Ribozyme motif structure mapped using random recombination and selection

QING S. WANG and PETER J. UNRAU

Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, B.C., V5A 1S6, Canada

ABSTRACT

Isolating the core functional elements of an RNA is normally performed during the characterization of a new RNA in order to simplify further biochemical analysis. The removal of extraneous sequence is challenging and can lead to biases that result from the incomplete sampling of deletion variants. An impartial solution to this problem is to construct a library containing a large number of deletion constructs and to select functional RNA isolates that are at least as efficient as their full-length progenitors. Here, we use nonhomologous recombination and selection to isolate the catalytic core of a pyrimidine nucleotide synthase ribozyme. A variable-length pool of ~10^8 recombinant molecules that included deletions, inversions, and translocations of a 271-nucleotide-long ribozyme isolate was constructed by digesting and randomly religating its DNA genome. In vitro selection for functional ribozymes was then performed in a size-dependent and a size-independent manner. The final pools had nearly equivalent catalytic rates even though their length distributions were completely different, indicating that a diverse range of deletion constructs were functionally active. Four short sequence islands, requiring as little as 81 nt of sequence, were found within all of the truncated ribozymes and could be folded into a secondary structure consisting of three helix–loops. Our findings suggest that nonhomologous recombination is a highly efficient way to isolate a ribozyme's core motif and could prove to be a useful method for evolving new ribozyme functions from pre-existing sequences in a manner that may have played an important role early in evolution.

Keywords: systematic deletion; random recombination; RNA motif isolation

INTRODUCTION

The striking modularity of RNA structures suggests that nonhomologous or random recombination may be a fruitful avenue to isolate and study functional RNA structures. This approach, which involves the deliberate jumbling of sequence elements, has been used previously to study protein and DNA function (Bittker et al. 2002, 2004) but has not been applied to an RNA system. Since RNA modularity ranges in scale from the macroscopic subunits of the ribosome (Yusupov et al. 2001) down to the subcomponents of small ribozymes (Szewczak and Moore 1995; Krasilnikov et al. 2004; Westhof and Massire 2004), this technique combined with a method to select functional sequence variants appears likely to be generally useful in the identification of RNA motif elements.

The mapping of a ribozyme core motif conventionally requires phylogenetic or chemical interference constraints to guide the manual removal of extraneous sequence. This process is both tedious and difficult to perform in an unbiased manner. If a phylogeny is available, the in vitro selection of short functional isolates from a large mutant library containing extensive chemically synthesized point deletions can quickly minimize a ribozyme motif (Chapple et al. 2003). Where phylogeny is lacking or is limited, the manual isolation of a functional RNA core motif can lead to the omission of important elements of the system (Khvorova et al. 2003; Uhlenbeck 2003). As modular RNAs having essential core sequence elements interspersed with regions of extraneous linking sequence often result from artificial in vitro selection experiments (Ekland et al. 1995; Illangasekare et al. 1997; Sabeti et al. 1997; Chapple et al. 2003; Knight and Yarus 2003) and are common in naturally occurring RNAs, the ability of random recombination and selection to isolate the core motif of a previously uncharacterized nucleotide synthase ribozyme was of great interest to us.

A pyrimidine nucleotide synthase ribozyme family originally isolated from a pool containing 228 nt of random sequence (Unrau and Bartel 1998) appeared likely to be found in a modular arrangement because of the length of

the initial library (Sabeti et al. 1997). Moreover, this family B pyrimidine nucleotide synthase ribozyme lost function when truncated at either end of its sequence, making it an ideal target to test the ability of random recombination to extract a ribozyme's core motif. Fragments of this ribozyme were generated by digesting its DNA genome. These blunt-ended pieces were then randomly religated together, resulting in a pool containing a larger number of deletions, inversions, and translocations of the initial ribozyme sequence. In vitro selection on this pool was then used to efficiently isolate the essential elements of the ribozyme. As this approach lacks the bias intrinsic to the manual truncation techniques normally applied to isolate functional domains, we propose that it could be applied to study a variety of artificial or naturally occurring RNA motifs.

RESULTS

Previously, an in vitro selection for pyrimidine nucleotide synthase ribozymes isolated three [4S]U synthase ribozyme families called A, B, and C (Unrau and Bartel 1998). These ribozymes when ligated to a 3'-terminal phosphoribosyl 1-pyrophosphate (pRpp) all react with free 4-thiouracil to produce a tethered 4-thiouridine product. Further selections and kinetic isotope studies have characterized the secondary structure and mechanism of family A. Glycosidic bond formation by this family involves highly dissociative chemistry as compared to many protein-mediated nucleotide synthesis reactions, and appears mechanistically similar to that of some glycosylases (Chapple et al. 2003; Unrau and Bartel 2003). In contrast to family A, family B requires manganese in addition to magnesium ions to function, implying that its chemistry may be of a different character than that of family A.

Attempts to manually truncate the 271-nt-long family B isolate B.01 (GenBank accession number: AF051883) from the 5'-terminus indicated that the conserved 5' sequence required for PCR amplification was essential for ribozyme function, and, indeed, removing the 5'-triphosphate with calf intestinal phosphatase inactivated the ribozyme. As the 3'-terminus, which was derivatized with pRpp, must be proximal to the ribozyme active site, it appeared that this construct could only be truncated internally. To obtain the core sequence of the B.01 ribozyme, we used nonhomologous recombination to generate a double-stranded DNA (dsDNA) pool based on its DNA genome (Fig. 1). The resulting DNA pool was then transcribed into RNA and used as the starting point for in vitro selection.

