

RNA

Evolutionary dynamics and population control during in vitro selection and amplification with multiple targets

H. Shi, X. Fan, Z. Ni and J. T. Lis

RNA 2002 8: 1461-1470

References

Article cited in:

<http://www.rnajournal.org/cgi/content/abstract/8/11/1461#otherarticles>

Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#)

Notes

To subscribe to *RNA* go to:
<http://www.rnajournal.org/subscriptions/>

METHOD

Evolutionary dynamics and population control during *in vitro* selection and amplification with multiple targets

HUA SHI, XIAOCHUN FAN, ZHUOYU NI, and JOHN T. LIS

Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York 14853, USA

ABSTRACT

Iterative cycles of *in vitro* selection and amplification allow rare functional nucleic acid molecules, aptamers, to be isolated from large sequence pools. Here we present an analysis of the progression of a selection experiment that simultaneously yielded two families of RNA aptamers against two disparate targets: the intended target protein (B52/SRp55) and the partitioning matrix. We tracked the sequence abundance and binding activity to reveal the enrichment of the aptamers through successive generations of selected pools. The two aptamer families showed distinct trajectories of evolution, as did members within a single family. We also developed a method to control the relative abundance of an aptamer family in selected pools. This method, involving specific ribonuclease digestion, can be used to reduce the background selection for aptamers that bind the matrix. Additionally, it can be used to isolate a full spectrum of aptamers in a sequential and exhaustive manner for all the different targets in a mixture.

Keywords: applied molecular evolution; aptamers; *in vitro* selection; RNA; SELEX

INTRODUCTION

RNA molecules not only carry information for their own replication, but also fold into well-defined shapes that may possess certain functional traits. The method known as *in vitro* selection or SELEX (Ellington & Szostak, 1990; Robertson & Joyce, 1990; Tuerk & Gold, 1990) attempts to generate novel affinity reagents by applying genetic selection directly to population of RNA molecules through a process that emulates Darwinian evolution. The novel ligands, or aptamers, generated by this process are capable of binding to a wide variety of targets with high affinity and specificity, which are critical to both basic research and therapeutic applications (Joyce, 1994; Gold et al., 1995; Wilson & Szostak, 1999). While its products fulfill numerous practical goals, the process of *in vitro* selection also serves as a model system of theoretical interest.

The SELEX method has been most frequently used to generate aptamers against single targets. However,

modification of this procedure to allow selection of independent aptamers against multiple targets in a single selection experiment would be useful to drug discovery and functional genomics studies. Although some experiments have been performed against complex targets such as multiunit enzymes (Brown & Gold, 1995, 1996), viral particles (Pan et al., 1995), organelles (Ringquist et al., 1995), and entire cells (Morris et al., 1998), there has not been systematic experimental development of a method to select multiple aptamers. In these experiments with complex targets, multiple families of aptamers are sometimes obtained, including ones that recognized different targets. Nonetheless, they recognized the most abundant or easily recognizable target sites—not necessarily the most desirable ones.

In a formal sense, almost all SELEX experiments involve multiple targets. First, as in the experiments mentioned above, the target can be a mixture of multiple distinct molecules or supramolecular assemblies; second, a single target molecule can contain multiple discrete binding sites on its surface, but separation of the different target sites is physically impossible. In both of these cases, aptamers specific to all the different components in the mixture are desirable. In two other cases, however, selection can yield aptamers with spec-

Reprint request to Hua Shi, Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York 14853, USA; e-mail: hs11@cornell.edu; or John T. Lis, Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York 14853, USA; e-mail: jtl10@cornell.edu.

ificity to components of the system that are not of interest. First, a presumably pure target preparation might contain unknown contaminants, which may select unwanted aptamers that come to dominate the selected population. Second, even the traditional single-target selection scheme in most cases requires a partitioning matrix (e.g., filters used to collect RNA–protein complexes) that could select undesired aptamers with the potential to overrun the selected population. The partitioning matrix, although not an intended target and not necessarily at equilibrium with the aptamer candidates, is exposed to the entire pool of candidates during the SELEX process at the partitioning step. Because the candidates bound to the matrix are copartitioned with those bound to the designated target(s), the matrix acts as a de facto target that is practically inseparable from the intended target(s) in an actual experiment. Therefore, a systematic analysis of a conventional single-target selection that uses a partitioning matrix can yield insights into the process of selection against multiple targets.

Without effective negative selection, matrix-binding aptamers may thwart the isolation of aptamers for the intended targets. We encountered this problem in a selection for aptamers against the *Drosophila* protein B52. In that experiment, we carried the original pool through nine cycles of selection and amplification, using nitrocellulose filters as the partitioning device (Shi et al., 1997). In the final cycle, we included a single negative selection step to eliminate matrix-binding aptamers by passing the aptamer pool through a nitrocellulose filter before amplifying the (unbound) aptamers. After cloning this pool of aptamers, we tested the B52-binding capability of isolated individuals. Among 66 individuals tested in one sampling only 4 showed high-affinity specific binding. Their sequences contained a consensus that was designated the B52 Binding Sequence (BBS) (Shi et al., 1997). From the 62 individuals that did not bind B52, 21 were sequenced and all fell in a different sequence family whose members contain three to five G-triplets. In most cases, four G-triplets (often preceded by an A) occur within a sequence segment 25 nt in length. Sequences meeting these criteria were termed “Multi-G Motif” (MGM). Two sequences that match or resemble MGM consensus were reported by Tuerk et al. (1992) to exhibit significant retention on nitrocellulose filters independent of the protein target. Our results indicated that two families of aptamers were selected by two distinct but coexisting targets: B52 and the nitrocellulose filter.

Here we track the evolving RNA populations in this SELEX experiment by following both the “phenotype,” or binding activity, and the “genotype,” or sequence, of the two aptamer families. Based on the concept of “negative selection according to genotype,” we developed a method to control the evolving sequence populations during selection, which can reduce or eliminate the

matrix-binding aptamers. Moreover, this scheme can be adapted to isolate exhaustively aptamers for multiple targets in a mixture.

