Directed Evolution of an RNA Enzyme

Amber A. Beaudry and Gerald F. Joyce

An in vitro evolution procedure was used to obtain RNA enzymes with a particular catalytic function. A population of 10^{13} variants of the *Tetrahymena* ribozyme, a group I ribozyme that catalyzes sequence-specific cleavage of RNA via a phosphoester transfer mechanism, was generated. This enzyme has a limited ability to cleave DNA under conditions of high temperature or high $MgCl_2$ concentration, or both. A selection constraint was imposed on the population of ribozyme variants such that only those individuals that carried out DNA cleavage under physiologic conditions were amplified to produce "progeny" ribozymes. Mutations were introduced during amplification to maintain heterogeneity in the population. This process was repeated for ten successive generations, resulting in enhanced (100 times) DNA cleavage activity.

Biological systems have a remarkable capacity to generate a diverse array of macromolecular catalysts. These catalysts, composed of either protein or RNA, appear to be the result of Darwinian evolution operating at the organismal level. The sequence and corresponding function of a macromolecular catalyst is a reflection of alternatives that have been adopted over evolutionary history and that confer selective advantage to the organism.

The need for catalysts that operate outside of their native context or catalyze reactions that are not represented in nature has resulted in the development of "enzyme engineering" technology. The usual route taken in enzyme engineering has been site-directed mutagenesis—alteration of the gene that encodes the catalyst—usually guided by structural and phylogenetic information (1). Site-directed mutagenesis has been improved by in vitro selective amplification techniques for generating large numbers of mutants with subsequent selection of some desirable property. Individual macromolecules are selected, and those selected are then amplified to generate a progeny distribution of favorable mutants. The process is repeated until only those individuals with the most desirable properties remain. This technique has been applied to RNA molecules in solution (2–4), RNA's bound to a ligand that is attached to a solid support (5), peptides attached directly to a solid support (6), and peptide epitopes expressed within a viral coat protein (7).

The process of Darwinian evolution, by which enzymes arise in nature, does not operate by generating a diverse population of variants and harvesting the most advantageous individuals. In biological systems, diversity is maintained by ongoing mutations, and the population is shaped by selection. Novel mutations augment existing variation, so that the evolutionary search is biased, in an appropriate fashion, by selection events that have already occurred (8). The more advantageous mutants, which are relatively abundant in the population, give rise to larger numbers of novel variants compared to the less advantageous.

We now describe our system of in vitro evolution as a method for enzyme engineering. We began with the *Tetrahymena* ribozyme, an RNA enzyme that catalyzes sequence-specific phosphoester transfer reactions that result in cleavage or ligation of RNA substrates (9–11). This enzyme can be used to cleave a single-stranded DNA substrate, albeit only under conditions of high temperature (50°C) or high $MgCl_2$ concentration (50 mM), or both (4). A kinetic study (12) showed that, even at 50°C, this reaction is inefficient compared to the "native" reaction with an RNA substrate. Under physiologic conditions (37°C, 10 mM $MgCl_2$), the DNA cleavage reaction is almost undetectable. In our study, we used directed evolution, maintaining a population of 10^{13} ribozymes over ten successive generations, to obtain ribozymes that cleave DNA with improved efficiency under physiologic conditions. We have complete access to genotypic and phenotypic parameters for the entire population over the course of its evolutionary history.

In vitro evolution. Darwinian evolution requires the repeated operation of three processes: (i) Introduction of genetic variation; (ii) selection of individuals on the basis of some fitness criterion; (iii) amplification of the selected individuals. Each of these processes can be realized in vitro (3). A gene can be mutagenized by chemical modification (13), incorporation of randomized mutagenic oligodeoxynucleotides (14, 15), or inaccurate copying by a polymerase (16, 17). The gene product can be selected, for example, by its ability to bind a ligand or to carry out a chemical reaction (2–5). The gene that corresponds to the selected gene product can be amplified by a reciprocal primer method, such as the polymerase chain reaction (PCR) (18). A major obstacle to realizing Darwinian evolution in vitro is the need to integrate mutation and amplification, both of which are genotype related, with selection, which is phenotype related. In the case of RNA enzymes, for which genotype and phenotype are embodied in the same molecule, the task is simplified.