DNA pool characterization

To analyze the final recombined DNA pool, we sequenced 31 isolates at random from the pool. Twenty isolates were built from two fragments, seven had three fragments, and four had four or more fragments. The average length of the cloned molecules was 111 base pairs (bp) with a broad distribution ranging from ~50 to ~200 bp. The length distribution of the fragments used to construct these molecules was an asymmetrical bell shape skewed toward shorter fragment lengths. Of these fragments, 90% were between 11 and 74 bp in length with an average length of 34 bp (we excluded from this analysis fragments containing the 5'-primer sequence which was 37 bp long). Three translocations and two inversions were observed (note inversions are not possible in two-fragment molecules, and translocations require at least four-fragment molecules). In all, 47 fragment ligation sites were identified. Slightly over half (26/47) clearly resulted from blunt-ended ligation, while the remainder were ambiguous and could not be distinguished from the ligation of fragments possessing sticky-end overhangs. We found 10 ligation sites with 1 bp of ambiguity, five with 2 bp, and six with 3 bp (expected from chance if blunt-ended: 9, 2, and 0.5, respectively), suggesting that the formation of blunt ends by T4 DNA polymerase was partially incomplete.

The distribution of DNase I cut sites required to generate these molecules was also analyzed together with nine additional sequences from recombined DNA that was not gel-purified. A total of 126 cut site locations were identified (Fig. 1F) (if the cut site was ambiguous, the longest possible fragment was arbitrarily selected). DNase I cut nearly randomly. The increase in frequency near either end of the B.01 sequence presumably reflects the fact that many of the sequences consist of only two fragments and that all sequences must contain a 5'- and a 3'-primer sequence fragment. The total number of sequences in the starting pool can be estimated from the fact that 14 cycles of PCR were used to generate ~5 pmol of final pool DNA. Assuming completely optimal PCR conditions there cannot have been $< \sim 10^8$ ($6 \times 10^{23} \times 5 \times 10^{-12}/2^{14}$) DNA sequences in the initial pool.

In vitro selection strategy

From the above pool, ribozymes able to synthesize a pyrimidine nucleotide were isolated using a selection strategy modified from a protocol originally developed to isolate their B.01 progenitor (Unrau and Bartel 1998; Chapple et al. 2003). The selection, summarized in Figure 2, relies on the *N*-acryloylaminophenylmercuric acetate (APM) gel retardation of ribozymes possessing a covalently attached 4-thiouridine ([4S]U) product. These ribozymes must also satisfy defined length constraints. This was carried out by using RNA markers that spanned a length range from 70 to 274 nt and were derivatized at their 3'-ends with [4S]U. By Round 3, ribozyme activity was observed, and the selection was split as follows: The length distribution of the size-independent pool was left alone and was subjected only to decreasing incubation times. The second pool was subjected to size-dependent selection by excising progressively shorter and shorter
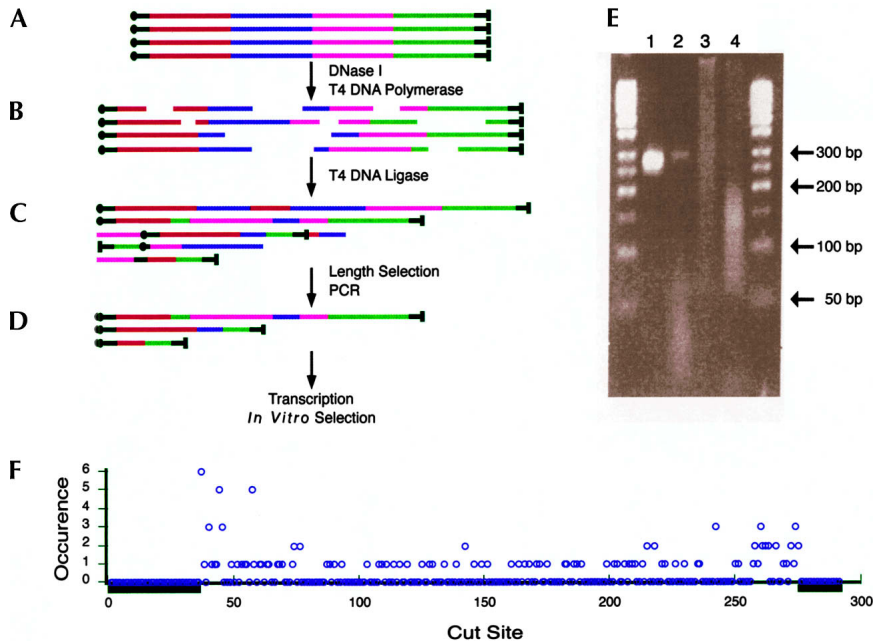
FIGURE 1. The random recombination scheme used to synthesize the starting pool. (*A*) Colored lines indicate arbitrary positions within the defined input B.01 DNA sequence (four copies shown). Black lines terminated with a dot and bar denote the 5′- and 3′-PCR primer-binding regions, respectively. The primers used to generate the pool are not phosphorylated. (*B*) dsDNAs are cut into randomly distributed fragments using DNase I. The resulting overlapping ends are removed or filled into blunt ends using T4 DNA polymerase. (*C*) T4 DNA ligase randomly reassembles the blunt-end fragments into new molecules with a broad length distribution. These fragments may now have a different location or orientation within the ligated construct. Nicked dsDNA with inner fragments containing 5′ hydroxyl primer sequence and dsDNA with inner fragments containing the majority of the primer region result in shorter DNA after PCR amplification. (*D*) After agarose gel purification of a suitable size range and PCR amplification, the final recombined DNA pool consist of molecules with 5′- and 3′-primer sequence at each end. This pool is then transcribed into the initial RNA pool and is ready for in vitro selection. (*E*) The recombined DNA pool construction steps summarized on a 3% agarose gel. (Lane *1*) Input 291-bp B.01 DNA; (lane *2*) DNA fragments after DNase I digestion; (lane *3*) DNA pool after T4 DNA ligase step; (lane *4*) the final recombined DNA pool after length selection and PCR. (*F*) Distribution of DNase I cut sites within the B.01 DNA sequence. A total of 126 cut sites (of 237 possible) were inferred from 40 recombined sequence isolates (see text). Long and short black bars indicate the 37-bp 5′-PCR primer and 17-bp 3′-RT-PCR primer locations, respectively.

