1-1-2005

Optimal Random Libraries For the Isolation of Catalytic RNA

Tricia M. Coleman
University of Southern Mississippi

Faqing Huang
University of Southern Mississippi, Faqing.Huang@usm.edu

[RNA Biology 2:4, 129-136, October/November/December 2005]; ©2005 Landes Bioscience

Research Paper

Optimal Random Libraries for the Isolation of Catalytic RNA

Tricia M. Coleman

Faqing Huang*

Department of Chemistry and Biochemistry; University of Southern Mississippi; Hattiesburg, Mississippi

*Correspondence to: Faqing Huang; Department of Chemistry and Biochemistry; University of Southern Mississippi; Hattiesburg, Mississippi 39406-5043 USA; Tel.: 601.266.4371; Fax: 601.266.6075; Email: faqing.h.huang@usm.edu

Received 09/12/05; Accepted 10/31/05

Previously published as a *RNA Biology* E-publication:
http://www.landesbioscience.com/journals/rnabiology/abstract.php?id=2285

KEY WORDS

RNA catalysis, thioester synthesis, ribozyme size-activity relationship, optimal RNA library, RNA molecular evolution

ABBREVIATIONS

nt	the number of nucleotides
N	the number of random nucleotides
CoA	Coenzyme A

ACKNOWLEDGEMENTS

This work was supported by an NSF grant MCB9974487 and a NASA grant NAG5-10668.

ABSTRACT

The relationship between ribozyme size and catalytic activity is of fundamental importance for RNA catalysis and molecular evolution in the RNA world. We have performed a series of competitive in vitro selection experiments to probe the relationship using RNA libraries containing size-heterogeneous random regions. Our experiments have established an inverse correlation between RNA replication efficiency (the combined efficiency of PCR amplification, transcription, and reverse transcription) and RNA size. A number of ribozyme sequences have been isolated from different RNA size groups under competitive selection conditions. Comprehensive kinetic analysis on isolated ribozymes has revealed that large ribozymes do not confer a significant catalytic superiority over smaller ones under most selection conditions, and actually impose two significant problems of replication inefficiency and RNA misfolding into inactive conformations. The fraction of a misfolded ribozyme population is defined as alpha. Large ribozymes tend to possess high alpha values, which may significantly reduce ribozyme performance. Our results suggest that a random region of around 60 nucleotides represents the optimal balance between ribozyme catalytic activity, RNA misfolding (alpha), and replication efficiency, and may therefore constitute the most advantageous RNA libraries for successful isolation of functional RNA sequences.

INTRODUCTION

In vitro selection techniques[1-3] have been used to isolate an array of catalytic RNA molecules and RNA aptamers. A critical aspect to the selection procedure is the preparation of initial libraries of randomized RNA molecules. One of the primary considerations when designing these RNA libraries is the length of the random region. Currently, diverse lengths of random regions, ranging from 30 nucleotides (nt) or less[4,5] to over 200 nucleotides,[6,7] are being used by different laboratories. The use of considerably different random region sizes reflects differing opinions on the advantages and disadvantages of short and long random regions. Currently, there is no agreement as to what lengths of random regions constitute optimal RNA libraries for successful isolation of functional RNA molecules.

There are distinct, yet disputed, advantages and disadvantages to both short and long random regions. Short random regions allow sampling more, if not all, of the possible sequence combinations, while only a minute fraction of the sequence space can be covered by a long random nucleotide region. The advantages of long random regions may include the combinatorial effects of sequence interactions and increased probability of finding large motifs required for complex activities.[8,9] In contrast, when RNA sequences get too short, the lack of folding into essential functional structure could impair catalytic activity. From existing literature, however, it is not clear whether large RNA sizes are necessary or advantageous for isolating RNA sequences with defined catalytic activities.

The question of the length of random RNA region vs. ribozyme catalytic activity is not only important for designing optimal random RNA libraries, but also bears significant implications for molecular evolution in the RNA world. On one hand, a number of relatively large ribozymes (>100 nt) have been isolated from random RNA libraries, e.g., the 274-nt RNA ligase ribozymes,[6] 173-nt polynucleotide kinase ribozymes,[10] ~140-nt ribozymes for C-C bond formation,[11] 274-nt ribozymes for nucleotide synthesis,[7] and 160-nt Diels-Alderase ribozymes before truncation.[12] On the other hand, no realistic mechanisms for replication of long RNA sequences by RNA catalysts have been established. Although impressive ribozyme polymerases have been isolated by in vitro selection

techniques that can extend template-directed RNA polymerization,[13] they are only capable of synthesizing small pieces of RNA (up to 14 nucleotides).

Experimental evidence to address the relationship between ribozyme size and catalytic activity is strikingly lacking. To tackle the problem, we have been performing in vitro selection experiments to isolate novel ribozymes from competitive RNA libraries containing size-heterogeneous random regions.[14,15] These libraries closely mimic true evolutionary conditions where the sequences in each size group compete for representation both within the individual groups and between different groups of varying random regions. In essence, a scenario develops where ribozymes with the best fit between catalytic activity and replication efficiency dominate during the selection. A relationship between random RNA size and catalytic activity has been established through the analysis of both RNA size population changes during selection cycles and catalytic activity of isolated ribozymes. From the current study, as well as our previous results,[14,15] we find that initial RNA libraries consisting of ~60 random nucleotides represent an optimal balance between ribozyme efficiency and replication efficiency.

MATERIALS AND METHODS

Optimization of selection conditions. Prior to the selection for thioester synthetase activity, transcription, reverse transcription, and PCR conditions were optimized to reduce potential bias toward any particular size group. Seven simulated selection cycles (background selection) were performed in the absence of a reaction step (i.e., no selection pressure). The starting size-heterogeneous random library was the same as described previously,[14] containing four differently-sized random pools with random regions of 30N, 60N, 100N and 140N. All the four pools contained identical PCR primer sequences: P1 (GCGAATTCAGTAATACGACTCACTATTAGGG-AAGTGCTACC) and P2 (TGGCTGGCCGCATGCCCG) (underlined sequence = T7 ϕ2.5 promoter, bold face A = transcription initiation site). Starting from a library containing equal molar ratio of dsDNA of the four sizes, RNA was transcribed under the standard conditions.[16] No apparent bias in transcription toward RNA sizes was observed. Following purification of RNA by membrane filtration (Millipore, M30), 2 pmol (0.2%) of the total RNA was used for reverse transcription. Different conditions, including using both AMV and M-MLV reverse transcriptases and different concentrations (10 to 30 units/10 μL reaction), and different reaction times (20 to 60 min at 42°C), were explored to minimize bias against large RNA. No significant difference was observed among different reverse transcription conditions. We thus opted to use the standard reverse transcription conditions.[16] Finally, PCR was performed under two different conditions: 5 units *Taq* DNA polymerase/100 μL reaction and 1 min extension at 72°C (standard conditions, background selection A), and 10 units *Taq* DNA polymerase/100 μL reaction with 4 min extension at 72°C (improved conditions, background selection B). When amplifying size-heterogeneous DNA, differences between the two PCR conditions became apparent. The improved conditions significantly reduce the bias towards long sequences.