Using a combination of two polymerase enzymes, we can amplify virtually any RNA (19). RNA is copied to a complementary DNA (cDNA) with reverse transcriptase (RT), and the resulting cDNA is transcribed to RNA with T7 RNA polymerase (T7 Pol) (Fig. 1A). Amplification occurs during transcription as a consequence of the ability of T7 RNA polymerase to generate 200 to 1200 copies of RNA transcript per copy of cDNA template (20). The amplification reaction is done in a single test tube at a constant temperature of 37°C, resulting in an increase of 10^3 to 10^6 times the original input of RNA after 1 hour (21).

The amplification was performed selectively in that individual RNA's in the population were required to catalyze a particular chemical reaction in order to become eligible for amplification (3, 4). The selection was based on the ability of group I ribozymes to catalyze a sequence-specific phosphoester transfer reaction involving an oligonucleotide (or oligodeoxynucleotide) substrate (Fig. 1B). The product of the reaction was a molecule that contained the 3' portion of the substrate attached to the 3' end of the ribozyme (EP). Selection occurred when an oligodeoxynucleotide primer was hybridized across the ligation junction and used to initiate cDNA synthesis. The primer did not bind to unreacted starting materials ($<10^{-8}$ compared to reaction products) and thus led to selective amplification of the catalytically active RNA's.

Mutations were introduced in two ways. First, at the outset, we used a set of mutagenic oligodeoxynucleotides that contained random substitutions at a fixed frequency of occurrence. These partially randomized oligonucleotides were produced on an automated DNA synthesizer with nucleoside 3'-phosphoramidite solutions that had been doped with a small percentage of each of the three incorrect monomers (15). Second, after each round of selective amplification, we introduced random mutations by performing the PCR under mutagenic conditions (17, 22). The RNA's obtained by selective amplification were subjected to

Departments of Chemistry and Molecular Biology, The Scripps Research Institute, La Jolla, CA 92037.

reverse transcription, the resulting cDNA's were PCR amplified, and the PCR products were transcribed to produce a progeny distribution of mutant RNA's.

Integration of the PCR with the selective RNA amplification procedure was useful in three other ways. First, it increased the overall amplification by about 10^3 times. Second, it simplifies the process of subcloning individuals from the evolving population. Normally, only a small portion of the DNA in the RNA amplification mixture is fully double-stranded, but with the PCR the amount of double-stranded DNA is greatly increased. Third, it returns the RNA to a form that can participate in

the RNA-catalyzed phosphoester transfer reaction. After phosphoester transfer, the ribozyme has the 3' portion of the substrate attached to its 3' end, and after selective RNA amplification, the substrate sequence remains attached (Fig. 1). However, by subsequent use of the PCR, followed by in vitro transcription, the original 3' end of the ribozyme is restored.

We refer to the entire series of events, beginning with a heterogeneous population of RNA's, proceeding with RNA catalysis in the target reaction, selective amplification of catalytically active RNA's, reverse transcription of the selective amplification products, mutagenic PCR, and in vitro

transcription to produce a progeny distribution of RNA's, as one "generation." Typically, a generation is completed in one to two working days, excluding time for analytic work. We refer to the initial population of mutant RNA's as "generation 0" and to subsequent progeny populations as "generation 1," "generation 2," and so forth. In principle, there is no limit to the number of successive generations that can be obtained (23).

