RNA (1998), *4*:1165–1175. Cambridge University Press. Printed in the USA.

Recombination, RNA evolution, and bifunctional RNA molecules isolated through Chimeric SELEX

DONALD H. BURKE[1] and JOHN H. WILLIS
Department of MCD Biology, University of Colorado, Boulder, Colorado 80309-0347, USA

ABSTRACT

Exchange of RNA structural domains through recombination can be used to engineer RNAs with novel functions and may have played an important role in the early evolution of life. The degree of function an RNA element retains upon recombination into a new sequence context is a measure of how deleterious or beneficial recombination will be. When we fused pairs of aptamers previously selected to bind coenzyme A, chloramphenicol, or adenosine, the chimerae retained some ability to bind both targets, but with reduced binding activity both in solution and on affinity resins, probably due to misfolding. Complex populations of recombined RNAs gave similar results. Applying dual selection pressure to recombined populations yielded the combinations that were best suited to binding both targets. Most reselected RNAs folded into the active conformation more readily than chimerae built from arbitrarily chosen aptamers, as indicated both by solution K_d measurements and affinity resin binding activity. Deletion/selection experiments confirmed that the sequences required for binding are fully contained within the respective domains and not derived from interaction between the domains, consistent with the modular architecture of their original design. The combinatorial nature of the recombination methods presented here takes advantage of the full sequence diversity of the starting populations and yields large numbers of bifunctional molecules (10^6 to more than 10^{12}). The method can be easily generalized and should be applicable to engineering dual-function RNAs for a wide variety of applications, including catalysis, novel therapeutics, and studies of long-range RNA structure.

Keywords: adenosine; aptamers; coenzyme A; chloramphenicol; fitness landscapes; macromolecular design; RNA structure; RNA World

INTRODUCTION

Recombination has been a powerful evolutionary force throughout biological evolution, as seen in the numerous amino acid sequence motifs that have been shuffled from one context into another to form natural mosaic proteins (for review, see Doolittle, 1995). Protein segments are routinely recombined to engineer chimeric proteins with novel functions. Evolutionary rates can be greatly accelerated when recombination is allowed in addition to point mutations (Holland, 1992; Stemmer, 1994a, 1994b; Crameri et al., 1998). Pre-mRNA processing is thought by some to have played a role in the natural evolution of these mosaic proteins through "exon shuffling" (Gilbert & Glynias, 1993; de Souza et al., 1996), and a similar process has been used to shuffle coding domains in filamentous bacteria to produce combinatorial peptide libraries for phage display (Fisch et al., 1996).

Large functional RNAs, such as self splicing introns and the RNA components of RNaseP, ribosomes, and spliceosomes, also tend to be organized into secondary structural domains. These domains assemble through base pairing, tetraloop–receptor interactions, counterion condensation onto divalent metal ion cores, adenosine platforms, H-bonding through the ribose 2′ hydroxyl, base triples, and other interactions (Pyle & Green, 1995; Cate et al., 1996; Strobel & Doudna, 1997). Normal intramolecular interactions that assemble the catalytic core of group I ribozymes also assemble these domains when they are present as separated pieces (Doudna & Cech, 1995). Recognition of the pre-tRNA substrate by the catalytic RNA subunit of RNaseP also involves tertiary RNA–RNA contacts (Yuan & Altman, 1995; Harris et al., 1997; Lee et al., 1997). However, there is little evidence that multidomain RNAs evolved piecemeal through recombination among the individual structural domains. Indeed, there are few examples of natural mosaic RNAs that would indicate whether recombination has been as powerful a force in RNA evolution as it has been for proteins.

Reprint requests to Donald H. Burke at his present address: Department of Chemistry, Indiana University, Bloomington, Indiana 47405, USA; e-mail: dhburke@indiana.edu.

On the other hand, there are compelling reasons to explore the potential for using recombination to engineer new RNAs. In addition to the expanding catalogue of functional RNAs found in nature, there are now hundreds to thousands of small functional RNA modules from in vitro selections (the SELEX protocol). The number of possible pairwise fusions among these elements grows with the square of the number of elements available, and each combination carries with it the possibility of additional functions arising from synergy between the recombined elements. For example, fusing an mRNA-cleaving ribozyme with signals that increase in vivo lifetime or direct the ribozyme to co-localize with its target could stimulate in vivo cleavage efficiency and lead to novel therapeutics (Westaway et al., 1995). RNAs that bind two substrates (e.g., FMN and NADH) could catalyze a reaction (e.g., electron transfer) between them (Ellington, 1993). "Molecular switches" have been generated by incorporating aptamer motifs into hammerhead ribozymes such that cleavage activity is regulated by addition of the appropriate binding partner for the aptamer (Tang & Breaker, 1997). Finally, if there was ever a time in which cellular or pre-cellular metabolism was driven by RNA catalysis (RNA World), recombination can be expected to have occurred with an appreciable rate. Synergy between the recombined products could have yielded RNAs with new functions, but only if each of the recombined elements retained significant function in the new sequence context.

We are interested in modeling random recombination among functional RNA elements as a means to generate RNAs with novel functions. When two RNA elements are thrown together, either through natural recombination or intentional engineering, it is difficult to know beforehand whether the joining will diminish the activity of one or both of the recombining partners. Each parental molecule can harbor "negative elements" that disrupt the other's activity by favoring alternative, inactive structures. Furthermore, not all of the binding modes represented in diverse aptamer populations will be suitable for evolving molecules with more advanced functions (Burke & Gold, 1997), particularly when the reactive moiety in the bound target is located within a good binding site for RNA. For example, an aptamer to flavin mononucleotide (FMN) might bury the flavin moiety, making it unavailable to participate in redox chemistry. Because there is no way of knowing a priori which variant from a given population would provide the best starting point for subsequent selections, there is a clear need for methodologies that allow rapid screening to identify the most fruitful combinations.

The "Chimeric SELEX" method described here simulates random recombination among functional RNAs derived from RNA populations with 70 or 80 positions of random sequence. Specifically, we make use of previously described RNA populations that bind chloramphenicol [Cam (Burke et al., 1997)], adenosine (Burke & Gold, 1997), or coenzyme A [CoA (Burke & Hoffman, 1998)]. These aptamers were selected from random initial pools that contained either 70 ("70N") or 80 ("80N") positions of random sequence. Among the diverse sequences from the Cam-binding populations ("70Cm" and "80Cm"), there were some that contained the motif in Figure 1A, which has a weak but suggestive similarity to the portions of 23S rRNA involved in Cam binding and peptide bond formation (Burke et al., 1997). All of the isolates from the adenosine-binding population (80S), which was selected to recognize S-adenosyl methionine (SAM), conformed to the sequence motif shown in Figure 1B. This element has also been isolated in selections for RNAs that bind ATP (Sassanfar & Szostak, 1993) and NAD^+ (Burgstaller & Famulok, 1994), and it recognizes many other adenosine derivatives. Of the two CoA-binding populations, the "80CoA" sequences were highly diverse, whereas the "70CoA" population was dominated by a motif that recruited secondary structural elements and 12 specifically required unpaired nucleotides from the primer binding sites (Fig. 1C). This RNA element also binds many adenosine derivatives, but its specificity differs significantly from that of apta-