products from post-transcription gels, and from the APM selection gels used to purify active sequences after ribozyme incubation with 4-thiouracil (as described in the Materials and Methods section).

Result of selections

The length distributions of the size-independent RNA converged each successive round to an interesting pattern (Fig. 3). During the first two selection rounds, the peak of the length distribution gradually increased from ~130 nt to ~150 nt. By Round 3, the active sequences were distributed over a region spanning from ~140 to ~220 nt. After this round, two peaks fixed at ~210 nt and ~145 nt long, separated by a moderately populated intermediate region, were predominant. The activity of Rounds 4–6 pool RNA was

less than twofold higher than the catalytic rate observed for B.01, whose apparent k_{cat}/K_m was 1.3 M^{-1} min^{-1} (Unrau and Bartel 1998).

During the size-dependent selection from Rounds 4 to 10, the input material for each round consisted of RNA between 70 nt and the shortest observable active sequence from the previous round (Fig. 4). As a consequence, the longest dominant transcripts decreased from ~145 nt to ~80 nt. The catalytic rates of Rounds 4 to 7 were similar to that of B.01 but by Round 8, first-order reaction rates had dropped to about half that of B.01 and were only recovered by decreasing the incubation times used for Rounds 11 and 12.

Active sequence analysis

In total, 64 isolates from Rounds 4 to 6 of the size-independent selection were sequenced. Sixty isolates aligned very well with B.01 and contained a diverse range of deletions; 15 of these picked at random were tested, and all were found to be active. The four sequences that did not align well to B.01 were found to be inactive. In addition, 18 isolates from Rounds 6 to 12 of the size-dependent selection were sequenced and screened for activity. Here only nine of the highly truncated isolates were found to be active, and all aligned well with B.01. All of the active ribozyme sequences are shown aligned to B.01 in Figure 5.

The inactive sequences from both selections are likely to be parasites that appear to survive by at least two mechanisms. Three of the four inactive sequences from the size-independent selection and eight out of nine of the inactive size-dependent isolates were observed to double in size after the T4 RNA ligation step, potentially circumventing the size-dependent selection step. The two remaining inactive isolates contained sequence (at least 19 nt long) that could hybridize with the reverse complement of nucleotides 210–229 of B.01, potentially leading to their survival as a result of incomplete dehybridization during the gel selection steps.

All 24 of the active ribozyme sequences have deletion regions that define four highly conserved "island" regions where no deletions were observed to occur. The 45 untested isolates from the size-independent selection never violated the island boundaries defined by the functional isolates tested. Any sequenced isolates that contained deletions ex-
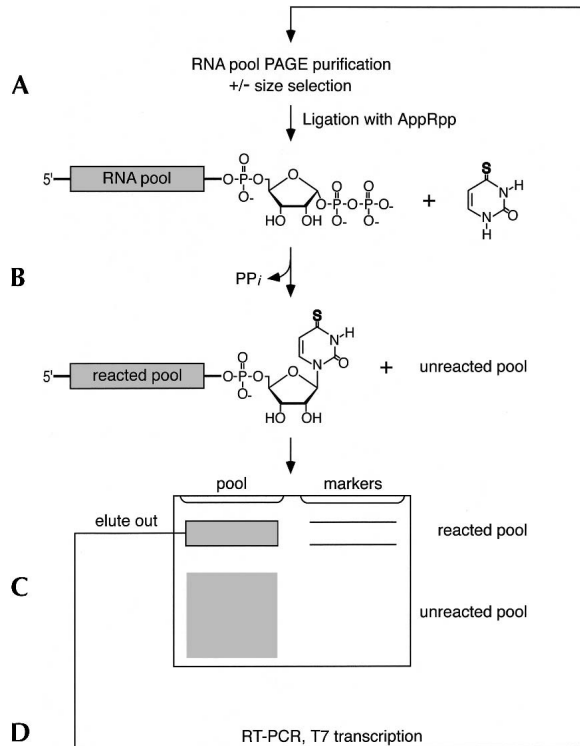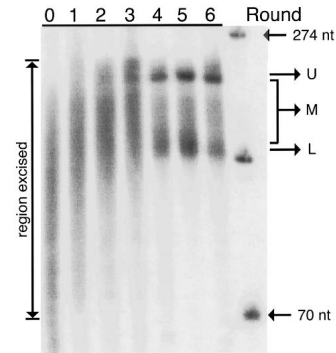
FIGURE 2. In vitro selection strategy. (*A*) The input RNA pool is transcribed and a suitable size range is selected using a denaturing polyacrylamide gel. The purified pool is ligated to pRpp using adenylylated pRpp (AppRpp) and T4 RNA ligase. (*B*) The resulting pool now tethered to pRpp at its 3′-terminus is incubated with 4-thiouracil. The reacted pool consists of ribozymes able to catalyze the synthesis of tethered ^{4S}U. (*C*) Active ribozymes in a size range indicated by synthetic markers terminated with ^{4S}U are purified using one or two APM gels, which retard the mobility of sulfur-containing RNA. (*D*) The eluted RNA pool is reverse-transcribed into DNA, ready for the next round of selection. Two selection pressures can be applied. The first is by varying the RNA size recovered from polyacrylamide and APM gels, and the second is by changing the 4-thiouracil incubation time.

