer, and the bound DNA was eluted twice using 30-min incubations with 200 μL of binding buffer containing 1 μM streptavidin (Sigma-Aldrich). The recovered elutions were precipitated with ethanol and amplified by PCR. Rounds 2−4: DNA from 100 μL of PCR reactions (see below) was bound to 40 μL of streptavidin-agarose as described above, followed by six washes with binding buffer. Rounds 5−8: DNA from 100 μL of PCR reactions was bound to 40 μL of streptavidin-agarose as above, followed by six washes with binding buffer. These were followed by two additional brief (30 s) washing steps with 200 μL of 1 μM streptavidin to remove binders with fast k_{off} in an effort to separate weak and tight binders. Rounds 9−10: DNA from 100 μL of PCR was bound to 40 μL of streptavidin-agarose as above, followed by six washes with binding buffer. These were followed by four additional brief washing steps with 200 μL of 1 μM streptavidin. The bound DNA was eluted twice using 30-min incubations with 200 μL of binding buffer containing 1 μM streptavidin.

Output DNA was amplified by PCR (13−15 cycles) with a 5′-phosphorylated reverse primer. PCR product yields were monitored by agarose gel electrophoresis. The reaction was purified using a PCR Purification Kit (Qiagen) and diluted to 1× final concentration of λ-exonuclease buffer (New England Biolabs). Five units of λ-exonuclease were added, and the solution was incubated for 30

(41) Jellinek, D.; Green, L. S.; Bell, C.; Janjic, N. *Biochemistry* **1994**, *33*, 10450–10456.
(42) Wong, I.; Lohman, T. M. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 5428–5432.

min at 37 °C. Phenol/chloroform extraction followed by size exclusion on a Centri-sep spin column (Princeton Separations) provided the single-stranded material for the next round of selection.

Capillary Electrophoresis. The DNA from 200 μL of PCR using 5′-fluorescein modified forward primer and 5′-phosphorylated reverse primer was purified by Qiagen spin column, λ-exonuclease digestion, phenol/chloroform extraction, and size exclusion as described above. The library was suspended in 10 μL of CE running buffer (25 mM Tris, pH 8.1, 3 mM $MgCl_2$), heat denatured at 95 °C for 5 min, and cooled on ice. Streptavidin (1 μM) in CE running buffer was added, and bound DNA was separated on a 40 cm bare-fused silica capillary with 75 μm inner diameter at 30 kV on the ProteomeLab PA 800 (Beckman Coulter).

Streptavidin Selection DNA Sequencing. After 10 rounds of selection, the library was subcloned into the PCRII TOPO vector (Invitrogen) and grown on carbecillin-agar plates at 37 °C overnight. Individual colonies were picked and amplified using the TempliPhi system (GE Healthcare) followed by a PCR reaction for DNA sequencing using BigDye Terminator Master Mix (Applied Biosystems). These reactions were purified using Centri-Sep Spin Columns and then sequenced on an ABI3730x Genetic Analyzer. Sequences were analyzed with Vector NTI software to identify consensus motifs.

IgE Clone 102 min Library Selection. Primers and spacers were chosen that did not interact with the 102 min sequence, as predicted by OMP. Sequences for the selection were as follows. Forward primer, 5′ACCTATCGTATCCTACCGA; reverse primer, 5′TGAGTCTACCTTACTCCAC; library, 5′ACCTATCGTATCCTACCGATTTgcacacacttcatccgtaccttctagtgggtgtgtgcTTTGTGGAGTAAGGTAGACTCA, where lower case bases imply a mixture that produces 79% of the indicated base and 7% of each of the other three bases. The library was synthesized and purified by PAGE by IDT. The IgE-binding selection was performed as described above, except for round 1 (which used 1 nmol DNA and ~200 pmol IgE on beads in 400 μL of PBSM) and round 2 (which used ~20 pmol DNA and ~100 pmol IgE on beads in 210 μL of PBSM. PCR was performed as described above, except primer annealing was at 55 °C for 30 s. Strand separation was performed as described above. The bead-binding assay was performed as described above, except 120 ng of phage λ DNA was used as the blocking agent.

Discussion

Overall, the patterned R*Y* DNA library described here outperformed a standard unpatterned library of the same length during *in vitro* selections for binding to three target proteins as measured by several criteria: enrichment of the R*Y* library during the selections, the frequency of R*Y*-derived clones isolated at the end of selections, and the binding potency of individual characterized aptamers arising from the R*Y* library. This improvement in functional potential likely arises from an increase in the average amount of secondary structure in the patterned library. The increased secondary structure of the patterned library is evident both from the increased likelihood of forming stretches of consecutive base pairs (as observed in the motif probability calculations) and also from the larger calculated average folding energy of R*Y* library members, which were one standard deviation or more higher than that of the N_{60} library.