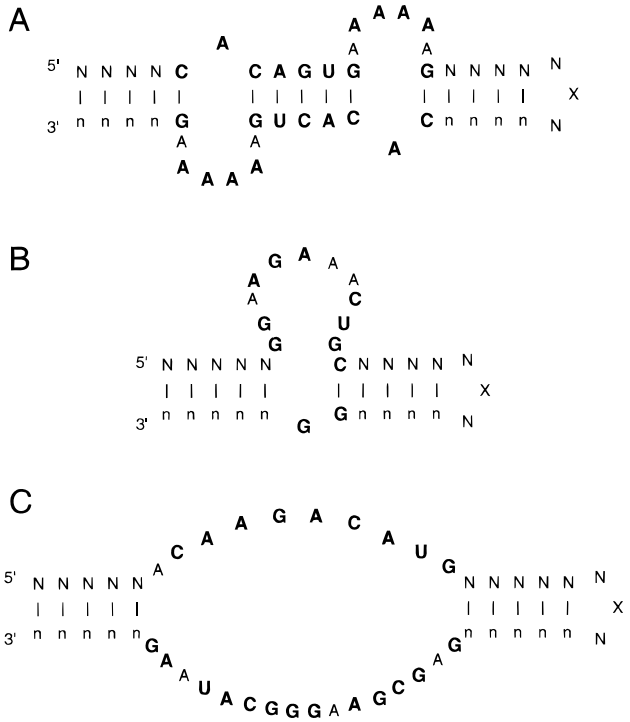

FIGURE 1. Schematic representation of some of the aptamer modules used in this study. **A:** One of the Cam-binding aptamer motifs present in both population 70Cm and 80Cm (Burke et al., 1997). **B:** Conserved adenosine-binding aptamer motif present in population 80S (Sassanfar & Szostak, 1993; Burgstaller & Famulok, 1994; Burke & Gold, 1997). **C:** Conserved CoA-binding aptamer motif found in population 70CoA (Burke & Hoffman, 1998). N, any nucleotide; n, its base pairing complement; x, variable number of nucleotides of any sequence. Nucleotides that were intolerant of mutations or were highly conserved are in bold print.

mers from the 80S population. The 70CoA and 80CoA populations also contained low frequencies of RNAs that required intact CoA for elution from the CoA resin and were not eluted by AMP alone (Burke & Hoffman, 1998).

The recombination event that initiates chimeric SELEX fuses every member of one population with function "X" with every member of another population with function "Y." In this paper, the binding activities of previously described aptamers in their original sequence context are compared with their activities in the context of larger, recombined molecules. We show that most such combinations are mildly to severely deleterious, but that additional cycles of selection/amplification enrich the population for those that best retain both activities. The deleterious effects of chimerae formation appear to be due largely to misfolding of the active element, although other factors are clearly involved in some instances. The process can be readily generalized to other systems and, because it is combinatorial, it generates vast numbers of bifunctional molecules. Similar bifunctional populations should be amenable to evolving sophisticated RNAs built from nucleic acid modules with two or more functions.

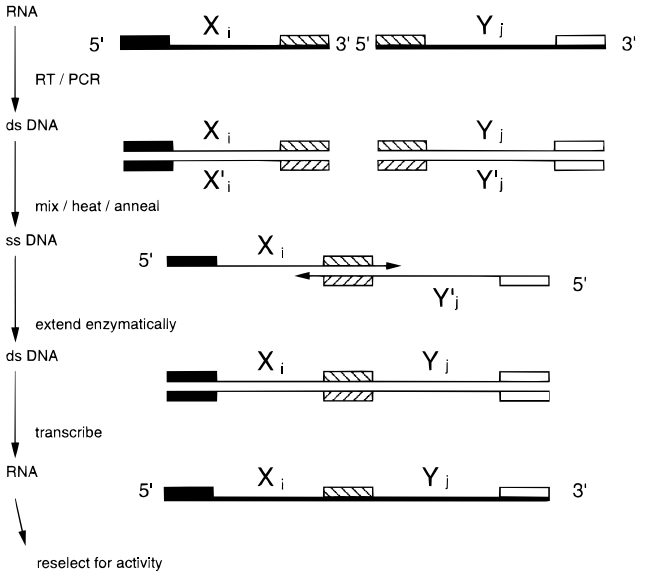

FIGURE 2. Overlap extension strategy for generating chimeric RNAs. Rectangles indicate primer binding sequences. Thick lines, RNA; thin lines, DNA.

RESULTS

Effect of chimerae formation on pairs of selected aptamers

To test the effect of arbitrarily recombining RNA aptamers, individual 70N and 80N isolates from the populations described above were joined using an overlap extension reaction shown in Figure 2. RNAs in each populations contain specific primer binding sequences at their 5′ and 3′ ends to facilitate pool amplification. The sequence at the 3′ ends of the 70N populations is identical to that at the 5′ ends of the 80N populations. When 70N and 80N RNAs are reverse transcribed, converted to double-stranded DNA, mixed, and denatured, the 70N top strand can reanneal with the 80N bottom strand. The recessed 3′ ends can then be extended with any DNA polymerase to yield 150N, chimeric molecules that contain the binding elements from each parental sequence. The two domains have been, in effect, recombined through their common sequence elements.

Table 1 shows the 22 pairwise combinations generated through the overlap extension reaction. Aptamers from several different selections were fused in order to minimize artifacts that could be introduced from any particular set. Series "Q" combines 70Cm and 80CoA individuals, series "R" reverses the order (70CoA with 80Cm), and series "S" combines 70Cm with 80S individuals. These particular aptamers were chosen because they were among the most active in binding to and eluting from their respective cognate resins (Burke et al., 1997; Burke & Hoffman, 1998). Each RNA was assayed for its ability to be retained

TABLE 1. Combinations of individual aptamers fused into chimerae.[a]

	70N parental aptamer					
80N parental aptamer	70Cm#6	70Cm#15	70Cm#45	70CoA#5	70CoA#16	70CoA#68
80CoA#18	Q#1	Q#3				
80CoA#32	Q#2	Q#4				
80Cm#19				R#1	R#4	R#7
80Cm#30				R#2	R#5	R#8
80Cm#33				R#3	R#6	R#9
80S#7	S#1	S#4	S#7			
80S#20	S#2	S#5	S#8			
80S#22	S#3	S#6	S#9			

[a]Sequences of parental aptamers are published (Burke & Gold, 1997; Burke et al., 1997; Burke & Hoffman, 1998).

on the corresponding affinity resins and to be eluted by free target molecules under conditions similar to those used in the original selections. The total amount of RNA that could be recovered in this way was generally reduced to between 10% (indistinguishable from background) and 70% (equivalent to original activity within experimental error) of that of the parental aptamer (Fig. 3). Not all combinations were affected to the same degree. For instance, 80Cm#30 and 80Cm#33 gave similar yields when assayed alone, but when combined with the same three 70CoA domains, the 80Cm#30-derived chimerae (R#2, R#5, and R#8) lost considerably more activity than did the 80Cm#33-derived chimerae (R#3, R#6, and R#9). The loss of binding activity is most likely caused by the formation of alternative base pairing patterns that preclude folding into the active conformation ["alternative conformation Hell" (Uhlenbeck, 1995)], such that the active conformation constitutes a smaller fraction of the population of chimeric molecules than it does in the parental 70N or 80N RNA. Alternatively, subtle conformational changes may distort the binding site and directly reduce the binding affinity. Either way, loss of affinity resin-binding activity directly reduces the selective fitness of these RNAs.