FIGURE 3. RNA pool distribution resulting from the size-independent selection. Equivalent amounts of input RNA from each round of selection were loaded on a 6% polyacrylamide gel, with RNA size markers (70, 137, and 274 nt) loaded on the *right*. Round 0 represents the initial recombined RNA pool. In each round, a segment of RNA from 70 nt to ~250 nt was excised from gels as indicated by the bars on the *left*. The 4-thiouracil incubation times from Round 1 to 4 were kept at 24 h and then decreased to 0.5 h for Round 5, and 0.05 h by Round 6. By Round 4 the RNA pool distribution showed two dominant bands, labeled upper (U) and lower (L), which were spanned by a middle region (M).

tending into these islands or having sequence differing from these island motifs were found to be inactive.

Two of these islands involve the primer-binding sequences required for reverse transcription and amplification and are both exactly the length of the primer sequence. The two inner islands are 20–21 nt and 24–25 nt in length. For the size-independent selection, the upper bands (U, Fig. 3), which corresponded to the first five active sequences by length, were built from two fragments, with one exception. The middle region (M, Fig. 3) consisted of three- and four-fragment sequences, while the lower band (L, Fig. 3) consisted of three-fragment molecules. The remaining shorter sequences, which resulted from the size-dependent selection, contained four to five fragments.

DISCUSSION

Recombined RNAs can be sorted into classes defined by the number of parental fragments they contain. As the number

of fragments grows, the average RNA length increases in a linear fashion, while the total number of recombined sequences increases exponentially. Prior to selection, the distribution for each fragment class should theoretically be centered on the mean length for each class (i.e., the fragment number times the average fragment length), but this was hard to reconstruct directly from our experimental data. The total number of possible two-fragment recombinants in our pool is ~10^4, as calculated from $[l_m(l_m - 1) - l_s(l_s - 1)]/2$, with l_s and l_m being the respective minimum and maximum pool lengths, which in our selection ranged from 50 to 250 nt. By a similar calculation there
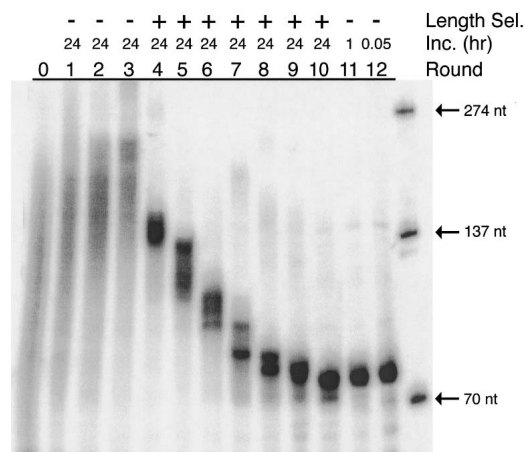
FIGURE 4. RNA pool distribution resulting from the size-dependent selection. Starting in Round 4, a gel slice extending from 70 nt to the shortest RNA dominating the previous round was used to select the RNA for the next round of selection. The 4-thiouracil incubation times used for each round of selection are indicated.
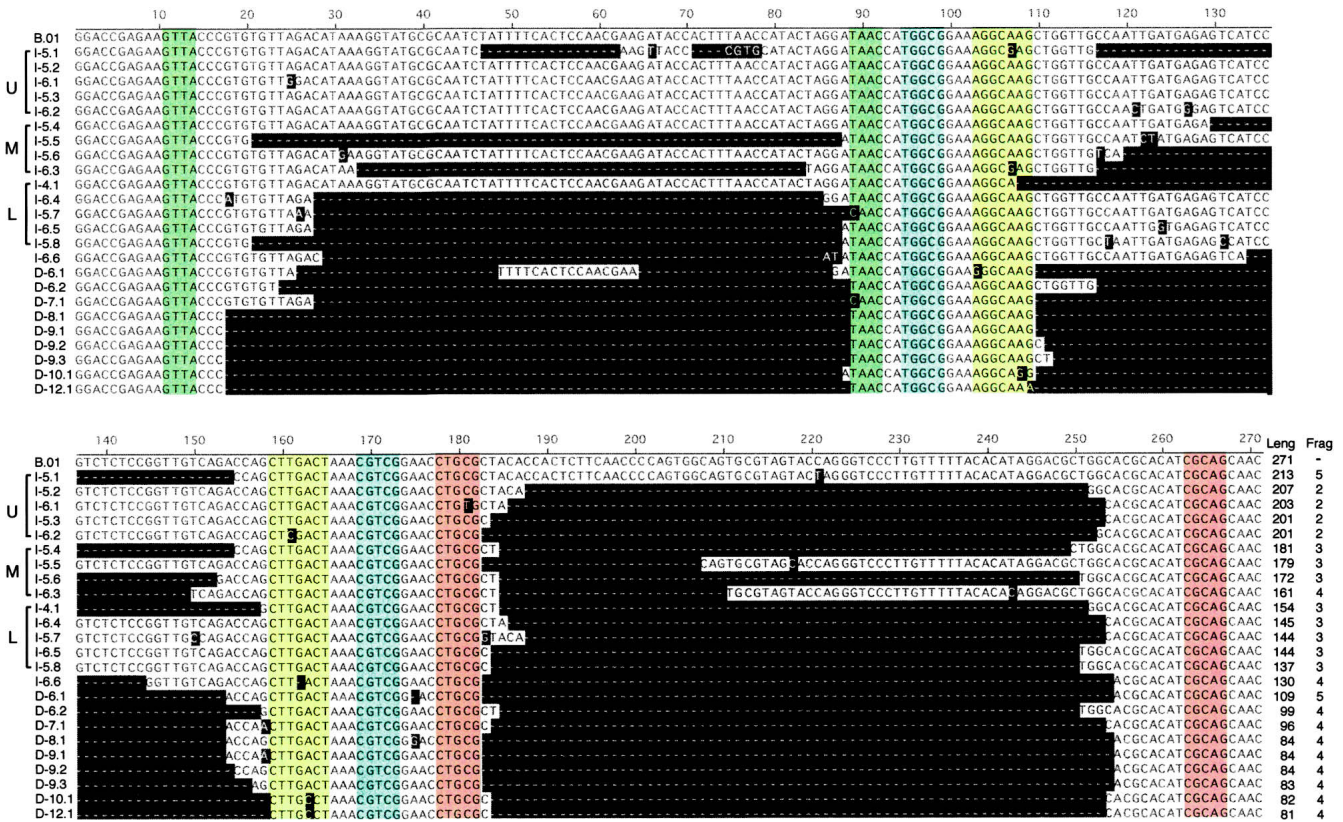
FIGURE 5. Sequence alignment of active ribozyme DNA sequence isolated from both selections and sorted by length. Ribozymes from the size-independent selection are indicated by I-*X.Y*, where *X* is the round of selection and *Y* is the isolate number. Sequences marked with a D correspond to isolates from the size-dependent selection. U, M, or L correspond to the range of lengths expected from the *upper*, *middle*, and *lower* bands observed in Figure 3, respectively. Pairs of vertically colored bars, within the four island regions, represent helical structural elements as predicted by the *Pknots* algorithm. The ribozyme length (Leng) and number of recombined fragments (Frag) for each isolate are shown in the *lower right*.