Improved catalytic function. The *Tetrahymena* ribozyme is a self-splicing group I intron derived from the large ribosomal RNA (rRNA) precursor of *Tetrahymena thermophila*. Its biological function is to catalyze its own excision from precursor rRNA to produce mature rRNA. This function has been expressed in vitro (9) and has been generalized to include various phosphoester transfer reactions involving RNA substrates (10, 24). For example, the ribozyme has been used as a sequence-specific endoribonuclease (11, 25), a reaction that proceeds with high catalytic efficiency ($k_{cat}/K_m = 10^8$ M^{-1} min^{-1}) (26). As mentioned previously, the ribozyme can also act as a sequence-specific endodeoxyribonuclease (4), although the efficiency of DNA cleavage is low ($k_{cat}/K_m = 200$ M^{-1} min^{-1}, determined at 50°C, 10 mM MgCl$_2$) (12). The efficiency of RNA-catalyzed DNA cleavage under physiologic conditions is even lower ($k_{cat}/K_m = 36$ M^{-1} min^{-1}, determined at 37°C, 10 mM MgCl$_2$; see below).

Our goal was to improve the catalytic efficiency of RNA-catalyzed DNA cleavage under physiologic conditions and thereby obtain ribozymes that could cleave DNA in vivo. It is not obvious how one should change the *Tetrahymena* ribozyme to convert it from an RNA-cleaving to a DNA-cleaving enzyme. Thus, we turned to directed evolution as a means to acquire the desired phenotype.

The ribozyme consists of 413 nucleotides and assumes a well-defined secondary and tertiary structure that is responsible for its catalytic activity (27–29). Phylogenetic analysis (30), supported by site-directed mutagenesis and deletion studies (31), points out a distinction between a conserved catalytic core (comprising about one-third of the molecule) and surrounding stem-loop elements that offer structural support but are not essential for catalytic activity. To generate the initial population of ribozyme variants, we introduced random mutations throughout the catalytic core of the molecule. Four synthetic oligodeoxynucleotides were prepared, each of which randomly mutagenizes 35 nucleotide positions at an error rate of 5 percent per position (Fig. 2). The degenerate oligodeoxynucleotides were incorporated into a DNA template that encodes the ribozyme, and the template was transcribed directly to produce the mutant RNA's (32). We began with 20 pmol (10^{13}

Fig. 1. Selective amplification of catalytic RNA. (**A**) Procedure for amplification of RNA. Reaction conditions: 10^{-5} to 10^1 nM RNA, 1 μM primers, 2 mM each nucleoside triphosphate (NTP's), 0.2 mM each deoxynucleoside triphosphate (dNTP's), 10 mM MgCl$_2$, 50 mM tris-HCl (pH 7.5), 5 mM dithiothreitol (DTT), AMV reverse transcriptase at 0.5 U/μl, T7 RNA polymerase at 5 U/μl; 37°C, 1 hour. (**B**) Procedure for selective amplification based on phosphoester transfer activity of a group I ribozyme. The 3' portion of the substrate, d(A$_3$(TA$_3$)$_3$), was transferred to the 3'-terminal guanosine of the ribozyme. Reaction conditions for RNA-catalyzed DNA cleavage: 1 μM *Tetrahymena* ribozyme (L-21 form), 10 μM d(GGCCCTCT-A$_3$(TA$_3$)$_3$, 10 mM MgCl$_2$, 30 mM *N*-[2-hydroxymethyl]-piperazine-*N'*-[3-propanesulfonic acid] (EPPS) (pH 7.5); 37°C, 1 hour. Selective amplification as in (A), with d((T$_3$A)$_3$T$_3$C) as primer 1 and d(ATCGATAATACGACTCACTATAGGAG-GGAAAAGTTATCAGGC) as primer 2. Subsequent selective cDNA synthesis with 0.2 pmol of the selective amplification product under

conditions as in (A), but omitting primer 2 and T7 RNA polymerase. Subsequent PCR amplification with 0.01 pmol of the selective cDNA synthesis product in a reaction mixture (100 μl volume) containing 0.2 μM d(CGAGTACTCCAAAACTAATC), 0.2 μM primer 2 (as above), 0.2 mM dNTP's, 50 mM KCl, 1.5 mM MgCl$_2$, 10 mM tris-HCl (pH 8.3), 0.01 percent gelatin, and 2.5 U of Taq DNA polymerase (Cetus), 30 cycles of 92°C for 1 minute, 45°C for 1 minute, 72°C for 1 minute. PCR products were purified by extraction with chloroform and isoamyl alcohol and by precipitation with ethanol, and used to transcribe RNA as described in the legend to Fig. 2.