Control selection. Size-heterogeneous RNA libraries were used to conduct a third selection (background selection C) to establish background levels. To determine whether the specific conditions used during the original thioester selection interfered with the cycling efficiency of the longer sequences, a set of seven control selection rounds were performed using the same initial size-heterogeneous RNA library.[14,15] Transcription, reverse transcription, and PCR steps were performed under the same conditions as in background selection B and the previous thioester-synthesizing ribozyme selection.[15] In this control selection, however, imidazole was used as a catalyst in the reaction step to convert nearly all CoA-RNA to Biotin-CoA-RNA after reaction with BiotinAMP.[17] CoA-RNA purified by thiopropyl Sepharose 6B was incubated at 25°C for 10 min with 1 mM BiotinAMP in 0.1 M imidazole (pH 7.0). The biotinyl thioester-CoA-RNA was then captured on a Neutravidin affinity column and reverse-transcribed directly.[15] After seven cycles, differently sized RNA, internally labeled with α-^{32}P-ATP, from each round was separated by PAGE and quantitated by phosphorimaging. Molar ratios were calculated by relative intensity of the different size groups corrected by the number of A's in these groups—1:1.48:2.35:3.00 for 30N, 60N, 100N and 140N respectively.

Isolation of 100N and 140N sequences. Although the 100N and 140N size groups seemed to have disappeared in the original thioester synthetase selection,[15] great efforts were made to isolate sequences from the 100N and 140N size groups. The PCR DNA at the end of the previous selection[15] was fractionated by PAGE. At the locations of the 100N and 140N size groups, gel pieces were cut. A carrier tRNA was added to the gel pieces and then RNA was recovered by elution and EtOH precipitation. Any 100N and 140N DNA in the gel pieces would be coprecipitated with the carrier tRNA. PCR was then performed on the recovered samples. Surprisingly, only the 30N and 60N size groups were visible by EtBr staining. The same procedure was repeated two more times in series until the 100N group became visible. Standard procedure was then used to clone and sequence the 100N size group. Thioester-synthesizing activity of the individual clones was assayed[15] to select active ribozymes. Efforts to isolate active 140N sequences were not successful.

Kinetic analysis. To evaluate the relationship between RNA size and catalytic activity, a comprehensive kinetic analysis was performed on all thioester synthetase ribozyme sequences from each of the three size groups —30N, 60N and 100N. Internally ^{32}P-labeled pure CoA-RNA (~ 0.1 μM) (purified by thiopropyl Sepharose 6B)[15] was incubated with various concentrations of the substrate BiotinAMP at 25°C in the selection buffer (20 mM HEPES, 100 mM NaCl, 20 mM $MgCl_2$ and 20 mM $CaCl_2$, pH 7.4). Aliquots were then taken at various time points, quenched with 40 mM EDTA and ethanol precipitated. Recovered RNA was incubated with 1μg streptavidin for 2 min and analyzed by gel-shift assay.[15] Apparent rate constant k and inactive ribozyme population α were obtained by curve-fitting thioester yield-time data to equation 1 using Sigma Plot program. Enzyme kinetic parameters k_{cat} and K_M were then derived from the k-[BiotinAMP] data based on the Michaelis-Menten equation by curve-fitting.

RESULTS

Optimization of conditions and control selections. We have previously isolated thioester-synthesizing ribozymes from size-heterogeneous pools containing random regions of 30, 60, 100 and 140 nucleotides (termed 30N, 60N, 100N, and 140N) (total number of nucleotides—62, 92, 146 and 186 nt, respectively) in an initial molar ratio of 1:1:1:1.[15] Prior to the ribozyme selection, a background selection (background selection A) with the size-heterogeneous pools was performed to determine if the selection conditions used in our laboratory favored any particular size group. Standard PCR conditions, with five units of *Taq* DNA polymerase per 100 μL PCR reaction and primer extension for 60–90 sec at 72°C, had been used previously.[14] Transcription and reverse transcription conditions were also standard.[16] Following seven rounds of PCR, transcription, and reverse transcription, the population of larger sizes was reduced significantly (Fig. 1A). Further rounds of background selection were not performed due to the level of the 140N size group falling below the detection limit of phosphorimaging. The relative abundance of different size groups after n rounds of selection can be described as:

$$\text{Relative Abundance} = \beta_i^n \quad (1)$$

where β_i is the combined amplification cycling efficiency (replication efficiency) of the size group i, which includes the efficiencies for PCR, transcription, and reverse transcription. Defining β_{30N} for the 30N size group as 1, the relative amplification cycling efficiencies of 60N, 100N, and 140N size groups were determined to be 0.88, 0.57 and 0.44, respectively, by curve-fitting to equation 1. After ten rounds of background amplification cycles, the 100N and 140N size group would be present in insignificant amounts relative to the two smaller sizes (30 and 60N), with the ratio of

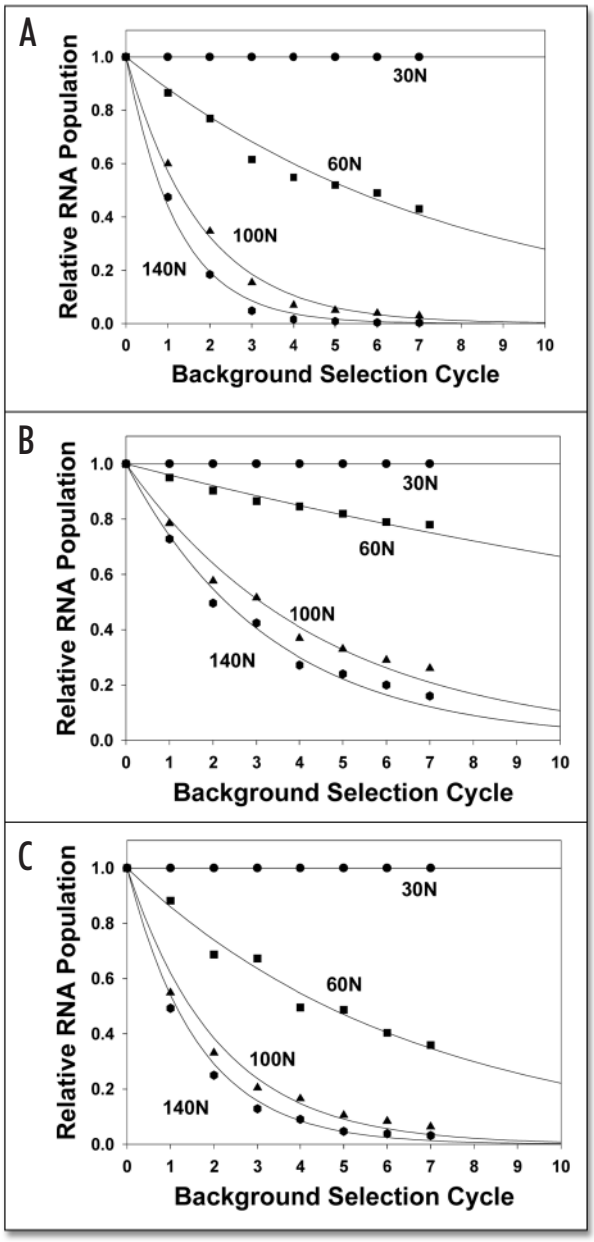

Figure 1. Relative RNA size population (background) for a size-heterogeneous RNA library containing four different random regions (30N, 60N, 100N and 140N) at successive amplification cycles. (A) Background selection A, seven cycles of PCR, RT, and transcription were performed under the standard PCR amplification conditions (used previously)[14] of 5 units *Taq* polymerase/100 μL PCR reaction, and primer extension time of 1 min at 72˚C. (B) Background selection B, optimized PCR amplification conditions to minimize bias toward short DNA amplification. The optimized conditions were 10 units *Taq* polymerase per 100 μL PCR reaction, and 4 min primer extension at 72˚C. (C) Background selection C, same PCR and transcription conditions as B, but adding a step of imidazole-catalyzed RNA thioesterification, and performing reverse transcription directly on Neutravidin beads. Solid symbols represent experimental data, and curves are from calculations based on Equation 1.

