Nucleic acid evolution and minimization by nonhomologous random recombination

Joshua A. Bittker, Brian V. Le, and David R. Liu*

Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138

Abstract

We have developed a simple method for exploring nucleic acid sequence space by nonhomologous random recombination (NRR) that enables DNA fragments to randomly recombine in a length-controlled manner without the need for sequence homology. We compared the results of using NRR and error-prone PCR to evolve DNA aptamers that bind streptavidin. Starting with two parental sequences of modest avidin affinity, evolution using NRR resulted in aptamers with 15- to 20-fold higher affinity than the highest-affinity aptamers evolved using error-prone PCR, and 27- or 46-fold higher affinities than parental sequences derived using systematic evolution of ligands by exponential enrichment (SELEX). NRR also facilitates the identification of functional regions within evolved sequences. Inspection of a small number of NRR-evolved clones identified a 40-base DNA sequence, present in multiple copies in each clone, that binds streptavidin. Our findings suggest that NRR may enhance the effectiveness of nucleic acid evolution and the ease of identifying structure–activity relationships among evolved sequences.

Our ability to subject combinatorial libraries of DNA or RNA sequences to iterated cycles of *in vitro* selection, amplification, and diversification has enabled the evolution of nucleic acids with tailor-made binding or catalytic properties^{1,2}. Existing methods for the *in vitro* evolution of nucleic acids have primarily used two methods for exploring the possible sequences of a given length (“sequence space”)—SELEX with or without intentional point mutagenesis.

SELEX^{3,4} consists of iterated cycles of selection and amplification without intentional mutagenesis between rounds. By itself, SELEX does not explore the sequence space of a nucleic acid in a purposeful manner, but rather enriches those active molecules already present within the starting pool of sequences, although in practice rare errors during DNA or RNA amplification introduce mutations at a low rate. In contrast, SELEX in combination with error-prone PCR⁵ or hypermutagenic PCR⁶ introduces point mutations that may enhance desired binding or catalytic properties^{7–10} at a frequency of ~1–10% per base per PCR reaction. Although point mutagenesis can potentially access more active neighbors in sequence space, the evolved molecules (“solutions”) found will still be quite similar to molecules in the initial pool (Fig. 1).

Exploring nucleic acid sequence space by random recombination may offer substantial benefits over simple enrichment or point-mutagenesis strategies. For example, random recombination between any positions may enhance binding affinities or catalytic activities by the repetition of active subsequences, enabling polyvalent interactions¹¹ between DNA sequences and ligands or substrates, or by combining nucleic acid motifs that each bind to a

different part of the target molecule. Random recombination, in theory, could also remove detrimental sequences, optimize the length between active motifs, and alter their relative order (Fig. 1). To date, a means of inducing random recombination during the *in vitro* evolution of nucleic acid aptamers or catalysts has not been reported. Unintentional recombination during SELEX can take place by template switching during PCR^{12,13} or reverse transcription¹⁴. However, as with the DNA shuffling methods^{15–18} used to recombine genes encoding protein libraries, these sources of recombination rely on DNA annealing and therefore require levels of sequence homology that may benefit only fairly mature pools containing similar solutions. Several newer methods^{19–22} allow nonhomologous recombination of genes encoding protein libraries, but none have been applied to nucleic acid evolution. Some of these methods allow only a single crossover^{19,23}, precluding internal deletion, internal insertion, or reordering of gene segments. Others require regions of homology between nonhomologous crossovers^{20–22}, or require the synthesis of oligonucleotides encoding each possible crossover^{20,22} and therefore necessitate knowledge of the DNA sequence of all starting species.

Here we report the development of a method for exploring nucleic acid sequence space by nonhomologous random recombination (NRR). We have compared the effectiveness of NRR with that of error-prone PCR for evolving a DNA aptamer using identical starting sequences and selection conditions. Our findings suggest that NRR may evolve nucleic acids substantially more effectively than pure SELEX or SELEX with error-prone PCR. In addition, the analysis of sequences evolved using NRR can provide greater insights into the structure–activity relationships of active clones than the analysis of sequences evolved using error-prone PCR. Using NRR, we successfully minimized a DNA aptamer both by inspection of a small number of NRR-evolved sequences and, independently, by controlling the size of the recombined molecules during the NRR process.

Results

A method for nonhomologous random recombination

We hypothesized that the ligation of 5′-phosphorylated, blunt-ended, double-stranded DNA fragments would provide an efficient and general strategy for the nonhomologous recombination of DNA in a size-controlled manner. This approach faced three major challenges. First, the addition of DNA ligases to double-stranded, blunt-ended fragments favors intramolecular circularization over intermolecular ligation. Second, the products of any useful nucleic acid diversification method must contain at their ends, but not internally, defined sequences that can serve as primer-binding sites for PCR amplification after selection. Third, the size of the recombined products must be controlled because sequences that are too large can be difficult to analyze or amplify, whereas sequences that are too small may not be able to form secondary structures that have optimal binding or catalytic properties²⁴.