RESULTS

In vitro evolutionary dynamics: Following the course of SELEX

To follow the course of enrichment of BBS through successive rounds of selection and amplification, we performed band shift assays, in which B52 protein was incubated with ³²P-labeled RNA transcripts from different generations in the presence of a large excess of tRNA. As shown in Figure 1A, a shifted band comigrating with that generated by the positive control, BBS #8 (Shi et al., 1997) was observed in as early as Generation 3, and peaked at Generation 6. (The slight increase in its intensity from Generation 8 to 9 is presumably a result of the negative selection against matrix binding.) This pattern of change in intensity of the shifted band suggested a more complicated dynamics than a simple monotonic exponential enrichment. Consistent with this, we also noticed apparently simultaneous enrichment of a second band, one that barely enters into the gel. Because this second band is not B52 dependent (data not shown), and there is little radioactivity at the same location in the lane containing the BBS positive control, its enrichment suggests that a different aptamer family might have been coselected.

To test whether the cryptic target responsible for this coselection was the partitioning matrix, we performed a filter-retention assay, which mimics the partitioning condition during the selection. Figure 1B shows an increase of bound RNA as a function of increasing cycles of selection in the absence of B52. This profile of enrichment is similar in its monotonic increase to the one seen in the non-B52 dependent band in Figure 1A. To address whether both results reflect the same aptamer family, we determined the sequences of the RNA in the non-B52 dependent band from Generation 7. Among 25 isolates, 11 were MGMs, 10 sequences were C rich (with at least one C-triplet), 2 were BBSs, and 2 did not match any consensus. The large and similar numbers of G-rich MGMs and C-rich sequences suggested that MGMs may form multimolecular networks with C-rich sequences in the solution, which would have been responsible for the extremely slow mobility on the gel. During the course of the experiment, C-rich sequences may have been enriched through indirect selection by the MGMs. This may be another instance of different RNA species interacting in a pool during selection, with one depending for its survival on the continued presence of another (Hanczyc & Dorit, 1998).

Binding activities of the RNA, as a phenotypic assay, only reveals the summation of all the sequences that contribute to that activity. To study the evolution of a