30N:60N:100N:140N = 1:0.28:0.004:0.0003, in the absence of catalytic activity. We therefore strove to improve selection conditions that would reduce the distinct bias against the longer RNA sizes. PCR was optimized under the conditions of 10 units of *Taq* polymerase/100 μL PCR reaction and the primer extension time of 4 minutes. Similar PCR conditions were used for ribozyme selection from long random RNA pools.[18] Under the optimized conditions (background selection B), the relative amplification cycling efficiencies of large size groups improved significantly (β_i = 1, 0.96, 0.80, 0.74 for the 30N, 60N, 100N and 140N size groups), although the bias against longer RNA pools could not be eliminated (Fig. 1B). After ten rounds, the relative size distribution would be 30N:60N:100N:140N = 1:0.66:0.11:0.05. Varying transcription or reverse transcription conditions did not reduce the bias further. Therefore, transcription and reverse transcription were performed under the standard conditions.[16] Although the relative abundance of 100N and 140N is low, the two large size groups should be easily detectable after 10 selection rounds.

Our previous selection of thioester-synthesizing ribozymes from size-heterogeneous random libraries resulted in near disappearance of 100N and 140N.[15] We thus sought to assess whether the specific procedure used to isolate the ribozymes could have led to the results. Two major differences existed between the above optimized background selection B (Fig. 1B) and the actual thioester-synthesizing ribozyme selection.[15] During the actual selection, CoA-initiated transcription was used to produce random RNA with a covalent 5′ CoA,[19] and the reverse transcription step was performed directly on Neutravidin beads. To compare the difference between the two sets of selection conditions, another control selection (background selection C) was performed under the same ribozyme selection conditions[15] minus the reaction step (RNA self-thioesterification). The reaction step between pure CoA-RNA and the substrate BiotinAMP was replaced by imidazole-catalyzed thioesterification, whose yield reaches >95%.[17] Therefore, this third control selection should establish the amplification cycling background of different RNA sizes under the same ribozyme selection conditions, but without the selection pressure for RNA catalytic activities.

Figure 1C shows the relative RNA size distribution under the background selection C. From equation 1, the combined relative amplification cycling efficiencies (including PCR, transcription, RNA thioesterification by imidazole catalysis, and direct reverse transcription on Neutravidin beads) were obtained as 30N:60N:100N:140N = 1:0.86:0.62:0.54. The results from background selection C closely resemble those from background selection A, indicating that RNA thioesterification by imidazole catalysis and direct reverse transcription on Neutravidin beads are also biased against large RNA sizes, in addition to PCR amplification. The optimized PCR conditions (background selection B) had reduced the bias against large RNA sizes, but the improvement in the PCR step was largely cancelled by imidazole-catalyzed RNA thioesterification and direct reverse transcription on Neutravidin beads. Even so, background selection C represented the best effort to optimize the selection conditions for the purpose of reducing replication bias against long RNA sequences and deriving the relationship between RNA size and catalytic activity by competitive in vitro selection experiments.

Isolation of 100N and 140N sequences. Seemingly disappearing from our previous selection,[15] the two larger size groups (100N and 140N) were actually present in extremely low concentrations after ten rounds of selection for thioester synthetase activity. Following exhaustive amplification and purification, 22 clones were ultimately isolated from the 100N size group. Attempts to isolate active sequences from the 140N groups failed. Of the 22 clones isolated from the 100N size group, 12 showed thioester synthetase activity above background. Figure 2 shows their sequences, along with previously isolated 30N and 60N sequences[15] for comparison.

Kinetic analysis. Kinetic analysis of the thioester-synthesizing activity for each isolated active sequence from the 30N, 60N and 100N size groups was conducted to determine their relative catalytic efficiencies and to establish a relationship between random region size and ribozyme efficiency. Kinetics were performed by streptavidin gel shift assay.[15] A representative gel shift kinetic assay is shown in Figure 3. We found that all analyzed sequences displayed a common property of a fraction of the ribozyme population being inactive. In fact, most published ribozymes exhibit the same property, most likely due to RNA misfolding.[20] We therefore define the inactive portion of the ribozyme population as α, whose value may change with different sequences and sizes. At each substrate (BiotinAMP) concentration, the observed rate constant (k) was derived from curve-fitting the following Equation 2. The resulting k data were then fitted to the Michaelis-Menten