Our approach (Fig. 2A) overcomes each of these challenges. A pool of starting DNA (for example, random, genomic, or defined sequences) is digested by DNase I. The resulting fragments are treated with T4 DNA polymerase, which can both fill in 5′ overhangs and degrade 3′ overhangs. The blunt-ended, 5′-phosphorylated, double-stranded DNA that is generated is then purified by preparative gel electrophoresis to isolate fragments of the desired size range. These blunt-ended fragments are treated with T4 DNA ligase in the presence of 15% (wt/vol) poly(ethylene glycol) (PEG). Under these conditions, intermolecular ligation is strongly favored over intramolecular circularization²⁵. As T4 DNA ligase catalyzes the efficient ligation of blunt-ended DNA independent of sequence²⁶, fragments recombine in a random and nonhomologous manner. To both control the average length of recombined molecules and install defined sequences at the ends of the diversified

library, a synthetic 5'-phosphorylated DNA hairpin is added in a defined stoichiometry to the ligation reaction. Because DNA molecules capped by ligation to the hairpin can no longer ligate with other molecules, increasing the concentration of hairpin decreases the average length of the recombined library. Digestion of the hairpin-terminated, recombined DNA pool using restriction endonucleases cleaves the end of each hairpin and cuts any undesired hairpin-hairpin dimers to provide a library of recombined, double-stranded DNA molecules flanked by a defined sequence at each end (Fig. 2A). These molecules are suitable for PCR amplification using a single primer sequence that anneals at both ends of each library member.

To test the ability of this method to recombine DNA nonhomologously, we subjected two pairs of unrelated DNA sequences of ~150–300 base pairs (bp) each to the NRR process described above. Each pair of parental sequences were digested to fragment sizes of 25–75 bp and recombined to target sizes of 200–300 bp. The average size of the recombined library could be controlled by modulating the stoichiometry of hairpin in the ligation reaction (Fig. 2B). After PCR amplification of the recombined library, individual daughter clones were subcloned into plasmids and sequenced. An average of 0.8 bases of homology was found at each recombination junction (crossover; $n = 124$), consistent with the theoretical minimum of 0.7 bases of homology ($2 \times \Sigma 0.25^n$) expected from random chance. As expected, the most frequent crossover events took place with zero bases of sequence homology (66 out of 124 crossovers). These results indicate that NRR allows the nonhomologous recombination of unrelated DNA sequences in a length-controlled manner.

Comparison of evolution by NRR and error-prone PCR

To determine how nonhomologous recombination affects the efficiency of nucleic acid evolution compared with point mutagenesis, we evolved DNA-based streptavidin-binding molecules (“aptamers”) by NRR or error-prone PCR using identical starting sequences and selection conditions (Fig. 3). A partially mature pool of streptavidin aptamers was generated by subjecting 5×10^{14} random DNA sequences of ~200 bases in length to three rounds of *in vitro* selection for streptavidin binding and PCR amplification. After three rounds of SELEX, two well-represented library members, S3–13 (a 200-mer not including the flanking 23-base primer sequences) and S3–16 (a 273-mer not including the primer sequences), were isolated and characterized. The affinities of S3–13 and S3–16 for streptavidin were $K_d = 89 \pm 14$ nM and 153 ± 28 nM, respectively (Fig. 4). These two parental sequences, which share no sequence homology, were diversified using either error-prone PCR or NRR to generate three libraries (Fig. 3). Error-prone PCR was used to generate a library of mutated S3–13 variants and a separate library of mutated S3–16 variants. The third library was generated by recombining S3–13 and S3–16 using NRR into daughter clones of similar length to the parents (~250 bp from 25–75 bp fragments). Each library was separately subjected to three rounds of SELEX for streptavidin binding under identical conditions to enrich for the sequences with the highest streptavidin affinities. The streptavidin affinities of individual clones from the resulting S3–13-based error-prone PCR pool, S3–16-based error-prone PCR pool, and NRR-recombined pool (designated 13E pool, 16E pool, and 13×16 pool, respectively), and the average affinity of each pool, were measured *in vitro* (Figs 3 and 4).

Error-prone PCR of S3–13 followed by three rounds of enrichment yielded a pool (13E) of sequences with an average streptavidin affinity comparable to that of the S3–13 parent (average 13E $K_d = 73 \pm 14$ nM, Fig. 4). Individual clones isolated from the 13E pool and assayed demonstrated streptavidin dissociation constants ranging from 51 ± 13 nM to 193 ± 43 nM (Fig. 4). These results suggest that one round of point mutagenesis followed by three rounds of SELEX cannot substantially improve the affinity of S3–13. Similarly, the evolution of S3–16 by error-prone PCR resulted in only a modest increase in average binding affinity (16E pool $K_d = 104 \pm 25$ nM; K_d values of individual assayed clones varied

from 65 ± 10 nM to 141 ± 44 nM, Fig. 4). In contrast, nonhomologous recombination of S3–13 and S3–16, followed by three rounds of SELEX, yielded aptamers with an average streptavidin affinity of $K_d = 13 \pm 3.6$ nM. The affinities of individual clones isolated from the recombined pool ranged from 3.3 ± 1.2 nM to 23 ± 9.2 nM (Fig. 4), representing an improvement of up to 27-fold and 46-fold over the S3–13 and S3–16 parental clones, respectively. The greatest streptavidin affinity among the assayed NRR-evolved clones was therefore 15- to 20-fold higher than the greatest affinities among the assayed error-prone PCR-evolved clones. These results demonstrate that although point mutagenesis provided only modest improvement during DNA-aptamer evolution for streptavidin binding, exploring sequence space by NRR yielded substantially more active molecules.