are ~10^6 three-fragment, 10^8 four-fragment, and 10^{10} five-fragment pool molecules possible in the same size range. As ~10^8 different recombinants were created, it is likely that nearly all two- and three-fragment molecules and a significant fraction of all four-fragment recombinants were contained within the recombined pool.

The conservation of two central islands observed in the size-dependent selection allows the gel band patterns observed in the size-independent selection to be analyzed and understood in terms of each recombinant RNA class. A two-fragment ribozyme can only tolerate deletions between islands that are separated by regions 71, 49, and 73 nt long (Fig. 5). As a consequence, the shortest functional two-fragment ribozyme must be at least 198 nt (271–73 nt) long. Theoretically, the number of two-fragment constructs decreases very steeply in this length region, and the transition from active to inactive sequences can therefore to be expected to be sharply defined at ~200 nt as was observed (upper band in Fig. 3, sequences labeled "U" in Fig. 5).

The same logic applied to three-fragment recombinants implies that this class of ribozyme must be at least 127 nt (271–73–71 nt) long. In contrast to the two-fragment ribo-

zymes, viable constructs should span a much broader size distribution as the fragment sizes required to produce a three-fragment ribozyme are compatible with the average fragment size used to generate the pool (average fragment length is ~35 nt), again consistent with observation (lower and middle bands in Fig. 3, sequences labeled "L" and "M" in Fig. 5). Finally, four-fragment ribozymes should most likely occupy the region between that of the active two- and three-fragment ribozymes with an average length of ~140 nt in the absence of size selection pressure. The functional isolates from the size-independent selection all conform to these length and size constraints very well (Fig. 5).

Ribozymes shorter than ~127 nt appear to be impossible to construct unless they are built from at least four fragments as a smaller number of fragments would lead to the partial deletion of at least one conserved sequence island. Since short four-fragment constructs must result from the ligation of smaller and rarer fragments, much fewer of these constructs are expected. Consistent with this expectation, ribozymes shorter than ~130 nt were observed only faintly by Round 6 of the size-independent selection (see Fig. 3), and the emergence of shorter ribozymes before their longer

and presumably more numerous relatives during the size-dependent selection was never observed. Despite the decreasing diversity suggested by the discrete bands formed during this selection (see Fig. 4), variability in this size range was still appreciable, and none of the nine short functional ribozymes were cloned more than once.

The shortest sequences were folded using the algorithm *Pknots* (Rivas and Eddy 1999), which is capable of determining higher-order structures that contain pseudoknots. The resulting structures were identical, and the predicted fold for the shortest isolate is shown in Figure 6. The ribozyme consists of three putative helix–loop structures, one of which contains an internal loop. The three blocks of nonessential sequence removed by recombination are located in the loop regions of helix–loops I, II, and III. This structure is consistent with every functional isolate sequenced. Although point mutagenesis was not deliberately introduced during the selection in order to study the effect of recombination, low levels of point mutation were still observed. Helix II contains an A:G mismatch in many isolates, but notably in the shortest sequences this is corrected to a C(43):G(34) pair (Fig. 6), presumably to stabilize the shortened helix–loop found in these isolates.