Clone	Sequence	Repeats
30N Size Group		
TES 1	AUUUCCCGGCUUUAAUACGUUAUUGACUAA	9
TES 2	AAUUCACUUCUCGGCUUUAAUACGUGUAAA	2
TES 4	AACUACUUUCCGGCUUUAAUACGCCUCAAA	
TES 10	UAUUCACUUCCCGGCUUUAAUACGGGCAAC	
TES 12	UUGAAAGGCCACAUUUCCUGCCCGGUACCU	
TES 16	UUGAAAUGAGCACAGGUCCGAGUCACAU	
TES 17	UUACACUUUCCGGCUUUAAUACGUCUCAGA	
TES 22	AUCGCCCUUUCCGGCUUUAAUACGUGCAAA	
TES 25	AUCGUCCUUCCCGGCUUUAAUACGUGCAAA	
TES 28	UUAAUGCACUUUAACUUCGCGCCCGACCCU	
TES 30	UUAGACUGUCCGGCGGUAAUACGUCUCAGA	
60N Size Group		
TES 31	ACACUUGCUGGUGUACGCGCCCCCUUGCGUACUCUGCCCUUCCGCGUCUACCCCGUCCAA	
TES 33	ACACUUGCUGGUGUACGCGCCCCCUUGCGUACUCUGCCCUUCCGCGUCUCCCCGUCCAA	7
TES 34	UACUCUAAGCACCUGCAGUAAAGUAGUUGCGGCACUCCCGGCUCUAAUACGGACACCAUC	
TES 35	GGAUACGCGAGCACCUUCUGGCUUAUUAGACUUUCUAAUAUCUGCCCUUCGAGCCCCCAU	
TES 36	UACUCCAAGCUCCGGCGGUAAAGUAGUCGCGGCACUCCCGGCUUUAAUGCGGACAUCAUC	
TES 37	GGAUAUGCGAGCACCUUCUGGCUUAUUAGACUUUCUAAUAUCUGCCCUUCGAGCCCCCAU	2
TES 40	AGUUACGCGAGCACCUUCUGGCUUAUUAGACUUUCUAAUAUCUGCCCUUCGAGCCCUCAU	
TES 41	ACACUUGCUGGUGUACGCGCCCCCUUGAGUACUCUGCCUUACGAGUCUCCCGUCCAACA	
TES 43	AGCUACGCGAGCACCUUCUGGCUUAUUAGACUUUCUAAUAUCUGCCCUUCGAAGCCCUCAA	
TES 44	GGAUACGAGAGCACCUUCUGGCUUAUUAGACUUUCUAAUAUCUGCCCUUCGAGCCCCAAU	
TES 45	GAAUAUCGCGAGCACCUUCUGGCUUAUUAGACUUUCUAAUAUCUACCCUUCGAGCCCCCUA	
TES 47	UACUCUAAUCUCCGGCACUAAAGUAGUCGCGGCACUCCCGGCUCUAAAUACGGACAUCAUC	
TES 48	UUCACCCACUCCGGGUGCGGUGUGCCCAAACCUCUUCCUACGCCGCAUUAGCCUAAGCAU	
100N Size Group		
TES61	CAGGCACGUGCCGCCCAUCUAUCCGGUGGGGUCUUUCAGAUAUGUGGUCGGAAGUAUGAUUCCGACGUCUACGGCAUGCGCGACCUACAACAGCACGCCUAGUCGGGCCUUAC	4
TES64	UCACGACAGCUUAGGUUGUCGUCCUCUAUAAUGGGGAGUAGGCGCCGAUAGGCGCUGGAAGGCACAUCGUGCCAGCACGUGCGACCCAGUGAAGGCUACUCAGUCCUCACUCGU	
TES71	AAUACGGUUAGGCAACUCGCUUAAACGCGGAGCAUCCAGUACUGUAAAUGGGCACUUAAACCGCCGACCCUCGGCACGUUCGGCCAGCAGUCCCUUGCGAAGAUCUUCUCAUG	
TES72	UCUCGACAGUUUAGGUUGUAGUCCUGUAUAAUGGGGAGUAGGUGCCGAUAGGCGCGUGCAAUAAAAUUGUGCCAGCACGUGCGACCCAGUGCGGGCUACUCAGUUCUCACUUGUG	
TES73	ACUCACUAUUAUUGAAGUGCUACCUCACGACAGUUUAGAUUGUCGUCCUCUAUAAUGGGGAGUACGCUCCCAGAGGCCCUGUUAACCAUAAAUGACCACAACCUAACCUGUCCG	2
TES75	GCCGCCCGGAUACCUGAGGUACGGAUAGUGCGUAGCAUCGGCCUGUGGCAGCGCGUAACGGAGUCCUCUACCGGCACGUGCGACCACACGAUUCCUAAGACCCUGCUGUGAUAG	
TES77	UUUUGCACUUCGAUAUCCGCCGGAGCUCUAAAUACGCUAUGUACCGCCCGCCGCCGAUACGGGCCAUCUUGACCGCACGUGUCAGCAACCAUUCUCCGCCUACCUAAUCAACAG	
TES80	ACAUUGGAGGAGGUCUCCUCCAUGCGUGUGCUAGCUCGGCCUAGCCUGGGUUCGGUCGGGUGUUCAUGCAUACGGCACGUUCAGCCAAUCCCGAAGCUACGGUUACCCAUGGUCG	

Figure 2. Isolated RNA sequences from the 30N, 60N and 100N size groups. RNA sequences from the 30N and 60N group have been shown previously.[15]

equation to derive the ribozyme kinetic parameters of k_{cat} and K_M.

$$\text{RNA Product Yield} = (1 - \alpha)(1 - e^{-kt}) \qquad (2)$$

The results for individual sequences are shown in Table 1. Also included are averaged kinetic parameters (repeated clones are accounted for their frequency) for each size group from the individual sequences, as well as those determined from the final pools of different sizes at the end of selection (10 rounds).[15] For the 30N group, TES12 and TES16 do not possess activity great enough to allow accurate determination of their kinetic parameters, and therefore are not included in this table. TES25 has the same sequence as TES22 except for a single mutation, and TES30 differs from TES17 by three mutations. TES25 and TES30 were thus assumed to behave like TES22 and TES17, respectively. For the 60N group, only four representative sequences were analyzed. For calculation purposes, we also determined apparent kinetic parameters (pool k_{cat}, K_M and α) for the final pools (after 10 rounds of selection) of different sizes. Because individual pools of different sizes at the end of selection contained inactive sequences, the pool k_{cat} values are lower than the averaged kcat values, and the α values are higher for the pools than the averaged α values. As can be seen from Table 1, the apparent pool K_M values are lower than the averaged K_M values.

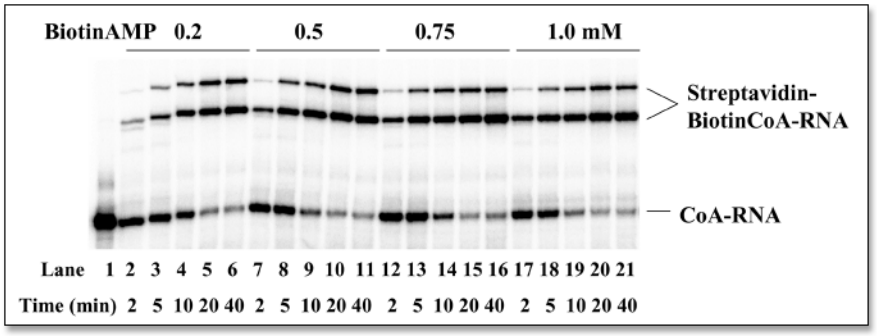

Figure 3. Representative kinetic assay of TES31 by streptavidin gel shift. Lane 1 is the control for unreacted CoA-RNA. The two upper bands are both streptavidin-biotin-CoA-RNA complexes (multimeric forms).[15]

Interestingly, the kinetic parameters (k_{cat} and K_M) of the most abundant size group (30N) are, on average, the poorest of the three groups if judging by the conventional term catalytic efficiency (k_{cat}/K_M). At the same time, the 100N group, which nearly disappeared during the selection,[15] contains quite good catalysts judged by the same term. In fact, TES75 and TES80 (100N ribozymes) are among the top-performing ribozymes in the literature based on their k_{cat} and K_M. However, the 100N group has higher α values than the two shorter 30N and 60N groups. When pool k_{cat}, K_M, α, and amplification cycling efficiencies of different size groups are taken into consideration, relative RNA size distributions between two different groups (A and B) after n selection cycles can be calculated under the selection conditions from the following equation:

$$\frac{A}{B} = \left[\frac{(\text{Product yield} \times \text{cycling efficiency})_A}{(\text{Product yield} \times \text{cycling efficiency})_B}\right]^n = \left[\frac{\left((1-\alpha)(1-e^{-\frac{k_{cat}[\text{substrate}]}{K_M+[\text{substrate}]}t})\times\text{cycling efficiency}\right)_A}{\left((1-\alpha)(1-e^{-\frac{k_{cat}[\text{substrate}]}{K_M+[\text{substrate}]}t})\times\text{cycling efficiency}\right)_B}\right]^n$$

Where A/B represents the relative RNA size population of group A over group B under the conditions: [substrate] = substrate concentration and