Using NRR to gain structure–function insights

Sequences of representative clones arising from the 13E, 16E, and 13×16 libraries are shown in Figure 5. Error-prone PCR introduced point mutations into the parental sequences at a frequency of 1.3% (27 mutations found in 2,087 sequenced bases). An examination of the sequences generated by error-prone PCR and selection does not provide insights into function because there are no obvious relationships between the nature of the point mutations and the affinities of the mutant clones. In contrast, an analysis of NRR-derived sequences indicates that nonhomologous recombination, internal deletion, repetition, reorienting, and reordering of sequence motifs commonly occur during NRR (Fig. 5). Moreover, examination of a small number of these sequences suggests clear structure–function relationships. Because nonhomologous recombination followed by selection freely deletes or disrupts nonessential subsequences, only those regions responsible for desired activity are expected to be highly conserved. Indeed, all sequenced 13×16 clones share a common 40-base subsequence derived from the S3–16 parent but are otherwise very different, suggesting that this conserved subsequence may be responsible for streptavidin affinity (Fig. 6).

We synthesized both strands of this 40-base region with or without the flanking 23-base primer sequences and measured the ability of each strand to bind streptavidin. Whereas one strand possessed no detectable streptavidin affinity, the other strand (containing 5′-TCTGTGAGACGACGCACCGTCGCAGGTTTTGTCTCACAG-3′) possessed substantial streptavidin binding affinity ($K_d = 62 \pm 6.6$ nM; $K_d = 105 \pm 21$ nM without the flanking primer sequences). The affinity of this 40-base motif is therefore 2.5-fold higher than that of the S3–16 parent and twofold higher than that of the 16E error-prone PCR-evolved pool, despite its fivefold to sevenfold smaller size. This minimal streptavidin aptamer is predicted²⁷ to fold into a stem–dumbbell motif (Fig. 6B), and the predicted secondary structures of all sequenced 13×16 clones also contain this same motif. To study the binding mode of this minimal aptamer, we assayed its ability to bind streptavidin in the presence of excess biotin. Biotin effectively blocks binding of the minimal aptamer to streptavidin (data not shown), suggesting that aptamers containing this minimal motif may bind to streptavidin at or near the biotin binding site. The rapid identification of a minimal DNA aptamer from a library evolved by NRR, without requiring additional mutagenesis experiments, suggests that NRR can reveal important structure–function information in addition to exploring sequence space more effectively than can point mutagenesis.

Nucleic acid minimization by NRR

The ability of NRR to generate recombined clones of defined average length may allow the removal of nonessential regions from a single parental sequence to generate partially minimized clones. To test this possibility, we subjected a single high-affinity clone from the 13×16 library (13×16#9, 235 bases not including flanking primer sequences, $K_d = 3.3 \pm 1.2$ nM) to NRR, recombining fragments 25–75 bp into clones 50–150 bp in length. The NRR-

diversified library was subjected to three rounds of SELEX under the same conditions used to select the 13E, 16E, and 13×16 libraries. The resulting enriched library (designated 13×16–9 min pool) had an average streptavidin binding affinity of $K_d = 89 \pm 16$ nM, similar to that of the minimal 40-mer (Fig. 4). The characterization of the four smallest individual clones (out of seven) isolated from this library revealed affinities ranging from $K_d = 56 \pm 17$ nM to 127 ± 23 nM and lengths of ~40% of the length of 13×16#9 (91–108 bp not including flanking primer sequences; Fig. 4). All clones minimized by NRR possess the minimal 40-mer motif. These results show that performing NRR under conditions that favor small recombined sequences enables the partial minimization of active nucleic acids by the deletion of nonessential fragments, without requiring sequence or activity data.

Discussion

We have developed a simple method for diversifying nucleic acids using nonhomologous random recombination that may represent a more effective means of evolving nucleic acids than existing methods such as SELEX or SELEX with error-prone PCR. NRR not only allows multiple recombination events between any DNA sequences, but also allows simultaneously the internal deletion, reordering, and duplication of motifs present in evolving nucleic acid pools. The NRR diversification method is sufficiently straightforward that a PCR-amplified, nonhomologously recombined library can be generated from starting DNA in a single day. Using NRR, DNA-based aptamers were evolved with streptavidin-binding affinities as high as $K_d = 3.3 \pm 1.2$ nM, whereas evolution using error-prone PCR under identical conditions for selection and enrichment resulted in aptamers with affinities >15-fold lower. These results indicate that substantially more active regions of sequence space may be accessible with NRR than with point mutagenesis.

In addition, NRR may prove useful in revealing structure–function relationships. A minimal 40-base sequence conferring streptavidin binding activity was discovered by simple inspection of NRR-generated sequences. NRR was also used to partially minimize an evolved streptavidin aptamer without the need for sequence information by subjecting a single active clone to NRR under conditions that favor smaller recombined sequences. Although it is possible to predict structure–function relationships by analyzing consensus sequence motifs generated through traditional SELEX approaches²⁸, NRR will help in elucidating structure–function relationships when such motifs are not obvious. It is also possible to reveal structure–function relationships by generating and assaying a series of progressively truncated mutants, and then resynthesizing, selecting, and assaying variants of minimized sequences at very high (~20%) mutation rates²⁹. In comparison, NRR is much simpler to execute, can be applied to multiple sequences during evolution rather than only to individual evolved sequences, and allows sequence rearrangements other than deletions to take place. On the other hand, the structure–function information revealed by evaluating mutations at every position in a sequence will be of higher resolution than that revealed by NRR.