To test whether misfolding is responsible for reduced resin-binding activity, ATP-binding activities of several RNAs were measured in solution using spin filtration (Jenison et al., 1994). Various concentrations of refolded RNA were incubated with [α-^{32}P]ATP and filtered through 30-kDa molecular weight cutoff membranes. Free ATP passes through the membrane, whereas RNA-bound ATP does not. Dissociation constants (K_ds) were determined graphically by plotting the fraction of the ra-

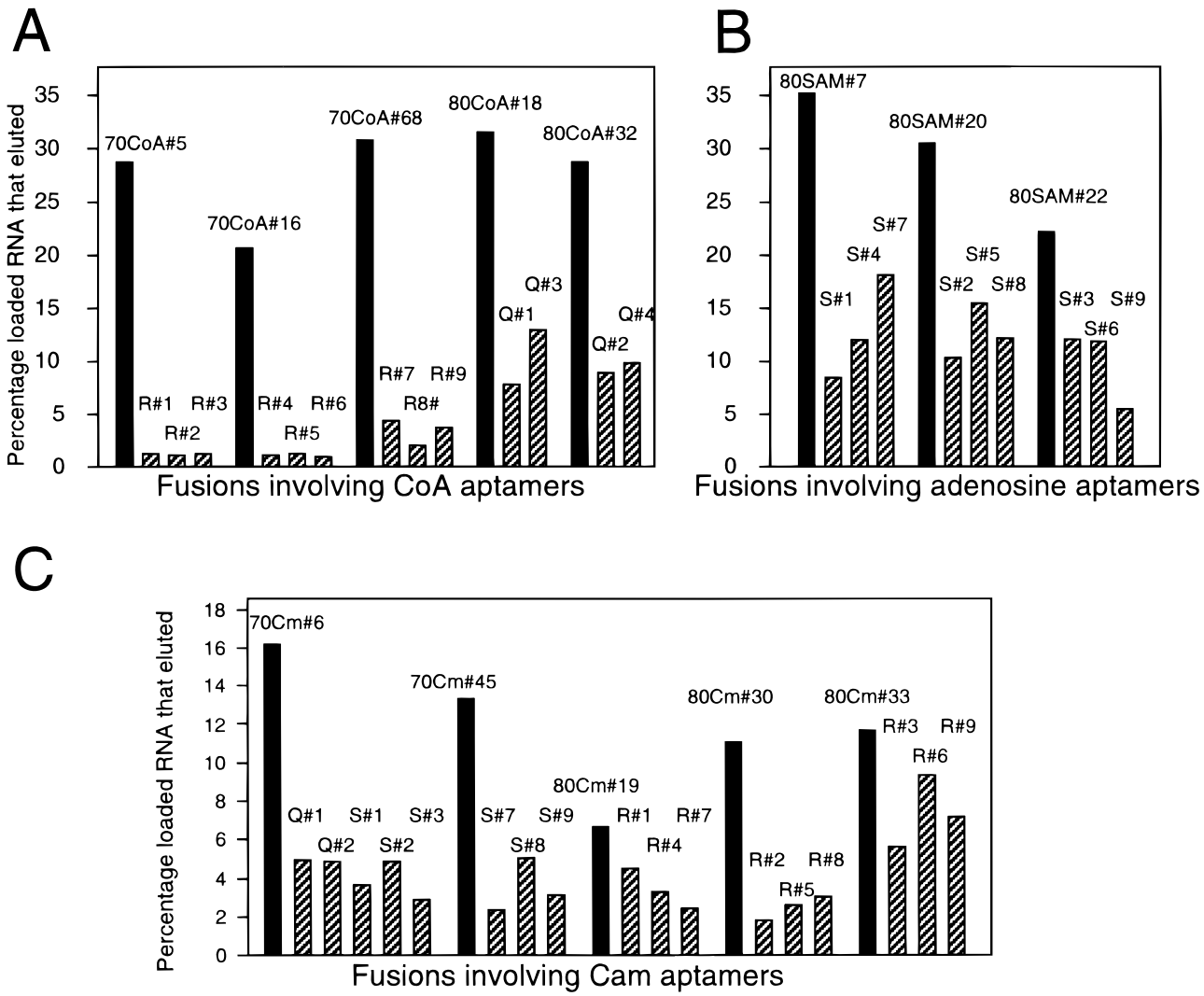

FIGURE 3. Effect of chimerae formation on binding activity, as measured by the fraction of all counts recovered from a cognate affinity resin that eluted upon addition of free target. Affinity resins were derivatized with (**A**) coenzyme A, (**B**) 5′ AMP (C-8 linkage), or (**C**) chloramphenicol. Solid bars, the activity of the aptamer alone; hatched bars, activities of the corresponding chimerae (compare Table 1).

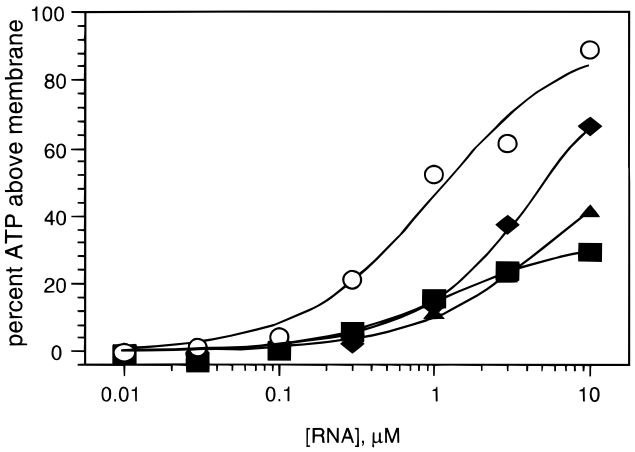

FIGURE 4. [α-^{32}P]ATP solution binding curves for 80N isolate SAM#20 (circles) and three SAM#20-derived chimerae (triangles, S#2; squares, S#5; diamonds, S#8). Theoretical curves were optimized using the equation given in Materials and Methods.

dioactivity retained above the membrane as a function of RNA in the assay and taking K_d to be equal to the ATP concentration at half the calculated plateau value (Fig. 4) as described in Materials and Methods. The 80S isolate SAM#20 binds with a K_d of 1.1 $\pm$ 0.3 μM, similar to that reported for comparable RNAs containing essentially the same binding element (Sassanfar & Szostak, 1993). Chimeric construct S#2, which contains the SAM#20 sequence, gave a similar K_d, but a calculated saturation plateau that was reduced to around one-third (36%) that of SAM#20. This is similar to the ratio observed in the fractions eluted from the resins, and is consistent with the notion that the differences are due to increased misfolding within the larger RNA. In contrast, both S#5 and S#8 gave calculated plateau values similar to that of SAM#20, but their K_ds were shifted approximately five- to sixfold. Their decreased elution activity may be due to some factors other than misfolding.