Fig. 2. Secondary structure of the *Tetrahymena* ribozyme (L-21 form) showing those regions that were randomly mutagenized (boxed segments). Transcription conditions: 2 pmol of DNA template (containing mutagenic oligodeoxynucleotides), 2 mM NTP's, 15 mM MgCl$_2$, 2 mM spermidine, 5 mM DTT, 50 mM tris-HCl (pH 7.5), 1500 U of T7 RNA polymerase; 60 μl volume; 37°C, 2 hours. RNA was purified by electrophoresis in a 5 percent polyacrylamide–8 M urea gel and subsequent column chromatography on Sephadex G-50.
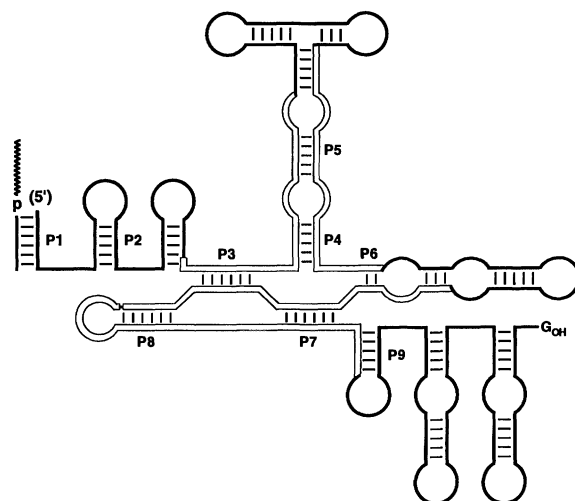
molecules) of material. Thus, the generation 0 population was expected to contain the wild-type ribozyme, all possible 1-, 2-, 3-, and 4-error mutants, and a sampling of the higher-error mutants (Table 1).

The evolution experiment spanned ten successive generations; each generation began with 20 pmol of RNA. The amount of RNA was quantified after selective amplification and after transcription (Fig. 3A). DNA cleavage activity for the population as a whole was monitored by a gel electrophoresis assay involving cleavage of (5'-^{32}P)-labeled d(GGCCCTCT-A$_3$(TA$_3$)$_3$) to yield d(GGCCCTCT) (Fig. 3B). It is expected that any given mutation would more likely be detrimental than beneficial, although there may be a substantial number of neutral mutations. Indeed, DNA cleavage activity for the generation 0 population is less efficient than for the wild type. The generation 1 population, having been selected for DNA cleavage activity under physiologic conditions, shows improved catalytic activity compared to generation 0 and is slightly improved over the wild type. Through successive generations there is continued improvement of phenotype. By

generation 7, the population as a whole cleaves DNA more efficiently at 37°C and 10 mM MgCl$_2$ than does the wild type at the high-temperature, high-MgCl$_2$ condition. Through generation 10 the rate of improvement has yet to level off.

RNA's from each generation were purified by polyacrylamide gel electrophoresis and Sephadex chromatography. To provide a more formal assay of DNA cleavage activity, we prepared d(GGCCCTCT-A$_3$(TA$_3$)$_3$[5'-^{32}P]A) substrate (33) and measured formation of both the ribozyme-d(A$_3$(TA$_3$)$_3$A) covalent intermediate and the RNA-catalyzed site-specific hydrolysis product d(A$_3$(TA$_3$)$_3$A) (Fig. 3C) (34). The hydrolysis product forms either by direct cleavage of the DNA substrate or by cleavage of the ribozyme-d(A$_3$(TA$_3$)$_3$A) covalent intermediate. Together, these reactions account for less than 5 percent of the cleaved substrate.