Table 1 **Kinetic parameters of isolated thioester-synthetase ribozymes and individual pools at the end of selection***

	Clone**	k_{cat} (min^{-1})	K_M (mM)	α	$(1-\alpha) \cdot k_{cat}/K_M$ ($M^{-1} \cdot min^{-1}$)
30N	TES1 (9)	0.29 ± 0.09	1.3 ± 0.39	0.24 ± 0.10	170 ± 80
	TES2 (2)	0.13 ± 0.04	1.0 ± 0.30	0.12 ± 0.06	110 ± 50
	TES4 (1)	0.10 ± 0.04	0.82 ± 0.33	0.21 ± 0.09	90 ± 60
	TES10 (1)	0.41 ± 0.16	1.5 ± 0.45	0.20 ± 0.08	220 ± 100
	TES17 (2)	0.10 ± 0.04	0.59 ± 0.24	0.19 ± 0.09	140 ± 80
	TES22 (2)	0.13 ± 0.04	0.55 ± 0.17	0.24 ± 0.10	270 ± 110
	TES28 (1)	0.49 ± 0.20	0.53 ± 0.21	0.36 ± 0.14	600 ± 340
	Average	0.24 ± 0.10	1.0 ± 0.4	0.22 ± 0.09	190 ± 80
	Pool	0.21 ± 0.06	0.43 ± 0.17	0.36 ± 0.07	310 ± 120
60N	TES31 (9)	0.40 ± 0.12	0.35 ± 0.14	0.19 ± 0.08	890 ± 160
	TES34 (3)	0.07 ± 0.03	0.69 ± 0.26	0.23 ± 0.11	80 ± 50
	TES35 (7)	0.13 ± 0.05	0.37 ± 0.11	0.15 ± 0.07	230 ± 140
	TES48 (1)	0.47 ± 0.20	1.2 ± 0.4	0.42 ± 0.17	230 ± 170
	Average	0.26 ± 0.10	0.45 ± 0.2	0.19 ± 0.08	470 ±160
	Pool	0.17 ± 0.05	0.17 ± 0.06	0.38 ± 0.08	620 ± 250
100N	TES61 (4)	0.18 ± 0.07	0.17 ± 0.07	0.35 ± 0.16	720 ± 360
	TES64 (1)	0.07 ± 0.03	0.31 ± 0.12	0.18 ± 0.11	190 ± 110
	TES71 (1)	0.39 ± 0.16	0.55 ± 0.17	0.44 ± 0.15	400 ± 200
	TES72 (1)	0.25 ± 0.08	0.48 ± 0.14	0.23 ± 0.09	400 ± 170
	TES73 (2)	0.22 ± 0.07	0.18 ± 0.05	0.20 ± 0.07	960 ± 410
	TES75 (1)	1.7 ± 0.7	0.93 ± 0.37	0.59 ± 0.26	740 ± 410
	TES77 (1)	0.18 ± 0.05	0.26 ± 0.08	0.46± 0.24	380 ± 150
	TES80 (1)	1.2 ± 0.4	0.57 ± 0.23	0.54 ± 0.20	970 ± 510
	Average	0.42 ± 0.17	0.37 ± 0.14	0.36 ± 0.14	770 ± 320
	Pool	0.16 ± 0.05	0.29 ± 0.10	0.43 ± 0.09	310 ± 110

*Kinetic analysis conditions: substrate, BiotinAMP; reaction buffer, 20 mM HEPES, pH 7.4, 100 mM NaCl, 20 mM $MgCl_2$ and 20 mM CaCl2; temperature, 25°C.
**The number of isolated repeats is indicated in parentheses.

t = reaction time. For the calculation of relative RNA size distributions of the 60N and 100N size groups under the selection conditions, A is defined as 30N and B is either 60N or 100N. The substrate concentration and reaction time were 1 mM and 30 min, respectively. The cycling efficiencies of the 30N, 60N and 100N from background selection C were 1, 0.88 and 0.62, respectively.

Figure 4A shows the 60N and 100N relative size distributions of background selection C (solid curves) and the thioester-synthesizing ribozyme selection (solid symbols).[15] Also included are the calculated RNA size populations (dotted curves) from Equation 3 and the pool kinetic parameters (pool k_{cat}, K_M, and α) from Table 1. For both 60N and 100N size groups, the first five rounds of thioester-synthesizing ribozyme selection followed closely the background distributions. The results were expected, because no catalytic activities appeared during these selection rounds.[15] However, the 60N and 100N populations went below the background levels from rounds 6 and 7 and stayed below until the end of the selection. At the same time, thioester-synthesizing ribozyme activities emerged.[15] The result indicated that the 60N and 100N pools were catalytically less efficient than the 30N pool under the selection conditions. To support the conclusion, calculated RNA size distributions (dotted curves) by using experimentally determined kinetic parameters from different RNA pools after ten rounds of selection were consistent with the selection results (solid symbols) after the emergence of catalytic activities.

Relative size populations of isolated active thioester-synthesizing ribozymes can be calculated according to equation 4:

$$\frac{A}{B} = \left[\frac{(\text{Product yield} \times \text{cycling efficiency})_A}{(\text{Product yield} \times \text{cycling efficiency})_B}\right]^n = \left[\frac{nB\left(\sum_{i=1}^{nA}\left[(1-\alpha_i)(1-e^{-\frac{k_{cat}[\text{substrate}]_f}{K_M+[\text{substrate}]_f}})\right] \times \text{cycling efficiency}\right)_A}{nA\left(\sum_{i=1}^{nB}\left[(1-\alpha_i)(1-e^{-\frac{k_{cat}[\text{substrate}]_f}{K_M+[\text{substrate}]_f}})\right] \times \text{cycling efficiency}\right)_B}\right]^n$$

Where nA and nB are the number of isolated active sequences from size groups A and B. While Equation 3 treats an RNA size group (containing a large number of both active and inactive sequences) as one ribozyme with its apparent pool k_{cat}, K_M, and α, equation 4 includes only a limited number of isolated active ribozyme sequences with individual sets of k_{cat}, K_M, and α. Figure 4B presents comparisons between the relative background distributions (solid curves) and calculated (from Equation 4) relative size populations (dotted curves) of isolated active ribozyme sequences. For the 60N group, the background level is lower than the calculated population, indicating the isolated ribozymes from the 60N group are, on the average, better than the average ribozyme isolated from the 30N group under the selection conditions. On the other hand, the background level of 100N is higher than the calculated population from ribozymes isolated from the 100N group, suggesting that an average 100N ribozyme is less efficient than an average 30N ribozyme under the selection conditions. The discrepancy between the two calculated relative size populations (pool vs. isolated ribozyme sequences) of the 60N group may indicate that the final 60N selection pool contained more inactive or less active RNA sequences than those in the 30N group.