Combining populations of aptamer domains

The overlap extension strategy was used to generate three populations of 150N chimeric molecules: 70Cm/80CoA, 70CoA/80Cm, and 70Cm/80S (designated "I," "F," and "P," respectively). Within the parental populations, no individual sequence made up more than 15% of the total number sampled, and 40–80% of the sequences were encountered only once within each set. On this basis, we estimate that there are between 10^3 and 10^6 different sequences in each parental population and 10^6–10^{12} unique combinations that can be formed between them. Primer extension analysis on the products of the overlap extension reation was used to determine the degree to which these combinations had been saturated. End-labeled primers complementary to the 70N 5′ fixed region were extended to the 3′ end of either the 70N or 80N domain, depending on whether recombination had taken place. The products were analyzed by gel electrophoresis, and the radioactivity in the two bands was quantified (data not shown). By this measure, 20–50% of the input DNA had been fused into chimerae of the appropriate size. Because approximately 0.5 μg of gel-purified parental DNA went into the reaction (6×10^{14} DNA strands), all of the combinations of individuals from each aptamer population are expected to be represented in the chimeric populations.

Reselected bifunctional populations

The chimeric populations were assayed for their abilities to be retained on both affinity resins and to be eluted by the respective free target molecules as above. Again, the amount of RNA that could be recovered in this way was reduced significantly relative to the parental populations (Fig. 5). However, after a small number of SELEX cycles, nearly full activity returned to both populations. Sequences of individual isolates from the final selected populations were diverse and showed no sign of having converged on a "winning" sequence (24 sequences from population F, 11 from I, and 11 from P, Fig. 6). The adenosine-binding motifs that had dominated the 70CoA and 80S populations are easily recognized in some of the sequences from the final selected chimeric populations, as are 23S rRNA-like sequences in some of the 70Cm and 80Cm domains. However, most isolates present novel sequence motifs in one or both domains, as expected in chimerae derived from diverse populations.

There were no duplicates of entire 150N molecules, but some of the 70N or 80N domains matched those in

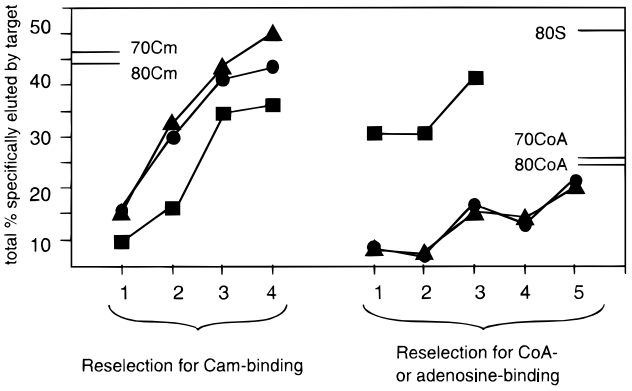

FIGURE 5. Percentage of input RNA eluted from cognate resins during each cycle of the reselection first for Cam-binding (left), then for AMP- or CoA-binding (right). Numbers on the *x*-axis indicate SELEX cycles. Triangles, population "F" (70CoA/80Cm); circles, population "I" (70Cm/80CoA); squares, population "P" (70Cm/80S). Aggregate activities of the parental, monofunctional RNA pools are shown by horizontal lines at the side.

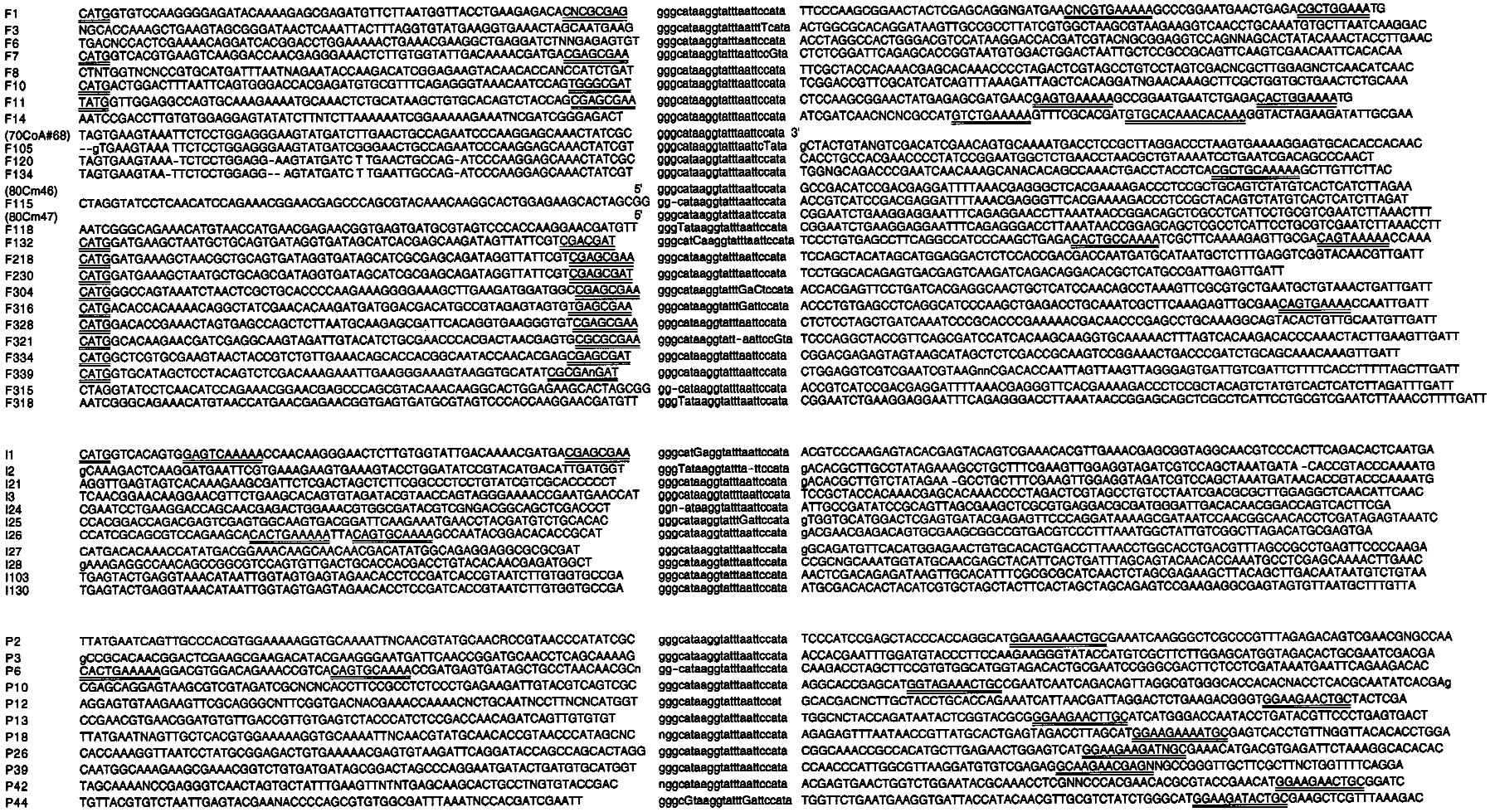

FIGURE 6. Sequences of the bifunctional aptamers from chimeric SELEX, divided into populations F (70CoA/80Cm), I (70Cm/80CoA), and P (70Cm/80S). Sequences corresponding to the motifs shown in Figure 1 are underlined. Three 70CoA (Burke & Hoffman, 1998) or 80Cm (Burke et al., 1997) aptamers that are homologous to the respective domains of particular chimeric isolates are shown for comparison. Nearly identical domains are aligned with gaps indicated with dashes. Overlapping sequences used in fusing the two domains are shown in lowercase except where mutated (uppercase for base changes, dashes for deletions). 5′ and 3′ primer binding sites are shown above. "n" refers to unreadable sequence positions.