All of the NRR-derived streptavidin aptamers possess two or more copies of the active 40-mer motif. The seven NRR-evolved aptamers in Figure 4 contain an average of 57 ± 39 nucleotides between 40-mer motifs, but there is no obvious correlation between inter-motif spacing and streptavidin affinity. Streptavidin is a tetrameric protein, and it is therefore possible that NRR-evolved aptamers achieve at least part of their enhanced affinity by simultaneously binding to two or more biotin binding sites. Thus, NRR may generate improved molecules by replacing nonessential regions with multiple copies of appropriately spaced active motifs, rather than through the creation of new active motifs. Among the assayed NRR-evolved clones, however, streptavidin affinities vary substantially, and neither the S3–16 parent nor any of the assayed point-mutated 16E clones (all of which have one

copy of the 40-mer motif) has greater streptavidin affinity than the minimal 40-mer motif. This suggests that the orientation of the active motif relative to flanking sequences, and subtle conformational differences between the NRR-evolved clones and the less active variants, also contribute considerably to the enhanced binding of the NRR-derived sequences in addition to the avidity effects discussed above.

Although we applied NRR to only one or two parental sequences in order to trace the parentage of each resulting daughter clone, it is possible that using NRR to diversify a library of many different parental clones may result in even larger improvements in desired activity. Moreover, it should be possible to adapt NRR to the evolution of RNA, which may prove especially useful in the *in vitro* evolution of multifunctional ribozymes^{30,31} that combine ligand-binding aptamers with catalytic motifs. Applying NRR to protein evolution is also a possibility. However, subjecting protein-encoding genes to NRR will result in up to a sixfold decrease in meaningful protein diversity per recombination event because of the one-in-three chance of recombining in-frame and the one-in-two chance of recombining in the proper orientation. Whether the possible benefits³² that NRR diversification offers the evolution of proteins can overcome the detrimental effects of this decrease in meaningful diversity is the subject of our ongoing studies.

Experimental protocol

General

Primer oligonucleotides were synthesized by standard methods and purified by reverse-phase high-performance liquid chromatography (HPLC). Agarose gels (1–3% wt/vol) were stained with ethidium bromide and visualized by UV transillumination. DNA was quantified by UV spectrophotometry or by gel electrophoresis, staining, and densitometry. Radioactivity from binding assays was quantified using a phosphorimager (Molecular Dynamics), and binding curves were fit to an arctangent function using Microsoft Excel to derive K_d values. Restriction endonucleases, T4 DNA ligase, Vent DNA polymerase, T4 polynucleotide kinase, and T4 DNA polymerase were obtained from New England Biolabs (Beverly, MA). PCR was carried out using Taq PCR Mastermix (Promega, Milwaukee, WI) on a PTC-200 thermal cycler (MJ Research, Waltham, MA). Individual sequences were cloned using the TOPO TA cloning kit (Invitrogen, Carlsbad, CA).

Hairpin and primer sequences

Hairpin and primer sets were changed occasionally to avoid contamination and had no substantial impact on the average streptavidin affinity of evolving pools. Contamination was monitored during each PCR reaction with a negative-control reaction lacking added template DNA.

Sequences were as follows: hairpin 1: 5'-phosphate-CTGTCCGGATACAAGCTTCAGCTGGGCCCCGCGCGGGCCAGCTGAAGCTTGTATCCGGACAG-3'; primer 1: 5'-CTGAAGCTTGTATCCGGACAG-3'; hairpin 2: 5'-phosphate-CCTCCGCGGCATCCGAATTCAGGCCTCCGGGCGCCCGGAGGCCTGAATTCGGATGCCGCGGAGG-3'; primer 2: 5'-CCTGAATTCGGATGCCGCGGAGG-3'.

Initial random pool

Template (5 nmol; 5'-GCCCCGCGGATGGGACGTCCC-N₄₀-CGCCCCGCGGCATCCGACGTCCC-3') and 5 nmol of primer (5'-GGGACGTCCGATGCCGCGGGCG-3') were annealed and extended with Vent DNA polymerase (94°C for 2 min 30 s; 65°C for 30 s; add polymerase; 75°C for 1 h). The 83-bp

product was digested with *FokI* to remove the ends, and the resulting 40-bp product was purified by gel electrophoresis on a 3% (wt/vol) agarose gel. The purified material was treated with T4 DNA polymerase to create blunt ends and purified by gel filtration (Centrisep columns, Princeton Separations, Adelphia, NJ). Double-stranded N₄₀ (57 pmol) was ligated to 57 pmol of hairpin 1 under intermolecular blunt-ligation conditions²⁵ (15% (wt/vol) PEG 6000, 50 μ M ATP in NEB T4 DNA ligase buffer without ATP using 120 Weiss units of T4 DNA ligase, 25°C, 1 h.) This ratio was empirically determined to give products averaging ~250 bp (200 bp of random region). The products were digested with *PvuII* to remove the hairpin ends. The resulting fragments were amplified under error-prone PCR conditions⁵ in 9.6 ml (94°C for 2 min 30 s; then cycled 40 times at 94°C for 30 s; 60°C for 30 s; 72°C for 1 min 10 s) to yield a library of $\sim 5 \times 10^{14}$ molecules with an average size of 250 bp.