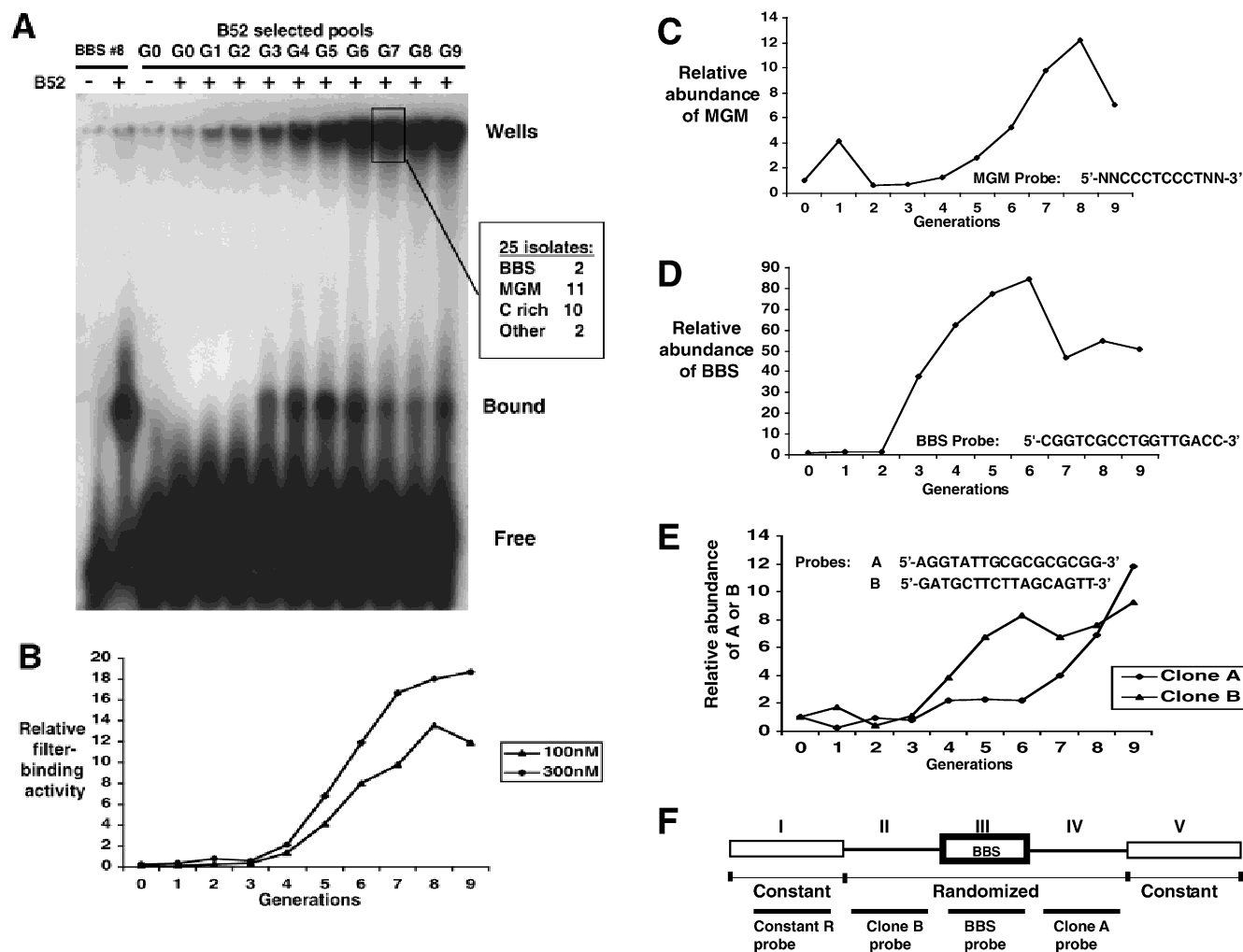


FIGURE 1. Tracking the enrichment of BBSs and MGMs from Generations 0 to 9 of SELEX against B52, partitioned by nitrocellulose filter binding. **A:** Band shift assay of RNA from different generations with the B52 protein on a 2.5% agarose gel. Binding reactions contain 10 pmol of B52 protein in 20- μ L mixtures with labeled RNA representing different generations. The band next to the well in G7 was gel purified, its constituents reverse-transcribed, cloned, and sequenced. Numbers in the inset summarize the sequences of 25 isolates. **B:** B52-independent relative filter-binding activity of different generations. Two different RNA concentrations were used in the filter-binding assay. The radioactive signals were quantified and plotted as fractions of input against generations. The decrease in the last generation may reflect the effect of the negative selection. **C, D:** Relative abundance of MGMs or BBSs, respectively, detected by representative probes in southern dot-blot analysis with DNA pools. The radioactive signals were quantified and plotted. Sequences of the probes are shown. Another BBS probe and several other MGM probes generated similar profiles as those presented here for the two families respectively. **E:** Abundance of two BBS clones (A and B) through the generations, detected as in **C** and **D** with the indicated probes. **F:** A schematic representation of an isolated BBS-containing individual. The positions of sequence segments corresponding to the probes are also indicated.

particular aptamer family, that is, a group of aptamers sharing not only the same activity but also the same consensus sequence, in a SELEX experiment, we used a dot-blot assay to examine the sequence abundance in pools representing different generations. The consensus sequence of the BBS family is highly homologous, and a 17-mer probe extracted from the sequence of the most popular clone should detect almost all family members at the hybridization stringency used here. Because the spacing between G-triplets of the MGM family varies considerably, it is not possible to find a sequence or a small set of sequences long enough to

cover all family members. For the MGM family, we designed a few short sequences as representative probes, each containing two G-triplets separated by one or two non-G nucleotides, and added two additional unspecified nucleotides (mix of all four) to each side. In all of the blot analyses, the 5' constant region was probed to normalize the signals relative to the total population. As shown in Figure 1C, D, the relative abundance of BBS and MGM sequences (genotypes) as a function of generations paralleled the enrichment of the respective B52 and filter-binding activities (phenotypes). Different MGM probes detecting different and possibly overlapping sub-

groups of the MGM family produced similar profiles as the one shown here. Because of the sequence variation among the family members and the resulting uneven hybridization efficiency, it is not possible to quantify precisely the abundance of a family, or of one family relative to the other, in one generation. However, the relative abundance detected by the same probe in different generations clearly showed the trends and patterns of enrichment.

The overall fitness of a particular aptamer clone within an aptamer family is not determined solely by the "true" aptamer moiety in the randomized region. The entire sequence of a BBS-containing individual can be regarded as comprising typically five segments arranged in a tandem array (I through V, depicted in Fig. 1F). In addition to the BBS sequence (III), the constant regions (I and V) and the randomized regions flanking the BBS (II and IV), as well as features resulting from the interactions among them and the BBS moiety, may also contribute to the capability of a particular individual to compete with the rest of the pool. To study the relationship between aptamer clones (a clone is the set of individuals derived from a single progenitor sequence in the original pool) and the family to which they belong, we followed the enrichment of two BBS clones, and compared their profile to that of the BBS family. One of them, designated Clone A, and represented by BBS #8 (Shi et al., 1997), has multiple isolates from the selected pools and showed highest affinity to B52 among the isolated individuals tested (Shi et al., 1997). Its abundance through the generations was revealed by using a probe derived from the sequence of its non-BBS portion in the randomized region (indicated in Fig. 1F) in a dot-blot analysis. As shown in Figure 1E, the enrichment profile of this particular BBS clone is very different from that of the general BBS family. Its abundance was relatively low when the family reached its peak at Generation 6, but kept increasing monotonically over the next few generations in spite of the decline of the family as a whole. Another clone, designated Clone B, was isolated only once as BBS #23 (Shi et al., 1997). The shape of its trajectory roughly parallels that of the general BBS family (Fig. 1E). Therefore, the evolutionary trajectory of a particular clone may or may not agree with the dynamic profile of the aptamer family to which it belongs, presumably as a result of internal competition within the family.

The BBS and MGM sequences each bind specifically to one of the two targets present during the binding and partitioning steps. The equilibrium dissociation constant of BBSs to B52 is 20–50 nM (Shi et al., 1997). It is difficult to determine the K_d for the MGM family's binding to the nitrocellulose filter, as the filter is far from being saturated even at the highest RNA concentration we were able to produce. With different RNA concentrations ranging from 10 to 1,000 nM, we estimated the K_d of the MGMs to the filter to be roughly in the range

of 10–100 μ M. The presence of four G-triplets in a typical MGM also suggested that the MGM sequences may form G-quartets (Kim et al., 1991) or similar structures that are active in binding to the filter. The original binding buffer used in our selection experiment contained 150 mM potassium acetate to mimic the intracellular concentration of potassium. This condition might have helped stabilize such structures (Katahira et al., 1995). In a binding reaction of MGM to the filter, we replaced K^+ with the same concentration of Na^+ or Li^+ , and observed a decrease of the filter-binding activity almost to the level of the unselected pool (data not shown).

In vitro population control: Manipulating the frequency of selected sequences

In conventional SELEX, the selection, either positive or negative, is based on the activities of RNA molecules; however, identification of the MGM consensus made it possible to selectively eliminate, or restrict, MGMs from the pool according to their sequences. To implement this negative selection scheme, we developed a population restriction method utilizing the enzymatic activity of RNase H, which digests RNA molecules selectively marked by complementary DNA sequences. A similar procedure had been used to implement a parallel destructive algorithm to solve an instance of the satisfiability problem related to the "knight problem" in chess (Faulhammer et al., 2000). There, as in many other molecular computation experiments, the nucleotide sequence of a nucleic acid molecule, as the physical embodiment of the information encoded therein, was used in itself as the criterion for selection. In contrast, here the sequence of the selected aptamers is treated as their genotype, as opposed to their binding activity phenotype. This scheme of negative selection according to genotype provides an independent means to control the relative size of different aptamer families, which are determined mostly by their phenotype in conventional SELEX. As shown in Figure 2A and discussed later, this method can be used to implement a general approach to select multiple, and possibly all, aptamer families to a mixture of targets.