other isolates or in aptamer sequences identified in the original CoA or Cam selections (Burke et al., 1997; Burke & Hoffman, 1998). The 70Cm domains of I#103 and I#130 are identical, and three point mutations distinguish the 70CoA domains of isolates F#218 and F#230, both of which conform to the motif shown in Figure 1C. The 70CoA domains of F#105, F#120, and F#134 differ by four to five point mutations from each other and from isolate 70CoA#68, which does not share the canonical motif in Figure 1C. Similarly, the 80Cm domain of F#115 differs from aptamer 80Cm#46 in five positions, whereas the 80Cm domain of F#118 differs from aptamer 80Cm#47 in three positions. Chimeric isolate I#1 was the only one that appeared to have both CoA- and Cam-binding sequences within one domain, suggesting that the original 70Cm isolate may have also contained the canonical CoA-binding element found in the 70CoA population. The functional significance of this apparent evolutionary convergence was not tested directly. The presence of 21 point mutations (including single-nucleotide insertions or deletions) within the overlapping primer binding sequences used to fuse the two domains suggests a net background mutation rate of approximately 1.9% per position subsequent to chimerae formation.

Binding activities of reselected aptamers

To determine whether the binding elements within the reselected chimeric RNAs were sensitive to sequence context in the same fashion as when individual aptamers are randomly recombined, the binding activities of the 150N chimerae were compared with those of separately amplified and transcribed 70N and 80N domains (Fig. 7A). Each set of three RNAs from each of six reselected chimerae was assayed for binding and elution from both cognate affinity resins. As expected from the way in which they were constructed, much more RNA from the 70N domain than from the 80N domain was specifically eluted from the CoA affinity resin (Fig. 7B). In sharp contrast to the arbitrarily recombined aptamer isolates, the fraction of the reselected 150N chimeric RNA that eluted specifically—and presumably folded properly—was approximately equal to or much greater than that of the 70N domain. Domain selectivity was reversed on the AMP affinity resin, whereas the 70N domains showed no binding activity and the 80N domains were readily eluted from the resin, again, as expected. Although the full-length P#42 RNA eluted (and presumably folded) as well as its 80N domain, the fraction of the full-length P#6 that could be eluted was much less than for its 80N domain, similar to the behavior observed for randomly recombined domains (Fig. 3). Binding and elution activity patterns were less consistent on the Cam affinity resin. Isolates F#7 and F#134 eluted equivalently within either the 150N or

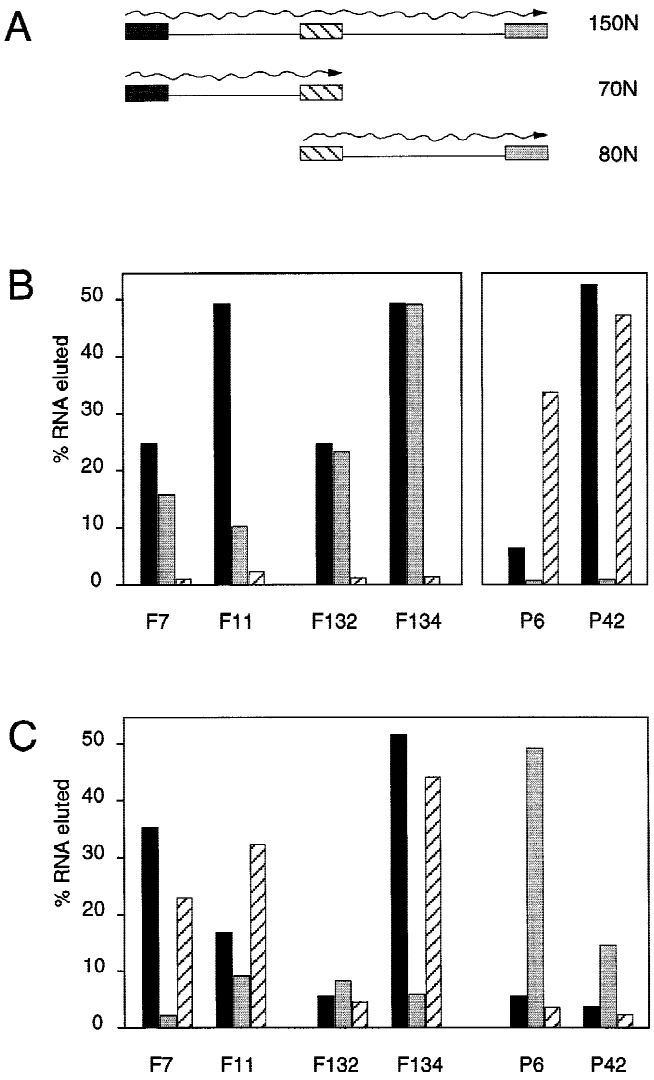

FIGURE 7. Elution activities of reselected 150N chimeric RNAs and their separately amplified and transcribed 70N and 80N domains (**A**). The assay was performed on (**B**) CoA or AMP affinity resins, and (**C**) Cam affinity resin.

80N sequence context (Fig. 7C), but F#11 was less active than its 80N domain and P#6 and P#42 were much less active than their 70N domains.

The solution binding affinity of chimeric isolate P#42 was indistinguishable from that of its adenosine-binding 80N domain within experimental error (K_d = 1.1 ± 0.2 and 1.5 ± 0.5 μM, respectively), as were those of chimeric isolate P#6 and its 80N domain (4 ± 2 and 2.1 ± 0.6 μM) (Fig. 8). Plateau values were also indistinguishable within each pair. Thus, for isolate P#42, chimera formation had no discernible effect on folding, as assayed both in solution and on the affinity resin. For isolate P#6, in contrast, chimera formation affected only the resin-binding activity, perhaps by sterically hindering access of the resin-bound AMP to the aptamer binding pocket.

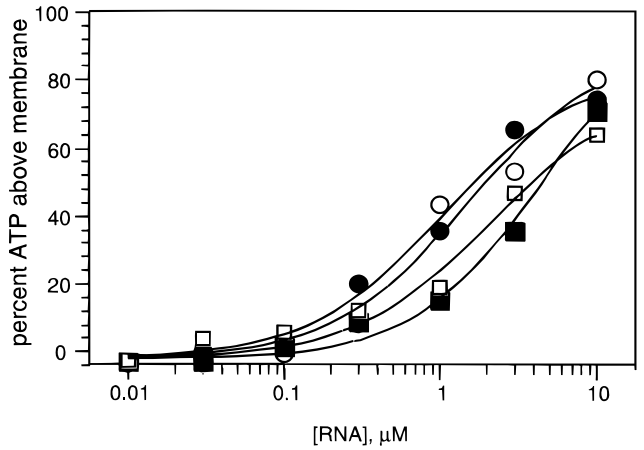

FIGURE 8. [α-^{32}P]ATP solution binding curves for chimeric isolates P6 (filled squares), P#42 (filled circles), and their respective 80N domains (open squares and circles, respectively). Theoretical curves were optimized using the equation given in Materials and Methods.