Fragmentation of sequences for nonhomologous random recombination

PCR-amplified products were digested with the appropriate type IIS restriction endonuclease (*Bcl*VI for primer 1 or *FokI* for primer 2) to remove the primer ends. Alternatively, if the sequence of an individual clone was known, primers were synthesized to PCR amplify the sequence without the hairpin ends. The resulting DNA (1–15 μ g) was digested with DNase I (Sigma, St. Louis, MO, 7.6 μ g/ μ l, 31.3 units/ μ g) in 10 mM MgCl₂, 20 mM Tris-Cl pH 8.0 for 1–5 min at room temperature using $\sim 2 \mu$ l of a 1:1,000 dilution of DNase I. The digestions were monitored by 3% (wt/vol) agarose gel electrophoresis. When the size of fragments reached the desired average, the reaction was extracted with phenol–chloroform and exchanged into T4 DNA polymerase buffer by gel filtration. The fragments were blunted with T4 DNA polymerase, and fragments of the desired size range were purified on a 3% (wt/vol) agarose gel.

Ligation and amplification of recombined DNA

Blunt-ended fragments were ligated together with hairpin 1 or hairpin 2 at a stoichiometric ratio empirically determined to generate the desired product length (typically, this was similar to the expected stoichiometry). Ligations were carried out under intermolecular blunt-ligation conditions²⁵. The ligation products were digested with the appropriate restriction enzyme to remove the hairpin ends (*PvuII* for hairpin 1 or *StuI* for hairpin 2). Digested ligation products were amplified by PCR using the appropriate primer (primer 1 for hairpin 1 or primer 2 for hairpin 2) at 1 μ M. PCRs were initially denatured at 94°C for 2 min 30 s, then cycled 40 times. Hairpin 1 PCRs were cycled as follows: 94°C for 30 s, 60°C for 30 s, 72°C for 30 s. Hairpin 2 PCRs were cycled as follows: 94°C for 30 s, 72°C for 1 min 30 s. All PCRs ended with a final 10 min extension at 72°C.

In vitro selections

For the three rounds of *in vitro* selection of the random library to generate clones including S3–13 and S3–16, the initial pool was denatured by heating to 95°C in deionized water (Millipore, Bedford, MA) for 5 min and chilling suddenly on ice. Buffer was added to a final composition of 150 mM NaCl, 50 mM Tris-Cl, pH 8.0, 10 mM MgCl₂ (“binding buffer”). Streptavidin–agarose (0.5 ml of a 50% (wt/vol) suspension, Sigma) was washed with binding buffer in a HR5–5 column (Amersham Biosciences, Piscataway, NJ). The library was passed through the column, followed by 50 ml of binding buffer. Desired sequences were eluted with 0.25 mg free streptavidin (Sigma) in 0.5 ml binding buffer, followed by another 1.5 ml of binding buffer. The eluted molecules were amplified by PCR as above. For the *in vitro* selection of sequences starting with parents S3–13 and S3–16 (using libraries diversified by error-prone PCR or NRR), the library and streptavidin–agarose were shaken for 1 h in 1 M NaCl, 50 mM Tris-Cl pH 8.0, 5 mM MgCl₂ (“stringent buffer”) at a final concentration of 1 nM for both DNA and streptavidin. The mixture was loaded into a

HR5-5 column and washed with 50 ml stringent buffer. Desired DNA molecules were eluted by shaking the washed beads with 0.125 mg free streptavidin in 0.9 ml stringent buffer at 25°C for 30 min.

Binding affinity assays

PCR-amplified DNA (1 pmol) was radiolabeled with 15 units T4 PNK and 10 μ Ci [γ -³²P]ATP (Perkin-Elmer Life Sciences, Boston, MA) in T4 PNK buffer at 37°C for 1 h. Labeled DNA was extracted twice with phenol–chloroform and purified twice by gel filtration to remove ATP. The DNA was denatured in water at 95°C for 5 min together with 2 μ g unlabeled human genomic DNA (to block nonspecific DNA binding) per 5 fmol labeled DNA, and then chilled on ice for 5 min. Labeled DNA (5 fmol) plus 2 μ g unlabeled human genomic DNA was added to varying amounts of streptavidin (0–1024 nM) in 50 μ l of 100 mM NaCl, 50 mM Tris-Cl, pH 8.0, 5 mM MgCl₂ (“assay buffer”). The DNA and streptavidin were incubated at room temperature for 2 h. A multiscreen-HA 96-well nitrocellulose filter plate (Millipore), which retains protein–DNA complexes much better than free DNA, was pre-washed with 125 μ l assay buffer and then loaded with each assay sample. Samples were rapidly filtered on a vacuum manifold, and membranes were washed twice with 250 μ l of assay buffer. The membrane from each well was punched out from the plate using a stylus and the bound radioactive label quantified by phosphorimager together with 1 fmol of unreacted probe as a standard.

Sequence analysis and structure prediction

Clones were sequenced by standard automated DNA sequencing methods. Sequences of NRR-derived clones were analyzed using MACAW33 to identify related subsequences. DNA secondary structures were predicted using mfold27.