To target RNase H activity specifically to the desired sequences in a heterogeneous sequence pool, the marking DNA (deoxyoligoribonucleotides) must be sufficiently long for it to anneal to the target sequence with specificity and for the resulting duplex to be recognized by the enzyme. We optimized parameters concerning marking DNA length and RNAase H type and conditions using three representative MGM sequences and a random control sequence, before extending this method to the sequence pools. Marking DNA oligos from 6 to 21 nt in length were tested with the *Escherichia coli* RNase H at 37 °C or a thermostable RNase H (Hybridase) at 45 °C. We found that *E. coli* RNase H

is more efficient but less specific than the Hybridase. A marking oligo at least 8 nt in length is needed in a reaction with *E. coli* RNase H at 37 °C. For the Hybridase at 45 °C, the oligo should be 12 nt or longer (data not shown).

Multiple sequence alignment of the MGM family revealed two challenges. First, because the space between the G-triplets varies in both nucleotide number and identity, the consecutive consensus of the family is no more than 3–4 nt in length. A degenerate population of oligos with a single G-triplet in the middle would contain insufficient information to specify the family members. Second, poor sequence similarity among the family members necessitated a set of representative marking oligos, rather than only one oligo. This heterogeneity of marking oligos and their targets would result in high degree of cross-hybridization during the “ear-marking” annealing step of the RNase H treatment, thus decreasing specificity and efficiency of this process. Based on their frequency in the MGM family, a group of five octamers was first chosen, each containing two adjacent G-triplets separated by one or two other nucleotides, in the hope that this may cover the majority of the family. Oligonucleotides with sequences complementary to this set (the “Octamarkers” in Fig. 2B) were used in treatments by the *E. coli* RNase H. For the Hybridase, we appended two additional unspecified nucleotides to each side to create the “Dodecamarkers” (Fig. 2B). Repeated annealing and RNase H treatments were performed to allow for more complete targeting by hybridization.

RNase H treatments using the optimal conditions identified with single MGM sequences showed a significant decrease in pool size in the MGM-dominated pool G9 but not the unselected G0 (Fig. 2C). This result is confirmed by comparing the amount of amplifiable RNA after the treatment by quantitative RT-PCR (data not shown). Once treated, the selected pool became more resistant to a second treatment (data not shown). Family-specific probing revealed a decrease in the relative abundance of MGMs and a correlated increase in the relative abundance of BBSs after the treatments (Fig. 2D). After two consecutive restriction treatments, we cloned the treated pools. Sequences of a sample of 25 isolates from each pool are shown in Figure 2E. In both pools, the sequences can be categorized into the following five groups: the BBS family, the MGM family, the C-rich family, individuals having a shorter random region (possibly resulting from deletion during repeated amplification or cloning), and sequences not belonging to the above families. The ratio of BBS to MGM is dramatically increased in both pools compared to the untreated G9 pool. None of the surviving MGM isolates contained any consecutive sequence segment that would be perfectly matched by any marking oligo used. The relative proportions of the four non-MGM groups are nearly identical in both treated pools. However,

there are more MGM sequences in the *E. coli* RNase H-treated pool than in the one treated with the Hybridase, indicating that treatment with longer marking oligos at a higher temperature by the Hybridase is more effective. Furthermore, from the surviving MGMs, a new representative consensus can be extracted in which two adjacent G-triplets are separated by three A or T residues. Based on this consensus, a new marking oligo, NCW13 (5'-NNCCCWWWCCCNN-3', W = A + T), was designed and used together with the dodecamarkers in the treatment of another pool selected by a different target. This further improved the efficiency of MGM removal: Cloning and sequencing of the treated pool revealed that the MGM abundance was reduced from nearly 100% in the pool to 3% after the treatment.

More BBS individuals isolated after the majority of MGMs were selectively eliminated from the selected pools allowed us to further define the sequence requirements of B52 binding. In the larger set of BBS clones now available, the alignable sequence segment in the randomized region remains the same 17-mer stretch defined previously (Shi et al., 1997); however, now we can see that the degree of “conservation” or convergence is very uneven: Three highly variable islands at or around positions 2, 9, and 16 reside in the otherwise uniform consensus. The structural implications of these observations are under further investigation.

In addition to mining less prevalent RNA aptamer families from a selected pool, as illustrated in Figure 2A, our *in vitro* population restriction method can be extended to a general approach for generating different ligands to multiple targets in a mixture (see Discussion). To test the efficiency and versatility of this method, we removed not only the MGM family but also the BBS family, successively or simultaneously with the MGMs, from the B52-selected G9 pool and some of its derivatives. In contrast to the MGM family, the BBS family has a long and highly homologous consensus sequence, and can be marked by two 17-mer sequences with only three positions different between them. To further increase the specificity, the reaction was carried on at a higher temperature, 55 °C, when only BBSs were to be eliminated. Depletion of either or both families was confirmed by southern dot-blot analyses with the marking oligos as probes (Fig. 2F).

DISCUSSION

A solution to the matrix-binding problem in SELEX

SELEX is an effective method to generate novel affinity reagents, or ligands, for targets ranging from small chemicals to macromolecular complexes. The most im-