Binding activity arises from non-overlapping sequences

The reselected chimerae were built from domains that had been preselected to bind their respective targets. The presence of well-characterized binding sequences in several of the isolates, along with the binding data in Figures 7 and 8, suggest that, for many reselected RNAs, this modular architecture has been preserved. However, it is possible that, during the course of the original recombination or the subsequent reselection steps, some RNAs mutated to form binding sites that required elements from both domains.

To test this possibility, we determined the 5′ and 3′ boundaries of the essential binding elements using deletion/selection experiments (Wallis et al., 1995; Burke & Hoffman, 1998). End-labeled RNA was partially alkaline hydrolyzed and applied to the appropriate affinity resins. Fractions collected during the wash or elution portions were pooled, normalized for radioactivity, and size-separated by gel electrophoresis. Typical data are shown in Figure 9A and summarized in Figure 9B. In every case, the binding elements are fully contained within the expected parental 70N or 80N domain. Furthermore, in those chimerae containing previously recognized binding sequences, the observed functional boundaries correspond to the edges of the sequences required to form the expected binding element (70CoA domains of F#7 and F#11 for CoA-binding, 80S domain of P#42 for AMP-binding).

The particular sequences that contribute to conformational entropy and stable misfoldings are often referred to as "negative sequence elements." Chimeric isolate P#6 appears to contain negative sequence elements near its 5′ end that inhibit activity in the full-length isolate. Whereas the original alkaline hydrolysis ladder was essentially uniform, RNA molecules truncated from the 5′ end to near position 65 were preferentially lost during the wash steps. There is a dramatic increase in signal for RNAs with larger 5′ deletions, and another jump in signal for RNA with 5′ termini between positions 90 and the functional boundary at position 128. These data all point to the presence of "negative elements" in the 70N domain 5′ to position 65 that inhibit AMP-binding activity derived from the 80N domain.

DISCUSSION

Chimeric SELEX in theory and practice

Sequence diversity following in vitro selection is greatly reduced relative to that at the beginning of the selection. Recombination among preselected elements in a two-stage selection restores nonrandom sequence diversity, increasing the available sequence space from 10^{15} molecules to 10^{30} molecules in the limit of noninteracting recombined elements. Such recombined populations may be able to access nucleic acid activities that are much more rare than those typically found by in vitro selections.

We have modeled the process of recombining RNA modules in the presence and absence of subsequent selective pressures. Most of the arbitrary combinations retained reduced levels of activity from each parental element, whereas additional cycles of selection quickly returned aggregate activity levels to near normal in recombined populations. Vast numbers of bifunctional molecules are produced in this way, because $10^{(x+y)}$ unique combinations can be generated between two populations containing 10^{x} and 10^{y} unique sequences, respectively. Only a fraction of these $10^{(x+y)}$ combinations retain full activity in both domains, reducing the number of strongly bifunctional combinations to $10^{(x+y-a)}$. The value of "*a*" will be a function of the particular sequences and selection criteria involved. In the experiments presented here, *a* appears to be between 1 and 3, given the rapidity with which binding activity returned in the chimeric populations during their reselection. Thus, chimeric aptamer populations built from two pools containing 10^{6} unique sequences each are expected to contain on the order of 10^{9} to 10^{11} different highly active bifunctional molecules. In the present case, such bifunctional populations would be good starting points for selecting catalytic RNAs that use both small molecules as substrates.

Hybrid RNAs are finding many uses, some of which are noted in the Introduction. The considerations noted above may be broadly applicable, because efforts to engineer dual-function RNAs may generally suffer from loss of activity in the recombined or assembled RNA elements, and screening paired combinations individually is laborious. Most strategies that fuse RNA modules will benefit from including a broad spectrum of