Given the length distribution of RNAs retained in the size-independent selection, it was somewhat surprising that functional isolates containing sequences in orders and orientations other than in the parental isolate were not found. We had originally hypothesized that while inversions might not be observed because of the formation of stable and inhibitory secondary structures, translocations might be possible. Further sequencing may reveal such sequences, but we found only one functional isolate that contained a 4-nt stretch (CGUG) that cannot be aligned perfectly (modulo deletion regions) with B.01 (Fig. 5). Formally, the secondary structure we propose can be threaded into two distinct topologies. The unobserved topology would replace the three hairpin loops with three interhelical connecting strands and

is obtained by swapping the two central island sequences. The context of these connecting strands appears critical to correctly orient helices I, II, and III and might preclude the easy formation of such a ribozyme. The loop-containing topology, however, contains ribozymes with remarkably homogeneous kinetic rates (longer isolates were no more than twofold faster and the shorter isolates no more than twofold slower than the B.01 parent). This implies that the observed topology efficiently prevents extraneous sequence from interfering with catalysis.

In conclusion, random recombination together with in vitro selection offers an efficient way to sample a large number of ribozyme deletion constructs simultaneously. A diverse range of active deletion constructs were isolated, and all were compatible with a single minimal motif (Fig. 6). Significantly, and in contrast to manual truncation (Ekland et al. 1995; Uhlenbeck 2003) and other deletion strategies (Chapple et al. 2003), 190 nt of sequence was removed from a 271-nt RNA without any prior knowledge of phylogeny or secondary structure. While ribozyme activity was not enhanced by this selection, the deliberate introduction of point mutations during the construction of a randomly recombined pool is likely to result in the simultaneous isolation of improved and precisely truncated ribozymes. The phylogenetic information provided by these point mutations would allow the rapid and accurate determination of RNA core motifs.

This technique could easily be applied to systematically study naturally occurring RNAs such as tmRNA (Williams and Bartel 1996; Williams et al. 1999) and riboswitches (Mandal and Breaker 2004; Winkler et al. 2004) that play important cellular roles. It might also be used to recombine pre-existing elements to create new RNA structures having different or enhanced functions and, as a consequence, increase our understanding of the evolutionary potential of RNA-based systems.

MATERIALS AND METHODS

General protocol

dsDNA concentration was estimated on an agarose gel by comparison with a previously calibrated 50-bp ladder (MBI Fermentas). PCR was accomplished on a PTC-100 Peltier Thermal Cycler (MJ research) using 10 mM Tris-HCl, 50 mM KCl, 15 mM MgCl$_2$, 0.01% gelatin (pH 8.3), 1 μM each primer, 0.1 U/μL *Taq* polymerase; the cycles were 94°C for 45 sec, 50°C for 85 sec, and 72°C for 115 sec. The PCR primers were 5′-*TTCTAATACGACTCAC TATA*GGACCGAGAAGTTACCC (T7 promoter in italics) and 5′-GTTGCTGCGATGTGCGT. The radiolabeled RNA pool (using [α-^{32}P]UTP during transcription) and synthetic markers (labeled using polynucleotide kinase and [γ-^{32}P]ATP) were detected using a phosphorimager (Storm 820, Molecular Dynamics). The length distribution of RNA pools was inferred using three known size markers (70-, 137-, and 274-nt RNA transcripts) as standards.
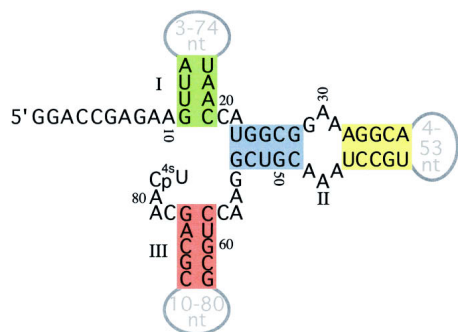
FIGURE 6. Proposed core motif secondary structure of the family B nucleotide synthase ribozyme family as predicted using the *Pknots* algorithm. The three regions of deletion are found within the three terminal loops of the structure, and their size ranges are indicated. The sequence corresponds to the shortest isolate (D-12.1) in the alignment shown in Figure 5 and is numbered from its 5′-end.

DNase I digestion

B.01 ribozyme DNA, isolated from a previous in vitro selection (Unrau and Bartel 1998), was PCR-amplified. Excess primers were removed using a Qiaquick PCR Purification Kit (QIAGEN) before DNA precipitation. DNA was resuspended in digestion buffer (40 mM Tris-acetate, 6 mM $MnCl_2$ at pH 7.5) at ~375 nM to which 5×10^{-4} U/µL DNase I (Amersham) was added. This enzyme, which had previously been diluted in storage buffer lacking Mg^{2+} (10 mM Tris-HCl, 10 mM $CaCl_2$, 50% glycerol at pH 7.5), was prepared from a 10 U/µL commercial stock. Digestion was performed in the presence of Mn^{2+}, as DNase I is claimed to produce blunt-ended fragments at random locations in the presence of this metal ion (Melgar and Goldthwait 1968; Campbell and Jackson 1980). After 1 h of incubation at 23°C, 6 mM EGTA was added (to specifically chelate Mn^{2+} in the presence of Mg^{2+} in subsequent steps) to the reaction, which was phenol-extracted twice and then ether-extracted. The digested DNA was stored at −20°C.