Acknowledgments

Funding was provided by US National Science Foundation CAREER award MCB-0094128 and by Harvard University. J.A.B. is a Howard Hughes Medical Institute predoctoral fellow. B.V.L. was supported by the Harvard College Research Program. We are grateful to P. Patten (Maxygen, Redwood City, CA) for helpful discussions.

References

1. Wilson DS, Szostak JW. *In vitro* selection of functional nucleic acids. Annu. Rev. Biochem. 1999; 68:611–647. [PubMed: 10872462]
2. Bittker JA, Phillips KJ, Liu DR. Recent advances in the *in vitro* evolution of nucleic acids. Curr. Opin. Chem. Biol. 2002; 6:367–374. [PubMed: 12023118]
3. Kopylov AM, Spiridonova VA. Combinatorial chemistry of nucleic acids: SELEX. Mol. Biol. 2000; 34:940–954.
4. Tuerk C, Gold L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 RNA polymerase. Science. 1990; 249:505–510. [PubMed: 2200121]
5. Cadwell RC, Joyce GF. Mutagenic PCR. PCR Methods Appl. 1994; 3:S136–S140. [PubMed: 7920233]
6. Vartanian JP, Henry M, Wain-Hobson S. Hypermutagenic PCR involving all four transitions and a sizeable proportion of transversions. Nucleic Acids Res. 1996; 24:2627–2631. [PubMed: 8758987]
7. Piganeau N, Jenne A, Thuillier V, Famulok M. An allosteric ribozyme regulated by doxycycline. Angew. Chem. Int. Ed. 2000; 39:4369–4373.
8. Santoro SW, Joyce GF, Sakthivel K, Gramatikova S, Barbas CF. RNA cleavage by a DNA enzyme with extended chemical functionality. J. Am. Chem. Soc. 2000; 122:2433–2439. [PubMed: 11543272]
9. Tsang J, Joyce GF. Specialization of the DNA-cleaving activity of a group I ribozyme through *in vitro* evolution. J. Mol. Biol. 1996; 262:31–42. [PubMed: 8809177]

10. McGinness KE, Wright MC, Joyce GF. Continuous *in vitro* evolution of a ribozyme that catalyzes three successive nucleotidyl addition reactions. *Chem. Biol.* 2002; 9:585–596. [PubMed: 12031665]
11. Rao J, Lahiri J, Isaacs L, Weis RM, Whitesides GM. A trivalent system from vancomycin-D-ala-D-Ala with higher affinity than avidin-biotin. *Science.* 1998; 280:708–711. [PubMed: 9563940]
12. Meyerhans A, Vartanian JP, Wain-Hobson S. DNA recombination during PCR. *Nucleic Acids Res.* 1990; 18:1687–1691. [PubMed: 2186361]
13. Paabo S, Irwin DM, Wilson AC. DNA damage promotes jumping between templates during enzymatic amplification. *J. Biol. Chem.* 1990; 265:4718–4721. [PubMed: 2307682]
14. DeStefano JJ. Kinetic analysis of the catalysis of strand transfer from internal regions of heteropolymeric RNA templates by human immunodeficiency virus reverse transcriptase. *J. Mol. Biol.* 1994; 243:558–567. [PubMed: 7525968]
15. Ness, JE.; Del Cardayre, SB.; Minshull, J.; Stemmer, WPC. *Advances in Protein Chemistry.* Arnold, FH., editor. Vol. Vol. 55. Academic Press; San Diego, CA: 2001. p. 261-292.
16. Stemmer WP. Rapid evolution of a protein *in vitro* by DNA shuffling. *Nature.* 1994; 370:389–391. [PubMed: 8047147]
17. Volkov AA, Arnold FH. Methods for *in vitro* DNA recombination and random chimeraesis. *Methods Enzymol.* 2000; 328:447–456. [PubMed: 11075359]
18. Zhao H, Giver L, Shao Z, Affholter JA, Arnold FH. Molecular evolution by staggered extension process (StEP) *in vitro* recombination. *Nat. Biotechnol.* 1998; 16:258–261. [PubMed: 9528005]
19. Sieber V, Martinez CA, Arnold FH. Libraries of hybrid proteins from distantly related sequences. *Nat. Biotechnol.* 2001; 19:456–460. [PubMed: 11329016]
20. Kolkman JA, Stemmer WP. Directed evolution of proteins by exon shuffling. *Nat. Biotechnol.* 2001; 19:423–428. [PubMed: 11329010]
21. Lutz S, Ostermeier M, Moore GL, Maranas CD, Benkovic SJ. Creating multiple-crossover DNA libraries independent of sequence identity. *Proc. Natl. Acad. Sci. USA.* 2001; 98:11248–11253. [PubMed: 11562494]
22. Tsuji T, Onimaru M, Yanagawa H. Random multi-recombinant PCR for the construction of combinatorial protein libraries. *Nucleic Acids Res.* 2001; 29:E97. [PubMed: 11600716]
23. Ostermeier M, Shim JH, Benkovic SJ. A combinatorial approach to hybrid enzymes independent of DNA homology. *Nat. Biotechnol.* 1999; 17:1205–1209. [PubMed: 10585719]
24. Wedel AB. Fishing the best pool for novel ribozymes. *Trends Biotechnol.* 1996; 14:459–465. [PubMed: 8987914]
25. Upcroft P, Healey A. Rapid and efficient method for cloning of blunt-ended DNA fragments. *Gene.* 1987; 51:69–75. [PubMed: 3596240]
26. Engler, MJ.; Richardson, CC. *The Enzymes.* Boyer, PD., editor. Vol. Vol. 15. Academic Press; New York: 1982. p. 1-29.
27. SantaLucia J Jr. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. USA.* 1998; 95:1460–1465. [PubMed: 9465037]
28. Pagratis NC, et al. Potent 2'-amino-, and 2'-fluoro-2'-deoxyribonucleotide RNA inhibitors of keratinocyte growth factor. *Nat. Biotechnol.* 1997; 15:68–73. [PubMed: 9035109]
29. Eklund EH, Szostak JW, Bartel DP. Structurally complex and highly active RNA ligases derived from random RNA sequences. *Science.* 1995; 269:364–370. [PubMed: 7618102]
30. Kuwabara T, Warashina M, Taira K. Allosterically controllable ribozymes with biosensor functions. *Curr. Opin. Chem. Biol.* 2000; 4:669–677. [PubMed: 11102873]
31. Soukup GA, Breaker RR. Nucleic acid molecular switches. *Trends Biotechnol.* 1999; 17:469–476. [PubMed: 10557159]
32. Bogarad LD, Deem MW. A hierarchical approach to protein molecular evolution. *Proc. Natl. Acad. Sci. USA.* 1999; 96:2591–2595. [PubMed: 10077554]
33. Schuler GD, Altschul SF, Lipman DJ. A workbench for multiple alignment construction and analysis. *Proteins.* 1991; 9:180–190. [PubMed: 2006136]