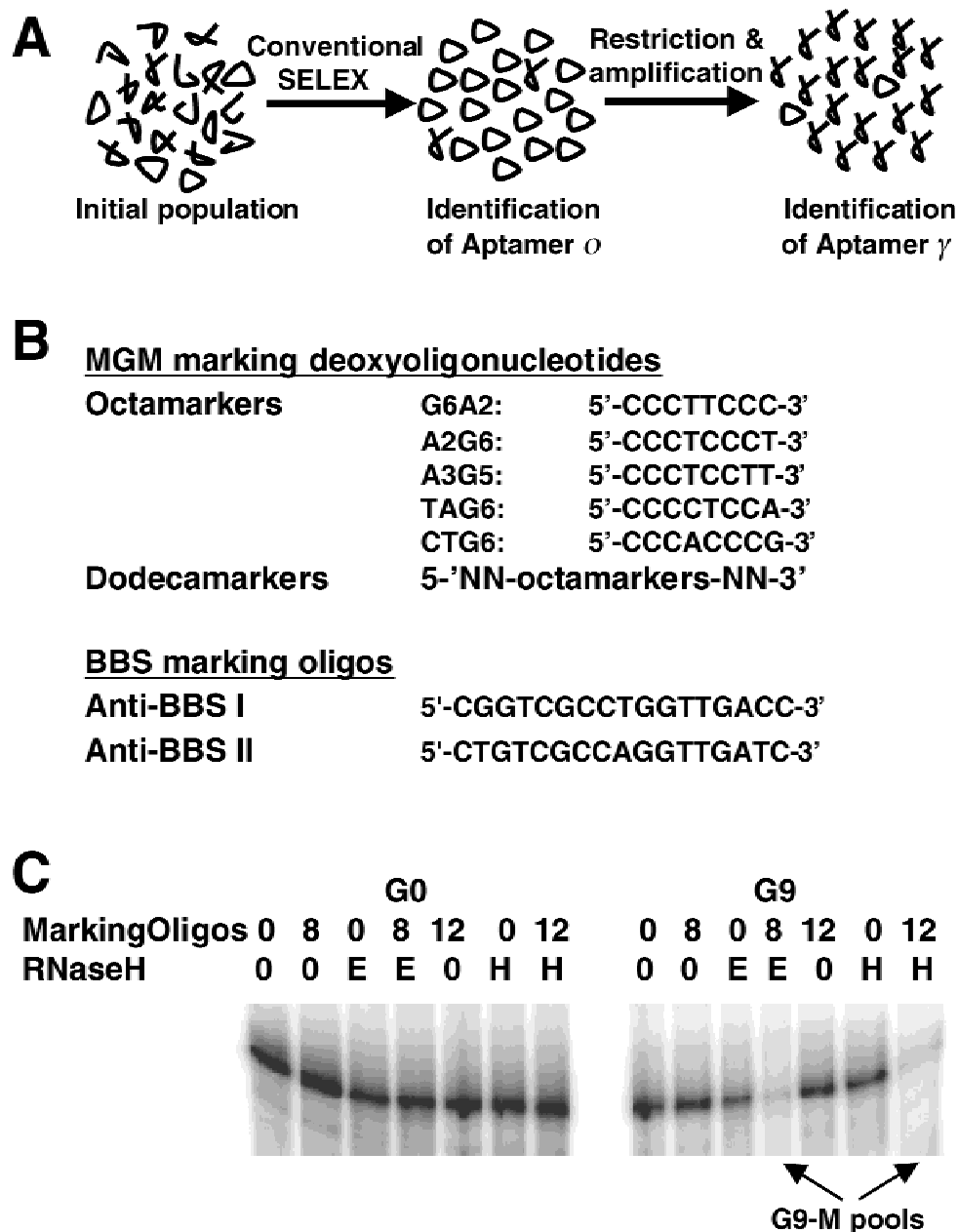
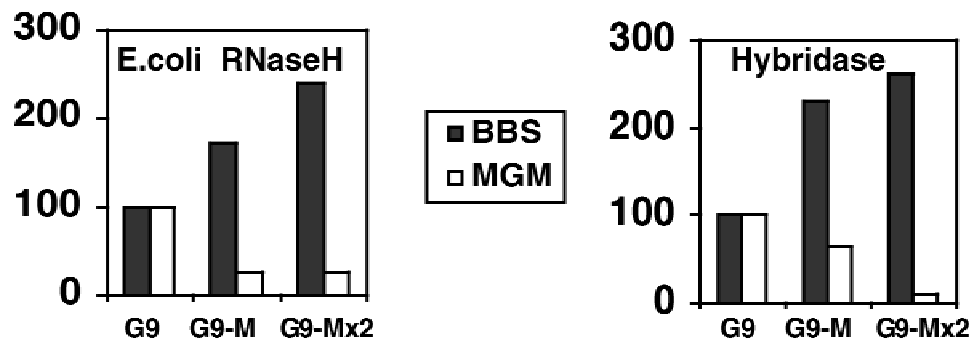


FIGURE 2. A method for population control in selected pools. **A:** A schematic diagram depicting the simplest case of exhaustive selection with two targets. Conventional SELEX results in the identification of Aptamer α , which may or may not be desirable. The method developed here allows both aptamer families, α and γ , to be identified. **B:** List of marking oligos used in RNA restriction treatments. The Octamarkers and the Dodecamarkers were used to mark the MGMs for elimination; the anti-BBS I and II were markers for the BBS family. **C:** Effects of MGM restriction treatment demonstrated by the decrease of the size of MGM-dominated G9 pool. The G9 RNA pool and the unselected G0 pool were treated with MGM marking oligos and RNase H. "8" and "12" indicate the Octamarkers and the Dodecamarkers, respectively. "0" indicates controls without markers added. "E" and "H" indicate the *E. coli* RNase H and the Hybridase, respectively. Decrease of total population size after treatment is revealed by the reduced intensity of the bands representing the pools. **D:** Effects of selective RNA restriction treatment revealed by the change of the corresponding family size. The abundance of the aptamers in the pools (in the form of DNA) was determined by the integrated densitometric data of southern dot-blot analysis using oligos with family consensus or representative sequences as probes. The levels of the BBS and MGM aptamer families detected by the probes in the untreated G9 pool were each set at 100. The actual ratio of BBS to MGM in this pool is about 1:10. G9-M is the G9 pool after a single treatment to remove MGMs. G9-Mx2 is the G9 pool that underwent two such treatments. **E:** Effects of MGM restriction treatment demonstrated by the sequences of isolated individuals. Samples from two double-treated pools are shown. Defining features of each family are highlighted in the sequences. **F:** RNA restriction treatments targeted to MGMs alone, BBSs alone, or both MGMs and BBSs. Abundance of the aptamer families is plotted as in **D**. G9-B is the G9 pool after a treatment to remove BBSs alone. G9-M&B is the G9 pool after a treatment to remove both MGMs and BBSs. (Figure continues on facing page.)