Our findings are consistent with and validate previous analyses that have noted the unusually structured nature of nucleic acid aptamers. For example, Szostak and co-workers studied a series of RNA aptamers for GTP and showed that active sequences arose from the most structured 33% of a random library, and that high-affinity aptamers are generally among the most structured 5% of a random unpatterned library.[25] For the N_{60} library used in this work, the 67th percentile of calculated folding energy is −17.4 kcal/mol, and the 95th percentile of calculated folding energy is −22.0 kcal/mol. In comparison, approximately 77% of the R*Y* library used in this work has a predicted folding energy of more than −17.4 kcal/mol, a 2.3-fold increase over N_{60}. Moreover, 34% of the R*Y* library has a predicted folding energy of more than −22.0 kcal/mol, representing a 7-fold increase over N_{60} of highly structured sequences thought to be common among high-affinity aptamers.

A similar analysis can reveal the percentage of each library that is at least as structured as the active aptamers isolated in the three selections described in this work. For the streptavidin selection, the average predicted folding energies for the libraries used were −9.8 and −13.0 kcal/mol for N_{60} and R*Y*, respectively. The 31 unique sequences related to the streptavidin binding hexaloop (out of 37 unique sequences total) isolated from round 10 of the streptavidin selection have an average predicted folding energy of −16.8 kcal/mol under the experimental conditions. Therefore, the characterized active aptamers are more than one standard deviation more structured than the average of the R*Y* library used in this work and more than two standard deviations more structured than the average of the N_{60} library. Only 2.2% of the sequences in the experimental N_{60} library are predicted to be at least as structured as the average streptavidin-binding aptamer; in comparison, 12.5% of the sequences in the R*Y* library are predicted to be at least this structured, representing a 5.8-fold increase over N_{60}. The range of folding energies sampled by the patterned R*Y* library, therefore, overlaps with the folding energies of the isolated aptamers much better than those accessed in the unpatterned, random library.

For the IgE selection, clone I9-102 had a predicted folding energy of −18.4 kcal/mol. In the N_{60} library used for the IgE selection, 25% of the sequences are predicted to be at least as structured as this aptamer; in contrast, 69% of the R*Y* library used in the selection are predicted to be at least as structured as I9-102, representing a 2.8-fold increase over N_{60}. Likewise, the eight assayed VEGF aptamers found in this work had an average calculated folding energy of −19.5 kcal/mol, and the three best aptamers had an average calculated folding energy of −21.2 kcal/mol. For the N_{60} library used for the VEGF selection, 17% of the sequences are at least as structured as the eight tested aptamers, and only 8% are at least as structured as the three best aptamers. Once again the predicted folding energies of the R*Y* library used in the selection more closely resembles the predicted folding energies of these aptamers; 59% of the R*Y* library sequences are predicted to be at least as structured as the average VEGF aptamer, and 41% of the R*Y* library sequences are predicted to be at least as structured as the three best aptamers, representing a greater than 5-fold increase over N_{60}.

The streptavidin-binding motif with its two stems, bulge, and loop sequences illustrates several of the design principles underlying the patterned library. The motif matched the pattern in several different frames, which collectively are 27-fold more likely to contain the motif than all of the frames of the N_{60} library (see the Supporting Information for a complete analysis). All of the frames in the R*Y* library that are compatible with the streptavidin-binding motif require the N_3 and N_4 loops that were included in the pattern design in order to accommodate the motif. This motif also illustrates the importance of the incorporation of small amounts of off-pattern bases into the R*Y* library. In the frame with the highest probability of containing the streptavidin-binding motif, the stem includes a mismatched R*−R* pairing that would require a 5% non-purine base at either position to form a base pair. In the other probable frames, the stem contains matched R*−Y* pairs, but the bulge and loop positions require non-RY residues. Thus, while the

R*Y* library was significantly more likely to contain the streptavidin binding motif than the random unpatterned library, a pure RY pattern does not contain the motif at all.

The IgE-binding motif is also more likely to occur in the designed R*Y* library than in an unpatterned library of the same length. The large size and rigid sequence constraints of the loop, however, did not ideally match the pattern. There were two frames with a significant probability of containing the binding motif; collectively, they were 3.4-fold more likely to contain the motif than all the frames of the N_{60} library combined. The IgE-binding motif also requires the incorporation of small amounts of off-pattern bases into the R*Y* library in order to accommodate the loops. Thus, while the R*Y* library is 3.4-fold more likely to contain the IgE-binding motif than a random, unpatterned library of the same length, the motif's requirements preclude its existence in the pure RY library, consistent with the observed loss of the RY library in the IgE and VEGF selections.