The relative frequencies of individual sequences within a defined size group are determined by their individual k_{cat}, K_M, α and the selection conditions, and should allow prediction based on Equation 3, where A and B represent two different sequences of the same length. The amplification cycling efficiencies of different sequences of the same length are similar (extensive cycling experiments of same-size ribozymes could not demonstrate differences in cycling efficiencies, unpublished results). Using the parameters from Table 1, we calculated the relative population of TES1 (repeated nine times) vs. TES28 (once) in the 30N group during the selection (Fig. 5A), as well as the population evolution of TES61 (repeated four times) vs. TES75 and TES80 (once). As shown in (Fig. 5B), TES1 and TES61 should dominate within their corresponding size groups, and isolated multiple clones of the two sequences support the prediction. Predicted ratios between TES1:TES28 is approximately 5:1 compared with the actual ratio of 9:1 from isolation. Calculated TES61:TES75 (TES80) ratio ranges from 30 to 90. These discrepancies may be due to the fact that only a limited number of clones were isolated and kinetically analyzed. Other factors such as varying cloning efficiencies of different sequences and relatively large errors in the kinetic parameters (Table 1) might also have played a role. Nevertheless, the prediction qualitatively describes the observed phenomenon of different frequencies in ribozyme selection.

DISCUSSION

By developing a unique competitive in vitro selection using size-heterogeneous random RNA libraries followed by extensive kinetic analysis of isolated RNA sequences, we have established the relationship between the lengths of random regions of RNA libraries and ribozyme catalytic activities.

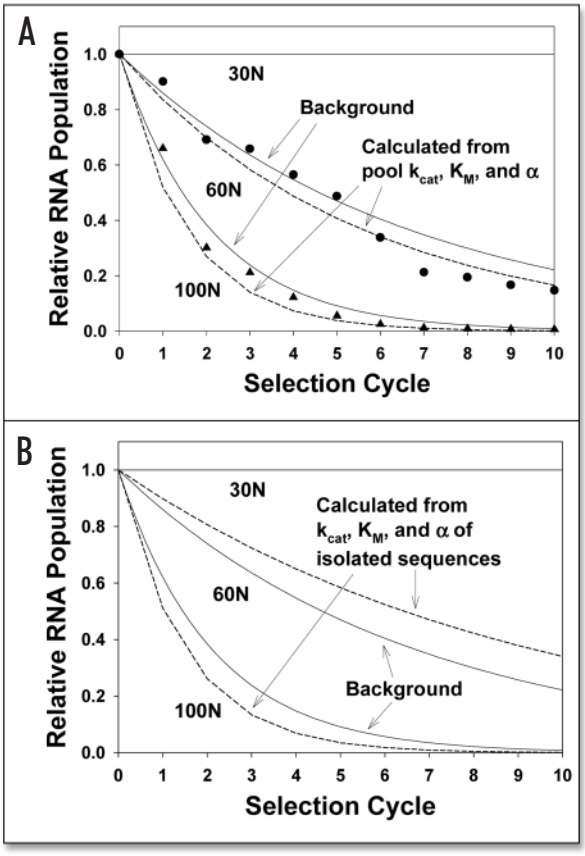

Figure 4. (A) Comparison of relative RNA size populations (solid symbols) between the thioester-synthesizing ribozyme selection[15] and background selection C (solid lines). Dotted lines are from calculations based on Equation 3 and pool kinetic parameters k_{cat}, K_M, and α from Table 1. (B) Background selection C (solid lines) vs calculated relative RNA size populations (dotted curves) of isolated active ribozymes from Equation 4 and individual kinetic parameters k_{cat}, K_M, and α from Table 1.

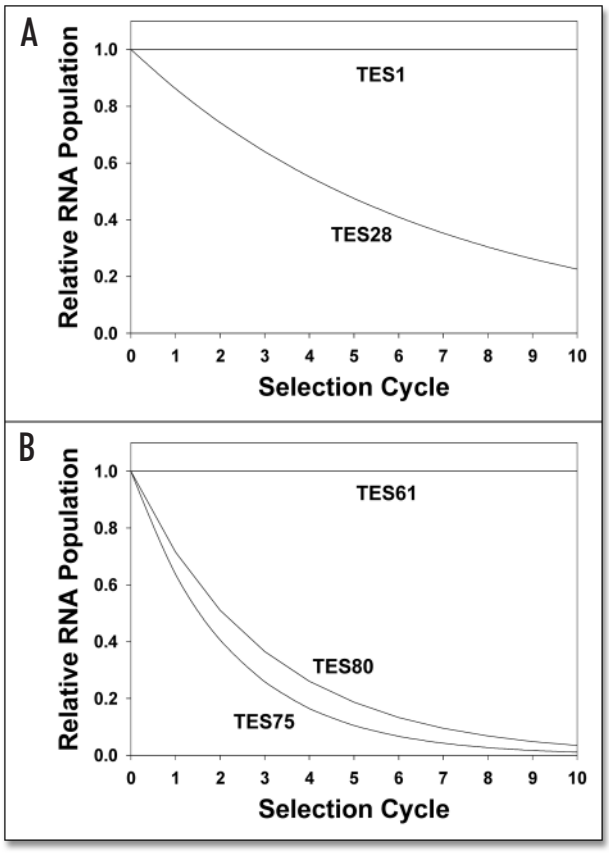

Figure 5. Calculated distributions of individual sequences within size groups based on determined kinetic parameters under the selection conditions. (A) TES1 vs TES28. While TES28 has higher k_{cat}/K_M than that of TES1, its moderately higher α significantly reduces its representation under the selection conditions. (B) TES61 vs TES75 and TES80. TES61 predominated after ten rounds of selection because it possesses lower α.

RNA Size vs. amplification cycling efficiency. In the original preparation and use of the size-heterogeneous pools,[14] the amplification cycling efficiency of different RNA sizes was not taken into consideration. The difference in replication efficiency between different sizes is significant under the previous PCR conditions (Fig. 1A). Optimization of PCR conditions can reduce the difference. However, the bias against long sequences cannot be eradicated (Fig. 1B and C). In a typical selection from size-homogeneous random libraries, the cycling inefficiency of the long sequences may be overcome by increasing the number of PCR cycles. No such compensation can be made for size-heterogeneous random libraries. The results from (Fig. 1) have important implications on RNA synthesis in the RNA world. In the in vitro selection experiments, efficient protein enzymes (the evolution products of billions of years) were used to reproduce RNA under the optimized conditions, and yet, long sequences impose significant replication difficulties under these competitive conditions. In the RNA world, RNA synthesis would presumably have been carried out by less efficient 21 and imperfect RNA polymerase ribozymes as is evidenced by the fact that to date no efficient polymerase ribozymes have been isolated[22] and the lack of an RNA polymerase in extant biology. It is therefore not difficult to imagine that even greater difficulties in long RNA reproduction

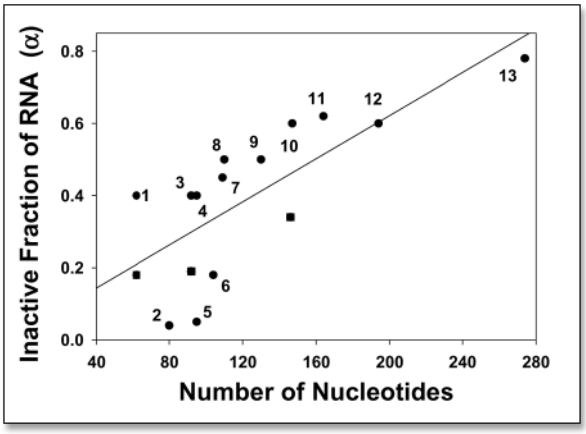

Figure 6. Correlation between ribozyme sizes (nt) and the fraction (α) of inactive RNA population. If the α value of a ribozyme was not explicitly given in the original publication, the active portion of the ribozyme was estimated from either stated activity or kinetic data in the publication. Averaged α values from Table 1 are represented as ■. Literature α values (•) were directly taken or inferred from published results: 1, ref. 14; 2, ref. 38; 3, ref. 14; 4, ref. 39; 5, ref. 35; 6, ref. 40; 7, ref. 41; 8, ref. 42; 9, ref. 43; 10, ref. 44; 11, ref. 33; 12, ref. 45; 13, ref. 7.