After ten generations, DNA cleavage activity for the population as a whole is 30 times higher than that of the wild type. Because selection is based on primer hybridization to the EP covalent intermediate (Fig. 1B), there is selection pressure against

the subsequent site-specific hydrolysis reaction. As a consequence, the efficiency of the hydrolysis reaction relative to the initial phosphoester transfer event drops from 4.9 percent for the wild type to 1.5 percent for the generation 10 population. There is selection pressure favoring accurate cleavage of the DNA at the target phosphodiester; inaccurate cleavage would result in partial mismatch of the primer used to initiate selective amplification. The accuracy of cleavage at first declines from 90 percent for the wild type to 45 percent for the generation 8 population and then rises to 60 percent for the generation 10 population. There are some individuals in the population that sacrifice accuracy for improved cleavage activity in order to enjoy an overall selective advantage (see below). Of course, the most favorable solution is an individual that has both high accuracy and high cleavage activity.

Evolutionary history. Although evolution in natural populations is an accomplished fact, evolution in vitro is a work in progress that allows the experimenter to access any time in evolutionary history. We obtained subclones from the evolving population at every generation (35). Generations 3, 6, and 9 were chosen for detailed analysis. DNA was prepared from 25 subclones at generations 3 and 6 and from 50 subclones at generation 9. The nucleotide sequence of the entire ribozyme gene was determined for each of these subclones (36) and shows how genotype changes over the course of evolutionary history (Fig. 4).

From generation 0 to generation 3, variation is discarded throughout much of the catalytic core of the ribozyme. The mean number of mutations per subclone decreased from 7.0 at generation 0 to 2.7 at generation 3. By generation 3, a small number of muta-