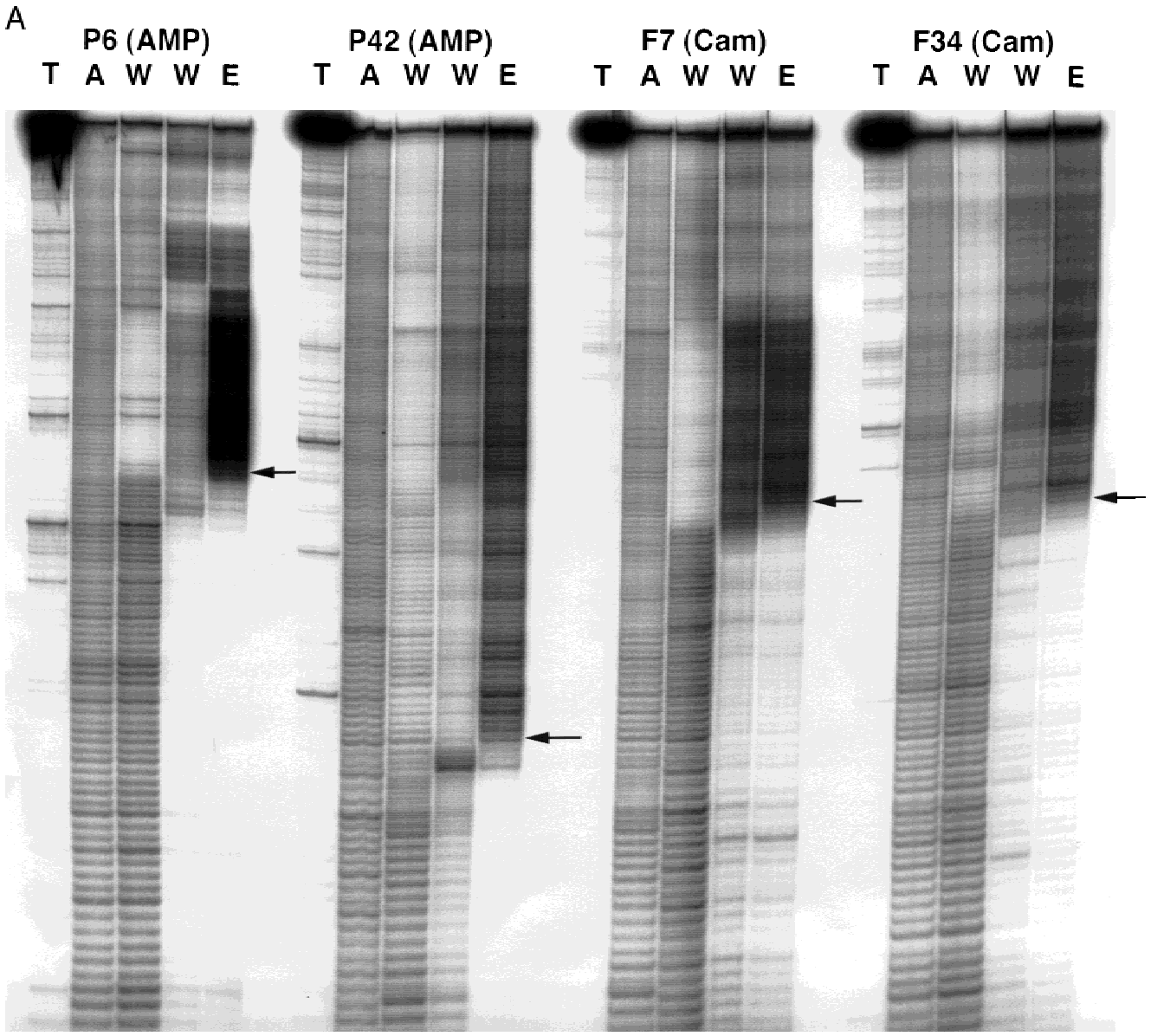

FIGURE 9. Determining the functional boundaries of binding elements within five chimeric isolates. **A:** Autoradiography of gel showing sample boundary data for four of these five RNAs using 3′ end-labeled RNA for the isolates and affinity resins indicated above, grouped into sets of five lanes each (marked T, A, W, W, E above each lane). Lane 1, size marker ladder generated by digesting the RNA with ribonuclease T1; lane 2, input alkaline hydrolysis ladder; lane 3, pooled first four wash fractions; lane 4, pooled last six wash fractions; lane 5, pooled elution fractions. Arrows to the right of each set indicate locations of 5′ boundaries. **B:** Summary of all boundary data, showing segments required for binding Cam (shaded boxes), CoA (cross-hatched boxes), or AMP (black boxes), with error bars to indicate uncertainty. Filled boxes under each RNA correspond to the primer binding sequences used in pool amplification or in constructing chimerae.

variants in the fusions and enriching for those that best retain all of the corresponding activities.

The deleterious effects of adding arbitrary sequences has also been observed for self-ligating ribozymes (Sebati et al., 1997). Aggregate catalytic rates for mixed populations was disrupted about fivefold, although the effect on particular isolates ranged from weak stimulation to 1,000-fold disruption in self-ligation rates. Those authors pointed out that a vigorous catalyst that occasionally spent time in inactive conformations would still perform well in activity assays and would be enriched in selections for catalysis. In contrast, aptamers must remain in the active conformation throughout the critical partition phase, when letting go of the target means being discarded with the wash.

"Seeding" and other methods

The overlap extension reaction at the heart of the recombinational step described here treats each functional RNA unit as a module by fusing the recombining partners end-to-end. Nucleic acid molecules with special or enhanced activities have been selected from populations seeded with variants of one particular aptamer motif, such as self-kinasing RNAs [derived from a pool of ATP-binding RNAs (Lorsch & Szostak, 1994)], self-alkylating RNAs [derived from a pool of biotin-binding RNAs (Wilson & Szostak, 1995)], or HIV-1 RT aptamers with exceptionally high affinity [from a pool of RT aptamers (Tuerk et al., 1992; C. Tuerk, pers. comm.)]. In each of these cases, random sequences were appended to a particular aptamer isolated in the first stage of the selection, usually with mutagenesis of the original aptamer. This approach is fundamentally different from the recombinational strategy we describe, because the "seeding" approach makes use of only one aptamer variant, ignoring other binding modes that may have yielded alternative solutions. Chimeric SELEX is unique in that it takes advantage of the full sequence diversity of preselected populations, rather than focusing on one or a small number of individual representatives. This can be particularly important if the predominant sequence element from the first stage of the selection is incompatible with the ultimate goal (e.g., an FMN aptamer that buries the flavin moiety and prevents it from participating in electron transfer).

Related strategies have also been described, such as PCR-mediated recombination (Judo et al., 1998) and "DNA shuffling" (Stemmer, 1994a). Both of these processes require that the exchanged elements share significant stretches of identical sequence, making them unsuitable for use with complex populations where recombination partners are exceedingly rare. Most aptamer populations could not be subjected to DNA shuffling or PCR-mediated recombination without destroying the motif and re-randomizing the population, because even representatives of a particular motif class often share only short stretches of primary sequence, and many selected populations contain diverse collections of motifs. Furthermore, in both PCR-mediated recombination and DNA shuffling, the recombinational end-points are unrelated to functional boundaries. Two or more populations can also be fused end-to-end at the DNA level through restriction sites, or at the RNA level by enzymatic ligation or through chemical linkers (T. Tarasow & B. Eaton, pers. comm.), given an appropriate means of recovering active fusions.

CONCLUSIONS

Joyce and colleagues have pioneered the use of point mutations introduced through error-prone amplification to optimize a sequence for a particular function or to evolve it to a closely related one (Beaudry & Joyce, 1992; Lehman & Joyce, 1993; Tsang & Joyce, 1994; Dai et al., 1995; Wright & Joyce, 1997). It should be similarly possible to employ recombinatorial methods to develop altogether new functions. Nature has made extensive use of genetic recombination to diversify protein function. Combining and recombining functional RNAs through Chimeric SELEX opens the door to generating numerous bispecific molecules whose overall functions are greater than the sum of their parts. As we shuffle functional domains from one context into another, we may be recapitulating in vitro a major form of evolution used by ribo-organisms during the early evolution of life.

MATERIALS AND METHODS

Starting pools and overlap extension reactions

Isolation and initial characterization of the aptamer populations used here and conditions for in vitro transcriptions, RNA purification and folding, and selection for column binding and retention have all been described in detail (Burke & Gold 1997; Burke et al., 1997; Burke & Hoffman, 1998). Templates for overlap extensions were double-stranded, gel-purified PCR products amplified from the same plasmid preparations used for sequencing. 70N and 80N DNAs were mixed and subjected to 10 cycles of heating and cooling in the absence of PCR primers (30 s at 94 °C, 3 min at 55 °C, 1 min at 68 or 72 °C). An aliquot of the overlap extension reaction was removed for primer extension analysis as described in the text and the rest was amplified by PCR. For fusions of individual aptamer domains, the overlap extension was performed in the presence of PCR primers, because quantifying reaction efficiency was not an issue.

Affinity resin binding and reselections

Surveys of affinity resin binding activities were performed by passing refolded, radiolabeled RNA over the appropriate affinity resins (bed volume = 125 μL), washing with 10 $\times$

125 μL binding buffer (50 mM Bis-Tris, pH 6.4, 200 mM NaCl, 10 mM $MgCl_2$ for AMP and CoA; concentrations doubled for Cam), and eluting with 6 $\times$ 125 μL binding buffer that contained free target (5 mM 5′AMP, 5 mM CoA, or 10 mM Cam). Radioactivity in each 125-μL fraction was determined by Cherenkov counting. For reselections, RNA was recovered by pooling the eluted fractions and precipitating with ethanol after adding 40 μg glycogen. Resuspended RNA was reverse transcribed, PCR amplified, and retranscribed for the next selection cycle. Bispecific aptamer populations from the chimeric SELEX were cloned into pUC18 or pERLAC (gift of Dr. D. Hoffman) for dideoxy sequencing.