D**Relative Abundance of BBS or MGM****E****G9-Mx2 (E. coli. RNaseH treated)****BBS**

UGGUCAACCAGGCGACCGCCACCCGCGCGCnAUACCU
 UGAGUGGUCAACCAGGCGACCCUCGUACUACCAGCUUCU
 UUCCCAAGAGUGGUCAACCUGGCGACCCUCGCCUAUCU
 UUCCCAAGAGUGGUCAACCUGGCGACCCUCGCCUAUCU

MGM

GACAAAUGGGCAAGGGUAGGGUAGCUAGGGGAUGUCGCA
 GUUUGGGUUAGGGCGGGCAACGCAUUGUCUUCUGAAC
 GUUUGGGUUAGGGCGGGCAACGCAUUGUCUUCUGAAC
 CAACGUAGAACCAUAAGGGUAUGGGUAGGGUUCACAGGG
 UGCUAGCACUGCAUACAGGGUAGGGUAGGGUUCAGGGUC
 GGUUGGCAAAACUAGGGUAGGGUUGGCAAAAAUAUUGCCU
 GCUGAGCAGUACAGUCAGGGUAAAAGGGGAUUCGGGA
 GCUAGGGAAAGGGUAGGGGAGAACAGCACAUACCCUCCCA
 GUUGGGUCAAGGGUAGGGUUCUGCAAAUUCUGCCAGCCCC

C-rich

UUGUUACUACCCCAACGACCCUGUACCCUGACCCUU
 CAGAAUCUUAGCGAAGCCUACCCAGCAUCUUAGUACCGG
 CCUGACCCCAUGGACCCCGCCCAUGUCCACCCUCGCGCA
 AGCUUAUCGUGAGCAGCCUAGCAACACAAUAUGGGCCCGU

Other

AGAUUAACGACAUGGGUCGAGUUGGACUGACUCAGUGAGU

Short randomized region

UUAC

No randomized region

G9-Mx2 (Hybridase treated)**BBS**

UACCCCAACCCGAUUAGCUUGGGGAAAUCAACUCGGCGACag
 AAUCAACCAGGCGACACGCCUCGGCAGCUCGCCGCGUGU
 ACACGAUCAACCUGGCGACAGUGAGCCUAUGUUGGCCAAC
 CUAACUACUAAGAAGCAUCCUGUACGAUCAACCUGGCGACag
 UGGUCAACCAGGCGACCGCCACCCGCGCGCAGAAUACCU
 UGGUCAACCAGGCGACCGCCACCCGCGCGCAGAAUACCU
 UGGUCAACUAGGGCGACCGCCUCAAGAUUUACUUCUUCUGU

MGM

GUUAGGGUAAGGGUAGGGAAUCCAUCCGUCGAAUCUCUAGU
 GUUGGGUUAGGGUUCAGGGAAUGCCAGAACAAACACAAACG

C-rich

CCACCAACAAGGACGCACGCCCCCAACCCACUCGGAACC
 AACAUACGUCGCACACUCUCAUAACCCACUAAACCACGU
 GAAUCACCCACGACCCGCAUUGCUCGGAAUACAA
 GGAUCCGCCACUUUCCUGCUACUCAAUACCCGCCUAUAC
 GAAUUCAGAUAGCCACCCGCGCGCAGAAUACCUAGUACU
 CAUUCAGCACCAACGGAGAACACGUCAAAUCCCCUCAC

Other

ACCUAGAUGAACCAUCAACGUUGUCAACGCGCUAGGACGA
 UCAAAUUAACUUUCGCGCUGGCCUACAGCGGACUCACCG

Short randomized region

UUAC

AUCUAUGUACACAGGGGAUGUCGCA

GAAGUUC

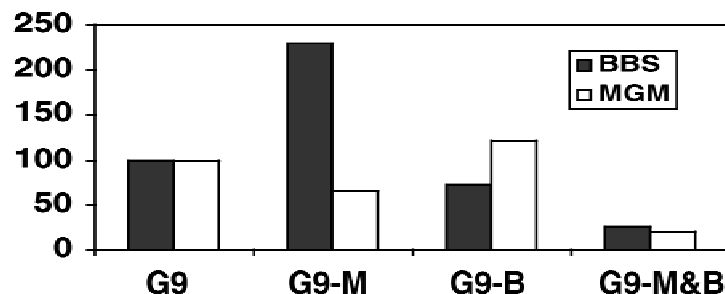
F**Relative Abundance of BBS or MGM**

FIGURE 2. Continued.

portant step in the SELEX procedure is the partitioning step in which the target-binding candidates are physically separated from the nonbinding ones. As demonstrated here and documented elsewhere (Conrad et al., 1996), the immobilization matrix used for the partitioning step can act as an unintended target to select unwanted aptamers, and these aptamers can come to dominate the selected pool. Several methods are commonly used to avoid this problem. First, matrix-binding aptamers may be eliminated by negative selection against the matrix, that is, collecting the fraction not bound to the matrix. In our experience, this is most effective when performed right after the recovery of bound species and before amplification. However, in early cycles of selection, when the copy number of each individual sequence is low, this may increase the chance of stochastic events in which particular sequences are lost. Second, because the number of binding sites on the surface of the matrix is fixed, increasing the amount of the intended target may change the ratio of the matrix-binding aptamers to target-binding aptamers in favor of the recovery of the latter in the partitioning step. However, the number of binding sites on the matrix can be extremely large compared to that on even the highest concentrations of targets, rendering this method ineffective. Third, alternating use of different types of matrix should theoretically eliminate molecules binding to either type of matrix. This method becomes less effective if the chemical difference between the commonly used matrices is not sufficient to discriminate against binding by less specific means such as hydrophobic interactions. In one of our experiments, we alternatively used nitrocellulose filter and agarose beads as the matrix, but saw no significant decrease of matrix-binding molecules in the selected pool (H. Shi & J.T. Lis, unpubl.). The sequence-directed RNA restriction method presented here provides another approach to the solution of this problem. We have demonstrated its efficiency in reducing the MGM family that binds to the nitrocellulose filters.