Fig. 3. Course of evolution over ten successive generations. (A) Changes in RNA population size over time: ●, after transcription (see legend to Fig. 2), quantitated by [^{3}H]uracil content; ○, at the start of each generation, based on 20-pmol portions; ■, after reaction with substrate, estimated by the assay shown in (C); □, after selective amplification (see legend to Fig. 1), quantitated by acid precipitation at 4°C of [α-^{32}P]GTP-labeled progeny RNA. (B) Cleavage of (5'-^{32}P)-labeled d(GGCCCTCT-A$_3$(TA$_3$)$_3$) substrate (S) in the absence of enzyme (−), in the presence of the wild-type Tetrahymena ribozyme (L-21 form), and in the presence of the population of RNA's obtained at each generation (Gn; n = 0,10). Reaction conditions: 0.5 μM ribozyme, 0.1 μM substrate (2.6 μCi/pmol), 30 mM EPPS (pH 7.5); either 10 mM MgCl$_2$, 37°C, 1 hour (low) or 50 mM MgCl$_2$, 2 mM spermidine, 50°C, 1 hour (high). Reaction products were separated by electrophoresis in a 20 percent polyacrylamide–8 M urea gel, an autoradiogram of which is shown. P, [5'-^{32}P]d(GGCCCTCT); P$_{+1}$, [5'-^{32}P]d(GGCCCTCTA). (C) Cleavage of [3'-^{32}P]dA-labeled d(GGCCCTCT-A$_3$(TA$_3$)$_3$[5'-^{32}P]A) under reaction conditions as in (B). S, EP, and P were separated by electrophoresis in a 20 percent polyacrylamide–8 M urea gel. Individual bands were cut from the gel and quantitated by Cerenkov counting. Data points are the average of five replicate experiments performed on three different days with two different preparations of substrate. Error bars correspond to ±1 SD.
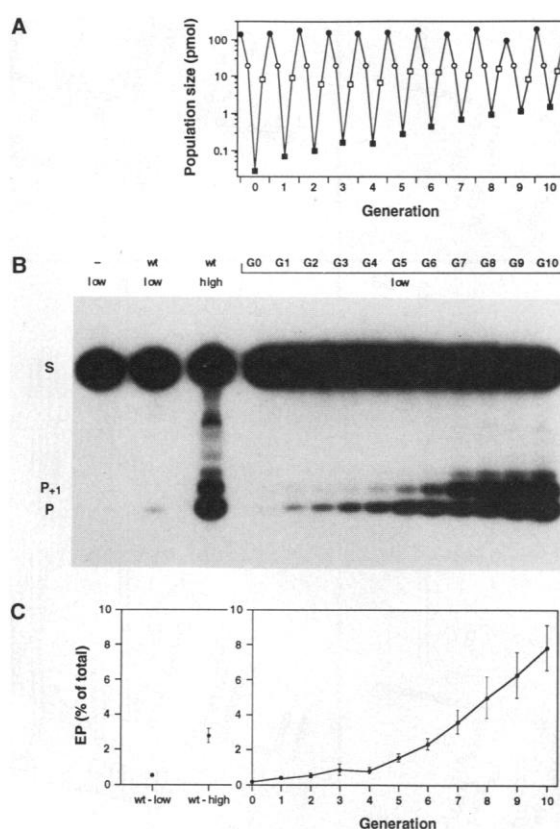
Table 1. Composition of the initial population (generation 0). The probability P of having k errors in a doped oligonucleotide of length v and degeneracy d is given by: $P(k,v,d) = [v!/((v-k)!k!)]d^k(1-d)^{v-k}$. We randomly mutagenized a total of 140 positions (v = 140) at a degeneracy of 5 percent per position (d = 0.05). The number of distinct k-error sequences of length v is given by: $N_k = [v!/((v-k)!k!)]3^k$. The expected number of copies per sequence is based on a population size of 20 pmol (1.2 × 10^{13} molecules).

Errors	Probability (%)	Sequences	Copies/sequence
0 (wt)	0.1	1	9 × 10^9
1	0.6	420	2 × 10^8
2	2.1	9 × 10^4	3 × 10^6
3	5.0	1 × 10^7	5 × 10^4
4	9.0	1 × 10^9	9 × 10^2
5	12.8	1 × 10^{11}	15
6	15.2	7 × 10^{12}	0.3
7+	55.4		

tions outside of the original zone of random mutation have occurred because of ongoing mutation events (Fig. 4B). The consensus sequence is still that of the wild type, although only one of 25 subclones has the entire wild-type sequence.

From generation 3 to generation 6 the dramatic accumulation of mutations at five positions within the ribozyme coincides with a threefold improvement in the phenotype of the population as a whole. From generation 6 to generation 9, these positions are further accentuated and aggregate phenotype improves another threefold. The mean number of mutations per subclone

rises to 4.6 at generation 6 and to 5.9 at generation 9 as a larger proportion of subclones adopt the common mutations and as mutations accumulate outside of the original zone of random mutation.

The most frequent mutation is an A→Y (Y = U or C) change at position 94 (94:A→Y). This mutation is present, as A→U, in only 1 of 25 subclones at generation 3. At generation 6 there are 15 of 25 occurrences, 12 as A→U and 3 as A→C; at generation 9 there are 35 of 50 occurrences, 22 as A→U and 13 as A→C. Position 94 is unpaired in the secondary structure of the wild-type ribozyme (27). Considering the

effect of site-directed mutations made at neighboring positions (37), the 94:A→Y change may alter the orientation of ribozyme-bound substrate relative to the catalytic core of the molecule.