The occurrence of this IgE binding motif would have been even higher in the R*Y* library if the unusually large binding loop had better fit the pattern. The large size of the binding loop forced the stem to extend into an N_3 region, and thus did not allow the stem to fully benefit from R:Y pattern-enhanced pairing. To better accommodate motifs with large loop sizes, a modified pattern could be used, with larger (8−10 base) N_m regions alternating with R*Y* patterned regions. A pattern with a combination of both large and small N_m regions might serve as the best general pattern for accommodating binding motifs for a variety of targets. Further experimental and computational analyses could lead to the development of optimized patterns for nucleic acid aptamer and catalyst evolution.

Buffer conditions, including divalent cation concentration, can affect nucleic acid folding energies and therefore nucleic acid structures. Commonly used selection conditions vary widely in the concentration of divalent cations, typically ranging from 0 to 10 mM. The selections performed in this study were conducted in the presence of 1−10 mM divalent magnesium, which could affect the general applicability of the patterned libraries discussed here. Increasing divalent cation concentration during selection in principle could stabilize folded structures and decrease the benefit of pattern-enhanced base pairing, although for many uses of aptamers the buffer conditions are constrained by the application. We observed that IgE aptamers from both the N_{60} and R*Y* libraries exhibit comparable binding activity in buffer supplemented with 1 mM or 10 mM $MgCl_2$ (Supporting Information Figure S11). Moreoever, although the magnitude of predicted folding energy is greater for all sequences in 10 mM $MgCl_2$ than in 1 mM $MgCl_2$, the relative distribution of predicted folding energies for 3000 randomly generated sequences from each library remains unchanged in the two buffer conditions (Supporting Information Table S19). In addition, the fraction of each library predicted to be at least as structured as the active aptamer sequences characterized in this work remains similar in 1 mM and 10 mM $MgCl_2$ (Supporting Information Table S20). Collectively, these findings suggest that patterned nucleic acid libraries may demonstrate improved functional potential across a range of divalent cation concentrations commonly used in selections.

The patterned libraries described in this work are unlikely to form G-quartets, which require stretches of three to four guanosines in a row, separated by loops.[21] While only a modest fraction of protein-binding DNA aptamers contain G-quartets, they have been more frequently observed in DNA aptamers that bind small molecules.[7,11] In general, these G-quartets form extremely compact structures into which some small-molecule ligands can intercalate. For aptamer evolution efforts in which the target is a small molecule capable of efficiently binding a G-quartet, the patterned libraries studied here may not be superior starting points compared with traditional unpatterned libraries.

Conclusion

We designed a patterned nucleic acid library with both an unusually high degree of secondary structure and the ability to accommodate loops and bulges typically found in aptamers and catalysts. Patterning with alternating pyrimidine- and purine-rich positions increased predicted average secondary structure compared with that of a standard unpatterned library of the same length while maintaining the ability of the libraries to be prepared in a single solid-phase synthesis with no splitting and pooling required. The patterned library did not incorporate any fixed sequences but instead was designed to allow binding motif flexibility by incorporating at least a small percentage of every base at every position.

The functional potential of the structured library was compared to that of an unpatterned N_{60} library by three different competitive selections for binding to streptavidin, IgE, or VEGF. In all three selections, the patterned library was enriched relative to the unpatterned library present in the same solution over the 9 or 10 rounds comprising the selection. Characterization of individual aptamer clones emerging from the three selections further revealed that the highest affinity aptamers observed arose from the patterned library for two protein targets, while in the third case, the highest affinity aptamers from the patterned and unpatterned libraries exhibited comparable affinity. A related library without any flexibility at the purine- and pyrimidine-rich positions failed to result in any active molecules, suggesting that the inclusion of a small amount of off-pattern bases into the patterned positions is necessary for binding. Calculations indicate that the consensus binding motifs elucidated for IgE and streptavidin aptamers are, respectively, 3.4- and ~27-fold more likely to occur in the patterned library than in the N_{60} library. Our findings indicate that nucleic acid libraries with greater average secondary structure can exhibit enhanced functional potential and collectively suggest that researchers performing nucleic acid selections should consider using patterned libraries rather than the standard unpatterned, random libraries to improve the frequency and activity of library members.

Acknowledgment. K.M.R. gratefully acknowledges a National Science Foundation Graduate Research Fellowship. This work was supported by NSF CAREER Award MCB-0094128, NIH/NIGMS R01 GM065400, and the Howard Hughes Medical Institute.

Supporting Information Available: Library design details; analysis of selection results; streptavidin binding motif probability calculation; IgE aptamer 9-102 minimization and motif analysis; IgE binding motif probability calculation; effects of divalent cation concentration on aptamer function and predicted structure. This material is available free of charge via the Internet at http://pubs.acs.org.

JA103023M