In our experience, MGM enrichment occurred in different selection experiments involving nitrocellulose filters, and when a large (10^{16}) initial candidate pool was used, even repeated negative selection against filter binding was not effective enough to get rid of them. To solve this problem during and after selection, we developed a scheme that combines three techniques. First, RNase H treatment is effective at eliminating identified MGMs for which marking DNA oligos of sufficient length could be designed. Second, after removing the majority covered by a representative set of marking oligos, the remaining filter-binding individuals can be more effectively eliminated by conventional negative selection. Third, because the MGM family's affinity to nitrocellulose filters is compromised in the absence of potassium ion, substituting sodium for potassium helps eliminate MGM sequences.

The process of molecular selection and amplification

The process of in vitro selection and amplification is driven by differential replication rates induced by a specific partitioning scheme that is dictated by the target. Interesting issues concerning the complex process of molecular evolution may emerge more clearly, and thus be better studied, in this kind of well-defined microsystem. Because of its simple formal representation as two-dimensional graphs, RNA secondary structure is often designated as the phenotype of the molecule in many theoretical studies (Fontana & Schuster, 1987; Schuster, 2001). However, the SELEX method makes it possible to study a functional trait as its phenotype: The binding activity of RNA to the target(s) is a direct function of three-dimensional molecular shape. In our study, the two distinct binding specificities, to the B52 protein and to the nitrocellulose filter, define two families of individuals in the RNA pool. For each family, the profiles of enrichment as represented by the binding activity or the sequence abundance generally agree with each other. MGMs were selected during a short contact time with the filter, that is, the time when the protein-binding mixture was passed through the filter, a kinetically less favorable condition. This may account for the time it took MGMs to reach an exponential rate of enrichment shown in the middle portion of the curve. The rate of enrichment of MGMs slowed in the later cycles, as one would predict there to be progressively less selective pressure between the winning individuals and the bulk pool. In contrast to the MGM family, BBS abundance as well as its activity increased dramatically during the early cycles, especially between Generations 3 and 4. This family was enriched initially at a higher rate than the MGM family, perhaps because its members have higher affinity for B52 than MGMs for the filter. Equilibrium binding for B52 in the early cycles was also facilitated by a long incubation with gentle thermocycles (Shi et al., 1997). The BBS trajectory turned downward later, presumably as a result of being out-competed by the MGM family. The MGMs might have been enriched by the high density of binding sites on the filter, expected to be much higher than the amount of B52 binding sites available. In addition, MGMs have lower informational complexity and therefore are statistically more abundant in the original pool than the BBS.

Exhaustive selection against a mixture of targets

In a multitarget selection experiment, it is important to ensure that different ligands can be selected to each of the targets in the mixture. The difference between the rate of enrichment of different ligand families during a selection experiment against multiple targets is deter-

mined by the differences in concentrations and binding affinities of ligands for the targets, as well as differences in other unidentified factors. All these variables (save the relative target concentration in some cases) are unknown at the onset of the experiment. Therefore, to ensure that aptamers against all potential targets are isolated, it is desirable to find an independent and convenient method that adjusts the rate of enrichment of any aptamer family.

The *in vitro* population control method described herein provides a means to isolate aptamers against multiple, and possibly all, targets in a mixture. When the selection yields one predominant sequence family, we can utilize its sequence information to selectively target and destroy this family in the pools representing any generation. Subsequent selective cycles allow the family with next highest rate of enrichment to dominate the population and be identified. Once the new family is identified, its sequence information can be used again to bring about its destruction and elimination from the population of selected RNAs. The simplest case of an exhaustive selection against two targets illustrated in Figure 2A can be extended to more complicated schemes. Ideally, iterative use of this method would ensure that any ligand family regardless of its initial relative rate of enrichment would have an opportunity to dominate the population and thus be identified in successive selection cycles. Our design and use of the NCW13 marking oligo is an application of this principle on subgroups within the MGM family. In our example, BBS and MGM represent both extremes of the consensus aptamer sequence in terms of their length and degree of homology. Successful ablation of both from the pool, either successively or simultaneously, demonstrated the efficiency and versatility of this strategy.

MATERIALS AND METHODS

Oligonucleotides, RNA, and RNase H

The deoxyoligonucleotides used as “markers” in the RNA restriction treatments and probes in the southern blot analyses were synthesized at the 50-nmol scale by Operon Inc. (Alameda, California). RNA was synthesized using the T7-MEGAscript™ *in vitro* transcription kit (Ambion, Inc., Austin, Texas) according to the manufacturer’s instructions. The [α - 32 P] UTP- (NEN Life Science Products, Boston, Massachusetts) incorporated RNA was prepared using the MAXI-script™ *in vitro* transcription kit (Ambion, Inc.) according to the manufacturer’s instructions. The *E. coli* RNase H and the thermostable Hybridase were purchased from Epicentre Corp. (Madison, Wisconsin).

Band shift assay

Band shift assay was performed as previously described (Shi et al., 1997). Each binding reaction contained 10 pmol of B52 protein.

Filter-binding assay

The setting of the filter-binding assay was modified from that of Wong and Lohman (1993) using BA-S85 nitrocellulose membrane and NA45 DEAE membrane (Schleicher & Schuell Inc., Keene, New Hampshire). Different amounts of RNA were normalized into 20 μ L of 1 \times binding buffer (50 mM Tris-HCl, pH 7.6, 0.5 mM MgCl₂, 150 mM salt—KCl, NaCl, or LiCl as indicated) to final concentrations in the range of 10–1,000 nM. The [α - 32 P] UTP-labeled RNA was used as a tracer and mixed with 400–600-fold excess amount of cold RNA. Each 20 μ L RNA preparation also contained 10 μ g of yeast tRNA as a nonspecific competitor. The RNA preparations were incubated at 80 °C for 5 min and then cooled down to the ambient temperature and kept there for 30 min before loading to the filter. The signals were quantified using the ImageQuant software (Molecular Dynamics Inc., Sunnyvale, California). The fraction of RNA retained on the nitrocellulose membrane (using the sum of signals on both membranes as total) was calculated and plotted.