Another frequent mutation, occurring in 4 of 25 subclones at generation 3, 6 of 25 subclones at generation 6, and 22 of 50 subclones at generation 9, is a G→A change at position 215. This mutation converts a G·U wobble pair to an A·U Watson-Crick pair within the catalytic core of the ribozyme. Among 87 group I intron sequences that have been analyzed, 39 have a G·U and 28 have a G·C, but only 4 have
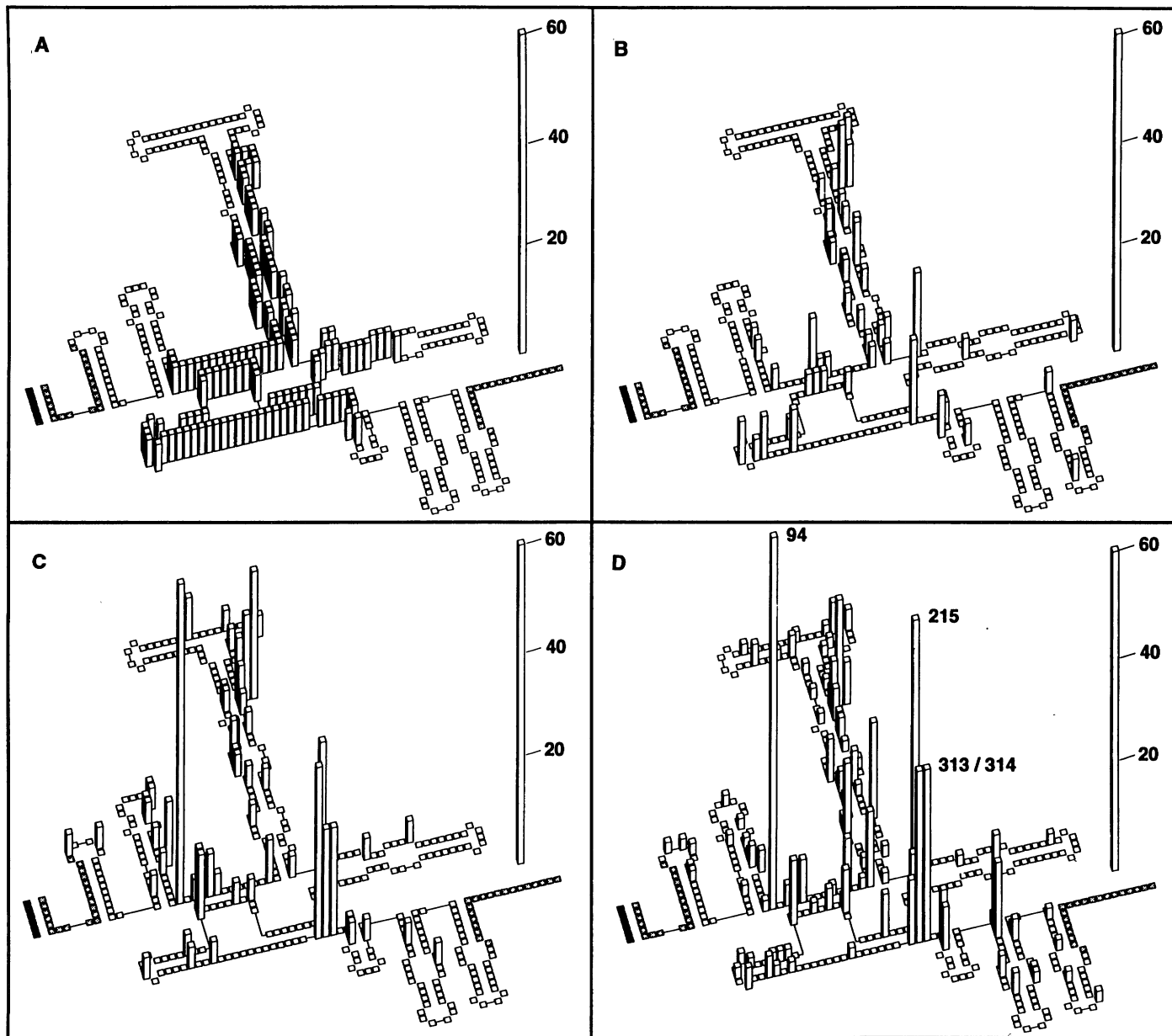
Fig. 4. Sites at which mutations occurred over the course of evolution, superimposed on the secondary structure of the *Tetrahymena* ribozyme. Box height corresponds to frequency of mutation (percent) at each nucleotide position. Nonmutable primer binding sites are shaded; substrate is shown in black. (**A**) Generation 0, based on expected composition of the initial population (Table 1); (**B**) generation 3, 25-subclone resolution; (**C**) generation 6, 25-subclone resolution; (**D**) generation 9, 50-subclone resolution. Labeled positions in (D) are sites of frequent mutation, which were prepared individually by site-directed mutagenesis.