Figure 1.

Diversification methods for nucleic acid evolution. Starting with parental sequences, pure SELEX enriches active sequences. Error-prone PCR yields parental sequences with point mutations. Homologous recombination methods, such as DNA shuffling, allow crossovers between sequences only in regions of homology. Nonhomologous random recombination allows many other changes, including random crossovers between any sequences, repetition, reordering, reorienting, elimination of subsequences, or any combination of these changes.

Figure 2.

Overview of the nonhomologous random recombination (NRR) method. (A) Starting DNA sequences are randomly digested with DNase I, blunt-ended with T4 DNA polymerase, and recombined with T4 DNA ligase under conditions that strongly favor intermolecular ligation over intramolecular circularization. (B) A defined stoichiometry of hairpin DNA added to the ligation reaction controls the average length of the recombined products. The completed ligation reaction is digested with a restriction endonuclease to provide a library of double-stranded recombined DNA flanked by defined primer-binding sequences.

Figure 3.

Nucleic acid evolution by NRR or error-prone PCR. A random pool of DNA was subjected to three rounds of SELEX for streptavidin binding. Two resulting sequences (S3–13 and S3–16) with moderate streptavidin affinity were diversified using either error-prone PCR or NRR, and the resulting three libraries were separately subjected to identical selection conditions.

Figure 4.

Streptavidin binding affinities of DNA aptamers evolved by error-prone PCR (EPPCR) or by NRR. (A) Representative streptavidin binding data. (B) Assay results showing average K_d value and standard deviation of three to six independent trials. 13E, clones from the error-prone PCR of S3-13; 16E, clones from the error-prone PCR of S3-16; 13 × 16, clones from recombining S3-13 and S3-16 using NRR; 13×16-9 min, clones from minimizing 13×16#9 using NRR.