would face ribozymes. Regardless of catalytic activities, large RNA sequences confer considerable disadvantage over short RNA sequences in terms of RNA replication.[23]

Ribozyme size vs. alpha. Typically, selected ribozymes cannot reach 100% reactivity because of an inherent inactive fraction of RNA population (we define it as α in Equation 2). Kinetic traps impede correct folding by non-native interactions and must be disrupted to achieve correct folding.[24-28] The longer an RNA sequence is, the greater the probability for misfolding into different inactive conformations.

An alternate explanation of the existence of α is that the ribozyme may be 100% active, but the reaction that the ribozyme catalyzes reaches an equilibrium between the substrates and products. While this may be true of some ribozymes, the reverse reaction does not occur appreciably for the described thioester synthesis reaction under our conditions. Addition of free AMP (one of the reaction products) to the reaction does not affect the extent of reaction (unpublished results). If the reaction were reversible, adding AMP to the solution would shift the equilibrium to the left, leading to lower product yields. Therefore, the α values in Table 1 reflect the inactive fraction of thioester-synthesizing ribozyme population.

Two interesting points emerged after kinetic analysis of the individual thioester synthetase sequences. On one hand, α of different sequences varies significantly within the same size group as well as among different size groups. On the other hand, statistical analysis reveals a correlation between α and ribozyme size. Longer sequences tend to have higher α values. To validate the existence of such a correlation, α values of isolated ribozymes (obtained directly from publications or inferred from published data) were plotted against the number of nucleotides, shown in Figure 6. It is clear from the figure that the α value of ribozymes gets higher when the length of ribozymes increases. The figure includes only primary isolates from in vitro selection and can be used to predict the statistic trend of isolated ribozymes of different sizes. Modified or optimized ribozymes after initial isolation may not follow the trend, and are thus excluded from the figure. In addition, different ribozyme sequences of similar sizes can have varying α values (Table 1). Therefore, specific ribozymes may deviate from the figure. The correlation between ribozyme size and α value suggests that long sequences in a selection experiment may not be necessary for isolation of ribozymes and excessive sequences may actually reduce the catalytic performance of ribozymes by increasing their inactive population (α).

Ribozyme sequence frequency vs. ribozyme efficiency. More efficient ribozyme sequences should gradually dominate pools during the selection. At the end of selection, frequencies of isolated clones should correlate with ribozyme efficiencies (determined by k_{cat}, K_M, and α) and the selection conditions. If ignoring the α effect, Table 1 seems to suggest the best ribozyme sequences do not necessarily appear in high frequencies. This observation is not unique to our selection.[22] We have shown that the discrepancies are caused by the term α and defined selection conditions, as described by Equation 3. It is possible that a good or even an excellent ribozyme by the conventional definition of k_{cat}/K_M escapes isolation by a limited number of clones under defined selection conditions due to its high α value. In fact, none of the sequences in the whole 100N size group were isolated originally in our earlier work.[15]

Ribozyme size vs. ribozyme efficiency. Catalytic efficiency of an enzyme is conventionally defined by k_{cat}/K_M. Since ribozymes have an additional parameter α that can significantly affect catalytic performance, α should be taken into account when comparing relative ribozyme catalytic efficiencies. We suggest using the term ribozyme efficiency $(1 - \alpha)k_{cat}/K_M$ for evaluating catalytic activities of ribozymes. From (Table 1), the ribozyme efficiency of different size groups does not exhibit significant difference (within 3-fold). For individual sequences, regardless of their sizes, k_{cat}/K_M and α tend to compensate for each other. However, since most selection experiments have been performed under permissive conditions (high substrate concentration and long incubation time) that reduce the effect of k_{cat}/K_M on product yields, the term $(1 - \alpha)$ becomes significant during selection, along with the cycling efficiency of the sequence. Because longer sequences have both higher α values and lower reproduction efficiencies, the 100N size group quickly lost its competitiveness during the selection. It should be noted that the catalytic activities of ribozymes are determined not only by their ribozyme efficiencies $[(1 - \alpha)k_{cat}/K_M]$ but also by reaction conditions. To compare the performance of different ribozymes under defined conditions, Equation 3 should be used.

Intuitively, long random RNA libraries have been argued to offer high success rates for selection because of their potential to fold into more complex structures.[6,8,9] However, current results and previous observations argue against this suggestion. First, the RNA misfolding problem is universal for ribozymes. Smaller RNA molecules somewhat circumvent the recognized folding problem in longer RNA sequences.[20,25-29] A given RNA sequence can fold into numerous secondary structures with similar folding energy, which may be active or inactive. Each possible conformation could account for an inactive fold, thereby increasing the likelihood that the molecule will misfold. Second, many in vitro selected ribozymes can be truncated into smaller ones, some by significant portions, and the resulting ribozymes retain the same catalytic efficiencies or perform even better.[6,12,18,30-35] It has also been shown that small motifs are optimal in sequence space for RNA cleavage activity.[36,37] Third, large sequences pose the inevitable disadvantage of replication inefficiency as shown quantitatively in Fig. 1. This reproduction advantage of small RNA over large RNA has been recognized and discussed.[23]

We believe that there exists an optimal length of relatively short random regions around 60 nt for the isolation of functional RNA sequences. Our previous coenzyme synthetase ribozyme selection resulted in the 60N group as the dominating population.[14] Although the 60N group was less abundant than the 30N at the end the competitive thioester-synthesizing ribozyme selection,[15] the isolated 60N ribozymes were calculated (by Equation 4 but omitting the cycling efficiency term) to perform slightly better than the 30N (1.04:1) under the selection conditions. On the other hand, the relative performance of the isolated 100N thioester-synthesizing ribozymes over the isolated 30N ribozymes was calculated to be 0.86:1 when factoring k_{cat}, K_M, α and cycling efficiency under the selection conditions. In addition, we have isolated ester-synthesizing ribozymes from the same initial random RNA libraries, and the 60N size group became dominant after seven cycles of selection (unpublished results). Therefore, three independent competitive selection experiments have yielded the same result—best performing ribozymes (on average) coming from the 60N group among the 30N, 60N, 100N and 140N size groups under the selection conditions. Although the conclusion of an optimal 60N random RNA library is drawn from three specific competitive selection experiments, the conclusion may be generally applicable to other competitive in vitro selection experiments.

Our conclusion of an optimal random RNA size of 60 nt is based on competitive selection experiments, and does not exclude the possibility of isolating highly efficient ribozymes from RNA libraries containing longer random regions under noncompetitive conditions. Without the selection pressure for reproduction efficiency under noncompetitive conditions, it is still an open question whether long random regions can lead to superior ribozymes or are necessary for complex functions. The question can be addressed by parallel selections (under the same conditions) of defined RNA functionalities from RNA libraries, each containing a random region of different length.