Generation of blunt-end fragments with T4 DNA polymerase

The digested material was diluted 1.5 times and supplemented to create a suitable environment for T4 DNA polymerase. The final buffer composition was 40 mM Tris- (HCl and acetate forms), 4 mM $MnCl_2$, 4 mM EGTA, 66 mM sodium acetate, 13.3 mM Mg acetate, 0.5 mM DTT, 0.2 mM dNTPs, 0.01% BSA (pH 7.9), and 0.1 U/µL T4 DNA polymerase (NEB). After incubating for 15 min at 11°C, the reaction was stopped by phenol/ether extraction and stored at −20°C.

Reassembling of fragments with T4 DNA ligase

The polished material was again diluted 1.2 times and supplemented to provide an optimal environment for T4 DNA ligase. The final buffer composition was now 50 mM Tris- (HCl and acetate forms), 3.3 mM $MnCl_2$, 3.3 mM $MgCl_2$, 11 mM Mg acetate, 3.3 mM EGTA, 55 mM sodium acetate, ~0.2 mM dNTPs, 0.01% BSA, 5.9% polyethylene glycol-8000, 1.2 mM DTT, 1.2 mM ATP (pH 7.6), and 0.1 U/µL T4 DNA ligase (Invitrogen). After 16 h of incubation at 23°C, the reaction was halted by addition of EDTA.

Agarose gel purification and PCR amplification

About 2 pmol of the ligated material was run on a 3% agarose gel containing ethidium bromide. Using a long-wavelength lamp (~365 nm), a slice between ~130 and ~250 bp, with ~60 µL volume, was cut out and heated at 90°C in 1 mL of water until melted, followed by three equal volume phenol extractions, one phenol-chloroform, and a final chloroform extraction. The extracted material (~1 pmol) was concentrated using a Microcon YM-3 centrifugal filter device (3000 MWCO; Millipore). After 14 cycles of PCR, that resulted in ~5 pmol of DNA, this material was diluted 100 times and amplified a further eight cycles. After phenol/chloroform and ethanol precipitation, this DNA pool was resuspended in water with a final concentration of ~500 nM and used to transcribe the initial RNA pool.

In vitro selection size-independent selection

During each round, the ribozyme pool was split in half. Radiolabeled pool RNA was used to facilitate the detection and purification of active ribozymes, while unlabeled pool RNA was mixed with radiolabeled synthetic markers to locate the regions to be excised.

RNA transcription and ligation to pRpp

In each round, ~10 pmol DNA template at ~100 nM was transcribed (40 mM Tris-HCl, 2.5 mM spermidine, 26 mM $MgCl_2$, 0.1% Triton X-100, 8 mM GTP, 5 mM ATP, 5 mM CTP, 2 mM UTP, +/− [α-^{32}P]UTP at pH 7.9, 5 U/µL T7 RNA polymerase) for 80 min at 37°C. The reaction was stopped with one volume of loading dye (90% formamide, 50 mM EDTA) and if unlabeled mixed with radiolabeled synthetic 70-nt and 274-nt RNA markers. After heating to 90°C for 5 min, the RNA was purified on a 6% polyacrylamide gel. The regions between 70 and ~250 nt as indicated by the synthetic markers were excised and eluted at 4°C into 300 mM NaCl overnight. The resulting material was ethanol-precipitated and resuspended in water. Pool RNA was derivatized with pRpp at its 3′-end using published protocols (Unrau and Bartel 1998; Chapple et al. 2003; Lau et al. 2004).

Ribozyme reaction and APM gel purification

To decrease cross-hybridization between different species potentially containing the reverse complement of other pool members, pool RNA was heated to 80°C for 5 min and cooled to 4°C in water before being incubated at 0.1 µM in 4 mM 4-thiouracil, 50 mM Tris-HCl, 150 mM KCl, 25 mM $MgCl_2$, and 0.5 mM $MnCl_2$ (pH 7.5) at 23°C. The RNA was then heated to 90°C for 5 min after addition of an equal volume of loading dye and loaded onto an APM gel (6 M urea/6% polyacrylamide gels, containing 120 µM APM) (Igloi 1988). This gel was also loaded with 70-nt and 274-nt marker RNAs derivatized at their 3′-ends with ^{4S}U using adenylylated ^{4S}U as previously described (Unrau and Bartel 1998; Chapple et al. 2003). These markers, which are shifted upward in an APM gel, served to delimit the location of active ribozyme sequences that are able to synthesize a terminal ^{4S}U residue. Pool RNA within this region was excised and eluted in the presence of 1 mM DTT before being ethanol-precipitated. A second APM gel was applied for Rounds 4–6 of size-independent selection and Rounds 10–12 of the size-dependent selection. Incubation times were 24 h for Rounds 1–4 of the size-independent selection, followed by 30 min for Round 5 and 3 min for Round 6.

Size-dependent selection

The size-dependent selection was identical to the size-independent selection except progressively shorter RNA molecules were used as input for Rounds 4–10 of this parallel selection. After transcription, a gel slice from the 70-nt-long marker up to the first observable RNA band was excised. After incubation with 4-thiouracil, the pool was also gel-purified using an APM gel, and a shifted region between the shifted 70-nt 4-thiouridine marker RNA and the shortest shifted RNA band observed on the gel was recovered.

Incubation times were decreased from 24 h to 60 min and 3 min for Rounds 11 and 12, respectively.