Southern dot-blot analysis

For each DNA pool representing a generation during the selection experiment, about 20–40 ng (0.3–0.6 pmol) sample in 2 μ L were prepared in 1 \times SSC and applied onto a piece of Hybond-XL membrane (Amersham Pharmacia Biotech UK Ltd., Buckinghamshire, UK). To prepare the probes, 20 pmol oligos were labeled with T4 polynucleotide kinase (New England Biolabs, Inc., Beverly, MA) in 20 μ L reactions in the presence of 60 μ Ci [γ - 32 P]-ATP (10,000 Ci/mmol; NEN Life Science Products). When detecting the constant region of the pool, 300 pmol cold probe oligo was added to the [γ - 32 P]-labeled probe. Hybridization was carried out at 37 °C overnight. The membrane was washed three times, each time in 0.1 \times SSC/0.2% SDS at room temperature for 10 min. The signals on the membrane were collected by exposing to a phosphorimage screen, and quantified using ImageQuant software (Molecular Dynamics Inc.).

RNA restriction treatment

In an RNase H reaction, for each unit of RNase H, 200 ng RNA (6 pmol) and 150 (30 \times 5) pmol marking oligos were used in 20 μ L RNase H buffer (50 mM Tris-Cl, pH 7.4, 100 mM NaCl, 10 mM MgCl₂). The reaction mix was preincubated at 72 °C for 3 min before adding the enzyme and transferring to a lower temperature specified by the type of RNase H and the T_m of the marking oligo. When the *E. coli* RNase H was used, the reaction was always held at 37 °C for 30 min and then stopped by addition of 10 mM EDTA. When the Hybridase was used, the reaction was held for 1 h at a temperature 2–3 °C below the T_m of the marking oligo. For example, 45 °C for the dodecamarkers, and 55 °C for the anti-BBS I and II. When more than one oligo were included in a single reaction, the lowest T_m was used to determine this temperature.

The effect of this restriction treatment was assessed in three assays. First, the effect on the entire pool was examined. A small amount (about 20–50 fmol) of radiolabeled tracer made from transcription with the same template preparation was included in the reaction, and a fraction of the reaction

mix was then run on a 8% polyacrylamide, 7 M urea gel. The radioactive bands on the gel were visualized by a phosphor-imager. Second, the treatment was also assessed from the decrease of the relative abundance of the sequence family under restriction, using southern dot-blot analysis (see above). Finally, a representative sample of a treated pool was cloned and sequenced to show its composition.

ACKNOWLEDGMENTS

The authors thank Drs. Volker Vogt and Mariana Wolfner for their critical comments on the manuscript. This work was supported by National Institutes of Health Grant GM 40918-1 to JTL.

Received June 29, 2002; returned for revision August 7, 2002; revised manuscript received August 14, 2002

REFERENCES

- Brown D, Gold L. 1995. Selection and characterization of RNAs replicated by Q beta replicase. *Biochemistry* 34:14775–14782.
- Brown D, Gold L. 1996. RNA replication by Q beta replicase: A working model. *Proc Natl Acad Sci USA* 93:11558–11562.
- Conrad RC, Giver L, Tian Y, Ellington AD. 1996. In vitro selection of nucleic acid aptamers that bind proteins. *Methods Enzymol* 267:336–367.
- Ellington AD, Szostak JW. 1990. In vitro selection of RNA molecules that bind specific ligands. *Nature* 346:818–822.
- Faulhammer D, Cukras AR, Lipton RJ, Landweber LF. 2000. Molecular computation: RNA solutions to chess problems. *Proc Natl Acad Sci USA* 97:1385–1389.
- Fontana W, Schuster P. 1987. A computer model of evolutionary optimization. *Biophys Chem* 26:123–147.
- Gold L, Polisky B, Uhlenbeck O, Yarus M. 1995. Diversity of oligonucleotide functions. *Annu Rev Biochem* 64:763–797.
- Hanczyc MM, Dorit RL. 1998. Experimental evolution of complexity: In vitro emergence of intermolecular ribozyme interactions. *RNA* 4:268–275.
- Joyce GF. 1994. In vitro evolution of nucleic acids. *Curr Opin Struct Biol* 4:331–336.
- Katahira M, Moriyama K, Kanagawa M, Saeki J, Kim MH, Nagaoka M, Ide M, Uesugi S, Kono T. 1995. RNA quadruplex containing G and A. *Nucleic Acids Symp Ser* 34:197–198.
- Kim J, Cheong C, Moore PB. 1991. Tetramerization of an RNA oligonucleotide containing a GGGG sequence. *Nature* 351:331–332.
- Morris KN, Jensen KB, Julin CM, Weil M, Gold L. 1998. High affinity ligands from in vitro selection: Complex targets. *Proc Natl Acad Sci USA* 95:2902–2907.
- Pan W, Craven RC, Qiu Q, Wilson CB, Wills JW, Golovine S, Wang JF. 1995. Isolation of virus-neutralizing RNAs from a large pool of random sequences. *Proc Natl Acad Sci USA* 92:11509–11513.
- Ringquist S, Jones T, Snyder EE, Gibson T, Boni I, Gold L. 1995. High-affinity RNA ligands to *Escherichia coli* ribosomes and ribosomal protein S1: Comparison of natural and unnatural binding sites. *Biochemistry* 34:3640–3648.
- Robertson DL, Joyce GF. 1990. Selection in vitro of an RNA enzyme that specifically cleaves single-stranded DNA. *Nature* 344:467–468.
- Schuster P. 2001. Evolution in silico and in vitro: The RNA model. *Biol Chem* 382:1301–1314.
- Shi H, Hoffman BE, Lis JT. 1997. A specific RNA hairpin loop structure binds the RNA recognition motifs of the *Drosophila* SR protein B52. *Mol Cell Bio* 17:1649–1657.
- Tuerk C, Gold L. 1990. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* 249:505–510.
- Tuerk C, MacDougall S, Gold L. 1992. RNA pseudoknots that inhibit human immunodeficiency virus type 1 reverse transcriptase. *Proc Natl Acad Sci USA* 89:6988–6992.
- Wilson DS, Szostak JW. 1999. In vitro selection of functional nucleic acids. *Annu Rev Biochem* 68:611–647.
- Wong I, Lohman TM. 1993. A double-filter method for nitrocellulose-filter binding: Application to protein–nucleic acid interactions. *Proc Natl Acad Sci USA* 90:5428–5432.