Solution binding assays and boundary determinations

Solution binding assays and boundary determinations were performed as described previously (Burke & Hoffman, 1997), with minor modifications. For the solution binding assays using Microcon 30 molecular weight cutoff filters (Amicon), spin time was 30–40 s, such that approximately one-third of the reaction mix passed through the membrane. Total [α-^{32}P]ATP was kept constant at 5 nM as before, and K_ds were determined graphically from plots of the fraction of the radioactivity (δ) retained above the size fractionation membrane as a function of added RNA concentration ([RNA]), using the equation $(\delta) = P \times \{[RNA]/([RNA] + K_d) + Bk\}$, where P is the optimized plateau value and Bk is the background retention in the absence of RNA (usually around 5%). For boundary determinations using the deletion/selection method, approximately equal counts were loaded onto each lane.

ACKNOWLEDGMENTS

We thank Drs. Larry Gold, Britta Singer, and Patrick Allen for comments on the manuscript, and Lori Scates for technical assistance in the early phases of this work. Support was provided by fellowship CHE-9302453 from NSF to D.B. and by additional funds from NeXstar Pharmaceuticals, Inc. J.W. was supported by NIH grant GM19963 to Larry Gold. We also thank Dr. Gold for providing space within his laboratory for the execution of this work.

Received April 6, 1998; returned for revision April 21, 1998; revised manuscript received June 15, 1998

REFERENCES

Beaudry AA, Joyce GF. 1992. Directed evolution of an RNA enzyme. *Science 257*:635–641.

Burgstaller P, Famulok M. 1994. Isolation of RNA aptamers for biological cofactors by in vitro selection. *Angewandte Chemie-Int Ed English 33*:1084–1087.

Burke DH, Gold L. 1997. RNA aptamers to the adenosine moiety of S-adenosyl methionine: Structural inferences from variations on a theme and the reproducibility of SELEX. *Nucleic Acids Res 25*:2020–2024.

Burke DH, Hoffman DC. 1998. A novel acidophilic RNA motif that recognizes coenzyme A. *Biochemistry 37*:4653–4663.

Burke DH, Hoffman DC, Brown A, Hansen M, Pardi A, Gold L. 1997. RNA aptamers to the peptidyl transferase inhibitor chloramphenicol. *Chem & Biol 4*:833–843.

Cate JH, Gooding AR, Podell E, Zhou K, Golden BL, Szewczak AA, Kundrot CE, Cech TR, Doudna JA. 1996. RNA tertiary structure mediation by adenosine platforms. *Science 273*:1696–1699.

Crameri A, Raillard SA, Bermudez E, Stemmer WP. 1998. DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature 391*:288–291.

Dai X, DeMesmaeker A, Joyce GF. 1995. Cleavage of an amide bond by a ribozyme. *Science 267*:237–240.

de Souza SJ, Long M, Schoenbach L, Roy SW, Gilbert W. 1996. Intron positions correlate with module boundaries in ancient proteins. *Proc Natl Acad Sci USA 93*:14632–14636.

Doolittle RF. 1995. The multiplicity of domains in proteins. *Annu Rev Biochem 64*:287–314.

Doudna JA, Cech TR. 1995. Self-assembly of a group I intron active site from its component tertiary structural domains. *RNA 1*:36–45.

Ellington AD. 1993. Experimental testing of theories of an early RNA world. *Methods Enzymol 224*:646–664.

Fisch I, Kontermann RE, Finnern R, Hartley O, Soler-Gonzalez AS, Griffiths AD, Winter G. 1996. A strategy of exon shuffling for making large peptide repertoires displayed on filamentous bacteriophage. *Proc Natl Acad Sci USA 93*:7761–7766.

Gilbert W, Glynias M. 1993. On the ancient nature of introns. *Gene 135*:137–144.

Harris ME, Kazantsev AV, Chen JL, Pace NR. 1997. Analysis of the tertiary structure of the ribonuclease P ribozyme–substrate complex by site-specific photoaffinity crosslinking. *RNA 3*:561–576.

Holland JH. 1992. Genetic algorithms. *Sci Am 267*:66–72.

Jenison RD, Gill SC, Pardi A, Polisky B. 1994. High-resolution molecular discrimination by RNA. *Science 263*:1425–1429.

Judo X, Wedel AB, Wilson C. 1998. Stimulation and suppression of PCR-mediated recombination. *Nucleic Acids Res*. Forthcoming.

Lee Y, Kindelberger DW, Lee JY, McClennen S, Chamberlain J, Engelke DR. 1997. Nuclear pre-tRNA terminal structure and RNase P recognition. *RNA 3*:175–185.

Lehman N, Joyce GF. 1993. Evolution in vitro of an RNA enzyme with altered metal dependence. *Nature 361*:182–185.

Lorsch JR, Szostak JW. 1994. In vitro evolution of new ribozymes with polynucleotide kinase activity. *Nature 371*:31–36.

Pyle AM, Green JB. 1995. RNA folding. *Curr Opin Struct Biol 5*:303–310.

Sassanfar M, Szostak JW. 1993. An RNA motif that binds ATP. *Nature 364*:550–553.

Sebati PC, Unrau PJ, Bartel DP. 1997. Accessing rare activities from random RNA sequences: The importance of the length of molecules in the starting pool. *Chem & Biol 4*:767–774.

Stemmer WP. 1994a. DNA shuffling by random fragmentation and reassembly: In vitro recombination for molecular evolution. *Proc Natl Acad Sci USA 91*:10747–10751.

Stemmer WP. 1994b. Rapid evolution of a protein in vitro by DNA shuffling. *Nature 370*:389–391.

Strobel SA, Doudna JA. 1997. RNA seeing double: Close-packing of helices in RNA tertiary structure. *Trends Biochem Sci 22*:262–266.

Tang J, Breaker RR. 1997. Rational design of allosteric ribozymes. *Chem & Biol 4*:453–459.

Tsang J, Joyce GF. 1994. Evolutionary optimization of the catalytic properties of a DNA-cleaving ribozyme. *Biochemistry 33*:5966–5973.

Tuerk C, MacDougal S, Gold L. 1992. RNA pseudoknots that inhibit human immunodeficiency virus type 1 reverse transcriptase. *Proc Natl Acad Sci USA 89*:6988–6992.

Uhlenbeck OC. 1995. Keeping RNA happy. *RNA 1*:4–6.

Wallis MG, von Ahsen U, Schroeder R, Famulok M. 1995. A novel RNA motif for neomycin recognition. *Chem & Biol 2*:543–552.

Westaway SK, Larson GP, Li S, Zaia JA, Rossi JJ. 1995. A chimeric tRNA(Lys3)-ribozyme inhibits HIV replication following virion assembly. *Nucleic Acids Symposium Series 33*:194–199.

Wilson C, Sostak JW. 1995. In vitro evolution of a self-alkylating ribozyme. *Nature 374*:777–782.

Wright MC, Joyce GF. 1997. Continuous in vitro evolution of catalytic function. *Science 276*:614–617.

Yuan Y, Altman S. 1995. Substrate recognition by human RNase P: Identification of small, model substrates for the enzyme. *EMBO J 14*:159–168.