Reverse transcription PCR

RNA from a round of selection was reverse-transcribed into cDNA using Reverse Transcriptase (Invitrogen) and PCR-amplified as previously described (Zaher and Unrau 2005). After phenol/chloroform extraction and ethanol precipitation as before, the DNA pool was transcribed into RNA and subjected to the next round of in vitro selection.

Sequencing and alignment

Sequences were cloned using a TA clone kit (Invitrogen) and sequenced using a DYEnamic ET terminator cycle sequencing kit (Amersham). Sequences were aligned using Megalign (DNA Star). Secondary structure was modeled using the *Pknots* version 1.01 algorithm (Rivas and Eddy 1999).

ACKNOWLEDGMENTS

REFERENCES

Bittker, J.A., Le, B.V., and Liu, D.R. 2002. Nucleic acid evolution and minimization by nonhomologous random recombination. *Nat. Biotechnol.* **20:** 1024–1029.

Bittker, J.A., Le, B.V., Liu, J.M., and Liu, D.R. 2004. Directed evolution of protein enzymes using nonhomologous random recombination. *Proc. Natl. Acad. Sci.* **101:** 7011–7016.

Campbell, V.W. and Jackson, D.A. 1980. The effect of divalent cations on the mode of action of DNase I. The initial reaction products produced from covalently closed circular DNA. *J. Biol. Chem.* **255:** 3726–3735.

Chapple, K.E., Bartel, D.P., and Unrau, P.J. 2003. Combinatorial minimization and secondary structure determination of a nucleotide synthase ribozyme. *RNA* **9:** 1208–1220.

Ekland, E.H., Szostak, J.W., and Bartel, D.P. 1995. Structurally complex and highly-active RNA ligases derived from random RNA sequences. *Science* **269:** 364–370.

Igloi, G.L. 1988. Interaction of transfer-RNAs and of phosphorothioate- substituted nucleic-acids with an organomercurial—Probing the chemical environment of thiolated residues by affinity electrophoresis. *Biochemistry* **27:** 3842–3849.

Illangasekare, M., Kovalchuke, O., and Yarus, M. 1997. Essential structures of a self-aminoacylating RNA. *J. Mol. Biol.* **274:** 519–529.

Khvorova, A., Lescoute, A., Westhof, E., and Jayasena, S.D. 2003. Sequence elements outside the hammerhead ribozyme catalytic core enable intracellular activity. *Nat. Struct. Biol.* **10:** 708–712.

Knight, R. and Yarus, M. 2003. Finding specific RNA motifs: Function in a zeptomole world? *RNA* **9:** 218–230.

Krasilnikov, A.S., Xiao, Y., Pan, T., and Mondragon, A. 2004. Basis for structural diversity in homologous RNAs. *Science* **306:** 104–107.

Lau, W.L., Cadieux, K.E.C., and Unrau, P.J. 2004. Isolation of fast purine nucleotide synthase ribozymes. *J. Am. Chem. Soc.* **126:** 15686–15693.

Mandal, M. and Breaker, R. 2004. An mRNA structure that differentially controls gene expression by binding guanine or adenine. *Biophys. J.* **86:** 32A.

Melgar, E. and Goldthwait, D.A. 1968. Deoxyribonucleic acid nucleases. II. The effects of metals on the mechanism of action of deoxyribonuclease I. *J. Biol. Chem.* **243:** 4409–4416.

Rivas, E. and Eddy, S.R. 1999. A dynamic programming algorithm for RNA structure prediction including pseudoknots. *J. Mol. Biol.* **285:** 2053–2068.

Sabeti, P.C., Unrau, P.J., and Bartel, D.P. 1997. Accessing rare activities from random RNA sequences: The importance of the length of molecules in the starting pool. *Chem. Biol.* **4:** 767–774.

Szewczak, A.A. and Moore, P.B. 1995. The sarcin/ricin loop, a modular RNA. *J. Mol. Biol.* **247:** 81–98.

Uhlenbeck, O.C. 2003. Less isn't always more. *RNA* **9:** 1415–1417.

Unrau, P.J. and Bartel, D.P. 1998. RNA-catalysed nucleotide synthesis. *Nature* **395:** 260–263.

———. 2003. An oxocarbenium-ion intermediate of a ribozyme reaction indicated by kinetic isotope effects. *Proc. Natl. Acad. Sci.* **100:** 15393–15397.

Westhof, E. and Massire, C. 2004. Evolution of RNA architecture. *Science* **306:** 62–63.

Williams, K.P. and Bartel, D.P. 1996. Phylogenetic analysis of tmRNA secondary structure. *RNA* **2:** 1306–1310.

Williams, K.P., Martindale, K.A., and Bartel, D.P. 1999. Resuming translation on tmRNA: A unique mode of determining a reading frame. *EMBO J.* **18:** 5423–5433.

Winkler, W.C., Nahvi, A., Roth, A., Collins, J.A., and Breaker, R.R. 2004. Control of gene expression by a natural metabolite-responsive ribozyme. *Nature* **428:** 281–286.

Yusupov, M.M., Yusupova, G.Z., Baucom, A., Lieberman, K., Earnest, T.N., Cate, J.H.D., and Noller, H.F. 2001. Crystal structure of the ribosome at 5.5 Å resolution. *Science* **292:** 883–896.

Zaher, H.S. and Unrau, P.J. 2005. Nucleic acid library construction using synthetic DNA constructs. In *Oligonucleotide synthesis methods and applications* (ed. P. Herdewijn), pp. 359–378. Humana Press, Totowa, NJ.