References

1. Ellington AD, Szostak JW. In vitro selection of RNA molecules that bind specific ligands. Nature 1990; 346:818-22.
2. Tuerk C, Gold L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science 1990; 249:505-10.
3. Robertson DL, Joyce GF. Selection in vitro of an RNA enzyme that specifically cleaves single- stranded DNA. Nature 1990; 344:467-8.
4. Giver L, Bartel D, Zapp M, Pawul A, Green M, Ellington AD. Selective optimization of the Rev-binding element of HIV-1. Nucleic Acids Res 1993; 21:5509-16.
5. Tuerk C, MacDougal-Waugh S. In vitro evolution of functional nucleic acids: High-affinity RNA ligands of HIV-1 proteins. Gene 1993; 137:33-9.
6. Bartel DP, Szostak JW. Isolation of new ribozymes from a large pool of random sequences. Science 1993; 261:1411-8.
7. Unrau PJ, Bartel DP. RNA-catalysed nucleotide synthesis. Nature 1998; 395:260-3.
8. Sabeti PC, Unrau PJ, Bartel DP. Accessing rare activities from random RNA sequences: The importance of the length of molecules in the starting pool. Chem Biol 1997; 4:767-74.
9. Carothers JM, Oestreich SC, Davis JH, Szostak JW. Informational complexity and functional activity of RNA structures. J Am Chem Soc 2004; 126:5130-7.
10. Lorsch JR, Szostak JW. In vitro evolution of new ribozymes with polynucleotide kinase activity. Nature 1994; 371:31-6.
11. Tarasow TM, Tarasow SL, Eaton BE. RNA-catalysed carbon-carbon bond formation. Nature 1997; 389:54-7.
12. Seelig B, Jaschke A. A small catalytic RNA motif with Diels-Alderase activity. Chem Biol 1999; 6:167-76.
13. Johnston WK, Unrau PJ, Lawrence MS, Glasner ME, Bartel DP. RNA-catalyzed RNA polymerization: Accurate and general RNA-templated primer extension. Science 2001; 292:1319-25.
14. Huang F, Bugg CW, Yarus M. RNA-Catalyzed CoA, NAD, and FAD synthesis from phosphopantetheine, NMN, and FMN. Biochemistry 2000; 39:15548-5.
15. Coleman TM, Huang F. RNA-catalyzed thioester synthesis. Chem Biol 2002; 9:1227-36.
16. Ciesiolka J, Illangasekare M, Majerfeld I, et al. Affinity selection-amplification from randomized ribooligonucleotide pools. Methods Enzymol 1996; 267:315-35.
17. Coleman TM, Li N, Huang F. A simple and efficient method to prepare thioesters in aqueous solutions. Tetrahedron Lett 2005; 46:4307-10.
18. Ekland EH, Szostak JW, Bartel DP. Structurally complex and highly active RNA ligases derived from random RNA sequences. Science 1995; 269:364-70.
19. Huang F. Efficient incorporation of CoA, NAD and FAD into RNA by in vitro transcription. Nucleic Acids Res 2003; 31:e8.
20. Uhlenbeck OC. Keeping RNA happy. RNA 1995; 1:4-6.
21. Doudna JA, Lorsch JR. Ribozyme catalysis: Not different, just worse. Nat Struct Mol Biol 2005; 12:395-402.
22. Lawrence MS, Bartel DP. New ligase-derived RNA polymerase ribozymes. RNA 2005; 11:1173-1180.
23. Joyce GF. Directed evolution of nucleic acid enzymes. Annu Rev Biochem 2004; 73:791-836.
24. Rook MS, Treiber DK, Williamson JR. Fast folding mutants of the Tetrahymena group I ribozyme reveal a rugged folding energy landscape. J Mol Biol 1998; 281:609-20.
25. Treiber DK, Rook MS, Zarrinkar PP, Williamson JR. Kinetic intermediates trapped by native interactions in RNA folding. Science 1998; 279:1943-6.
26. Treiber DK, Williamson JR. Exposing the kinetic traps in RNA folding. Curr Opin Struct Biol 1999; 9:339-45.
27. Pan J, Thirumalai D, Woodson SA. Folding of RNA involves parallel pathways. J Mol Biol 1997; 273:7-13.
28. Pan T, Sosnick TR. Intermediates and kinetic traps in the folding of a large ribozyme revealed by circular dichroism and UV absorbance spectroscopies and catalytic activity. Nat Struct Biol 1997; 4:931-8.
29. Herschlag D. RNA chaperones and the RNA folding problem. J Biol Chem 1995; 270:20871-4.
30. Chapman KB, Szostak JW. Isolation of a ribozyme with 5'-5' ligase activity. Chem Biol 1995; 2:325-33.
31. Ekland EH, Bartel DP. The secondary structure and sequence optimization of an RNA ligase ribozyme. Nucleic Acids Res 1995; 23:3231-8.
32. Illangasekare M, Yarus M. A tiny RNA that catalyzes both aminoacyl-RNA and peptidyl-RNA synthesis. RNA 1999; 5:1482-9.
33. Jenne A, Famulok M. A novel ribozyme with ester transferase activity. Chem Biol 1998; 5:23-34.
34. Landweber LF, Pokrovskaya ID. Emergence of a dual-catalytic RNA with metal-specific cleavage and ligase activities: The spandrels of RNA evolution. Proc Natl Acad Sci USA 1999; 96:173-8.
35. Illangasekare M, Yarus M. Small-molecule-substrate interactions with a self-aminoacylating ribozyme. J Mol Biol 1997; 268:631-9.
36. Salehi-Ashtiani K, Szostak JW. In vitro evolution suggests multiple origins for the hammerhead ribozyme. Nature 2001; 414:82-84.
37. Cruz RP, Withers JB, Li Y. Dinucleotide junction cleavage versatility of 8-17 deoxyribozyme. Chem Biol 2004; 11:57-67.
38. Jayasena VK, Gold L. In vitro selection of self-cleaving RNAs with a low pH optimum. Proc Natl Acad Sci USA 1997; 94:10612-7.
39. Illangasekare M, Yarus M. Specific, rapid synthesis of Phe-RNA by RNA. Proc Natl Acad Sci USA 1999; 96:5470-5.
40. Huang F, Yarus M. A calcium-metalloribozyme with autodecapping and pyrophosphatase activities. Biochemistry 1997; 36:14107-19.
41. Williams KP, Ciafre S, Tocchini-Valentini GP. Selection of novel Mg(2+)-dependent self-cleaving ribozymes. EMBO J 1995; 14:4551-7.
42. Tsukiji S, Pattnaik SB, Suga H. An alcohol dehydrogenase ribozyme. Nat Struct Biol 2003; 10:713-7.
43. Lohse PA, Szostak JW. Ribozyme-catalysed amino-acid transfer reactions. Nature 1996; 381:442-4.
44. Wiegand TW, Janssen RC, Eaton BE. Selection of RNA amide synthases. Chem Biol 1997; 4:675-83.
45. Zhang B, Cech TR. Peptide bond formation by in vitro selected ribozymes. Nature 1997; 390:96-100.