A S3-13 parent:
 CGGGGGTCCCGCTGCTGCTCAATGACGGCTCAGCTTCGGTGGGCTTTAACAGTAATCAATCATATGAGCAGTTTCAACGATCACCTACCCACACC
 GCTCGAATGTTTGCATAAACCTGGGTAGACTACGCATAATTGGGTATTGAGTCTCTTGTAGGACTAACCGGTTCTATCTCGAGGATTATTAGGTC
 13E#3:
 CGGGGAGCCCGCTGCTGCTCAATGACGACTCAGCTTCGGTGGGCTTTAACAGTAATCAATCATATGAGCAGTTTCAACGATCACCTACCCACACC
 GCTCGAATGTTTGCATAAACCTGGGTAGACTACGCATAATTGGGTATTGAGTCTCTTGTAGGACTAACCGGTTCTATCTCGAGGATTATTAGGTC
 13E#4:
 CGGGGGTCCCGCTGCTGCTCAATGACGGCTCAGCTTCGGTGGGCTTTAACAGTAATCAATCATATGAGCAGTTTCAACGATCACCTACCCACACC
 GCTCGAATGTTTGCATAAACCTGGGTAGACTACGCATAATTGGGTATTGAGTCTCTTGTAGGACTAACCGGTTCTATCTCGAGGATTATTAGGTC
 13E#6:
 CGGGGGTCCCGCTGCTGCTCAATGACGGCTCAGCTTCGGTGGGCTTTAACAGTAATCAATCATATGAGCAGTTTCAACGATCACCTACCCACACC
 GCTCGAATGTTTGCATAAACCTGGGTAGACTACGCATAATTGGGTATTGAGTCTCTTGTAGGACTAACCGGTTCTATCTCGAGGATTATTAGGTC
 S3-16 parent:
 TGACACAAAGACAGACAGGCTATCCAAGAACCTCTTACTCTGTGAGACGACGACCGGTCGAGGTTTTGTCTCACAGACGCTAAAAATACAGACATGC
 ACCAATGAACATGAGTTGACCGCTGTTCTTGAATTTATGGCCGATGTGGTAAGTACTTCTACTGTATCTTCGCGTACCTTAGGTTAAAGTTCTCTTT
 TTGGGAATGTGCTGCCCGGGCATCCGACGCTCCCTTTGGGGGTAGGTGCAACGGGAATCTTGAGGGATCATT
 16E#2:
 TGACACAAAGAC- GACAGGCTA-CCAAGAACCTCTTACTCTGTGAGACGACGACCGGTCGAGGTTTTGTCTCACAGACGCTAAAAATACAGACATGC
 ACCAATGAACATGAGTTGACCGCTGTTCTTGAATTTATGGCCGATGTGGTAAGTACTTCTACTGTATCTTCGCGTACCTTAGGTTAAAGTTCTCTTT
 TTGGGAATGTGCTGCCCGGGCATCCGACGCTCCCTTTGGGGGTAGGTGCAACGGGAATCTTGAGGGATCATT
 16E#3:
 TGACACAAAGAC- GACAGGCTATCCAAGAACCTCTTACTCTGTGAGACGACGACCGGTCGAGGTTTTGTCTCACAGACGCTAAAAATACAGACATGC
 ACCAATGAACATGAGTTGACCGCTGTTCTTGAATTTATGGCCGATGTGGTAAGTACTTCTACTGTATCTTCGCGTACCTTAGGTTAAAGTTCTCTTT
 TTGGGAATGTGCTGCCCGGGCATCCGACGCTCCCTTTGGGGGTAGGTGCAACGGGAATCTTGAGGGATCATT
 16E#4:
 TGACACAAAGAC- GACAGGCTATCCAAGAACCTCTTACTCTGTGAGACGACGACCGGTCGAGGTTTTGTCTCACAGACGCTAAAAATACAGACATGC
 ACCAATGAACATGAGTTGACCGCTGTTCTTGAATTTATGGCCGATGTGGTAAGTACTTCTACTGTATCTTCGCGTACCTTAGGTTAAAGTTCTCTTT
 TTGGGAATGTGCTGCCCGGGCATCCGACGCTCCCTTTGGGGGTAGGTGCTACGGGAATCTTGAGGGATCATT

B Parent 13:
 CGGGGGTCCCGCTGCTGCTCAATGACGGCTCAGCTTC
 GGTGGGCTTTAAACAGTAATCAATCATATGAGCAGTTTTC
 AACGATCACCTACCCACGCTCGAATGTTTGCATAAAC
 CTGGGTAGACTACGCATAATTGGGTATTGAGTCTCTTT
 GATGGACTAACCGGTTCTATCTCGAGGATTATTAGGTC
 Parent 13:
 CGGGGGTCCCGCTGCTGCTCAATGACGGCTCAGC...
 13x16#1:
 GAAACTGCTCATATGATTAGCCCGCTGCTGCTCAATGACGGCTCAGCTCTGTATTTTAGCGTCTGTGAGACAGAACCTGCGACCGGTGCTGT
 CGTCTCAGCTCTACTGTATCTTCGCGTACCTTAGGTTTACCCGCTGCTGCTCAATGACGGCTCTCTGTGAGACAAACCTGCGACCGGTGCTGT
 CTCACAGTAAGAGGTTCTTGGATA
 Parent 13:
 CGGGGGTCCCGCTGCTGCTCAATGACGGCTCAGCTTC
 GGTGGGCTTTAAACAGTAATCAATCATATGAGCAGTTTTC
 AACGATCACCTACCCACGCTCGAATGTTTGCATAAAC
 CTGGGTAGACTACGCATAATTGGGTATTGAGTCTCTTT
 GATGGACTAACCGGTTCTATCTCGAGGATTATTAGGTC
 13x16#5:
 CAGAACACCGGTGCAACTATTGTTATTGGTGCACTGTGAGACAAACCTGCGACCGGTGCTGTCTCACAGAGTA
 AGAGACTCTGTGAGACAAACCTGCGACCGGTGCTGTCTCACAGAGTA
 Parent 16:
 TGACACAAAGACAGACAGGCTATCCAAGAACCTCTTACTCTGTGAGACGAG
 CACCGGTGCGAGTTTTGTCTCACAGACGCTAAAAATACAGACATGACCAAT
 GAACAAATGAGTTCGACCGTGTCTTGAGTTTTATGGCCGATGTGGTAAGT...
 Parent 16:
 TGACACAAAGACAGACAGGCTATCCAAGAACCTCTTACTCTGTGAGACGAG
 CACCGGTGCGAGTTTTGTCTCACAGACGCTAAAAATACAGACATGACCAAT

Figure 5. Diversification by error-prone PCR and by NRR. (A) Representative sequences arising from diversification by error-prone PCR. Mutations are shown in red. (B) Representative sequences generated by NRR compared with parental sequences. The parentage of recombined subsequences generated by NRR is shown in various colors.

Figure 6.

Isolation of a minimal streptavidin binding motif by inspection of NRR-generated clones. (A) All of the sequenced clones evolved by NRR share a motif of ~40 bases despite their otherwise very different sequences. (B) This 40-mer, both with and without flanking primer binding sequences, has streptavidin binding activity and is predicted to fold into a stem–dumbbell structure. (C, D) The predicted secondary structures of full-length NRR-evolved clones invariably include one or more of these stem-dumbbell motifs.