an A·U at this location (29).

The most remarkable mutations are a G→U change at position 313 and an A→G change at position 314 that always occur together. These mutations are absent at generation 3, but are present in 5 of 25 subclones at generation 6 and 16 of 50 subclones at generation 9. The GA sequence normally present at positions 313-314 is thought to form a short duplex structure (the 5' half of the P9.0 pairing) that draws the 3'-terminal guanosine residue of the ribozyme into the catalytic core (38). We utilize the 3'-terminal guanosine as the nucleophile in the target phosphoester transfer reaction. The 313-314 mutations are expected to destroy the P9.0 pairing, yet confer selective advantage with respect to DNA cleavage (see below).

There is a frequent G → A change at position 312 that occurs only if the 313-314 mutations are not present. The 312:G → A change is present in 4 of 25 subclones at generation 3 and 8 of 25 subclones at generation 6, but only 5 of 50 subclones at generation 9. In terms of population frequency, the 312:G → A mutation declines as the 313-314:GA → UG mutations become more abundant.

Activity of evolved individuals. DNA from 14 subclones at generation 9 was transcribed to produce individual RNA's, which were purified by polyacrylamide gel electrophoresis and Sephadex chromatography. The catalytic behavior of these RNA's was studied with $(5'-^{32}P)$- and $(3'-^{32}P)$-labeled DNA substrates and $(5'-^{32}P)$- and $[\alpha-^{32}P]$-ATP-labeled RNA substrates having the sequence $GGCCC_U^TC_U^TC-A_3(_U^TA_3)_3$. The kinetic parameter most relevant to our selec-

tion criterion is the proportion of ribozyme molecules that become joined to the 3' portion of the DNA substrate after 1 hour at 37°C and 10 mM $MgCl_2$. These data and comparable data concerning reactions with a DNA substrate at 50°C and 50 mM $MgCl_2$ and with an RNA substrate at 37°C and 10 mM $MgCl_2$ are presented in Fig. 5.

There is considerable heterogeneity among the 14 individual RNA's with respect to DNA cleavage activity in the target reaction. All are more active than the wild type, with the best (clones 29 and 23) being about 60 times more active. The five most active individuals are more active under physiologic conditions than at the high-temperature and high-$MgCl_2$ condition. All 14 individuals show improved activity with the RNA substrate, even though the population has never been challenged with RNA. Improved RNA cleavage activity is largely due to enhanced activity in the site-specific hydrolysis reaction ($r = +0.93$), which allows enhanced turnover.

As mentioned previously, there is selection pressure against site-specific hydrolysis of the EP covalent intermediate in the reaction with a DNA substrate. In fact, all but one of the 14 individuals show decreased hydrolytic cleavage of the attached DNA compared to the wild type. All but one of the individuals show increased hydrolytic cleavage with the RNA substrate. Furthermore, there is a strong negative correlation ($r = -0.93$) between hydrolytic cleavage of DNA and RNA. The population is clearly divided into two groups: those with low DNA and high RNA hydrolysis activity and those with high DNA and low RNA hydrolysis activity (Fig. 5). All nine members of the former group carry

the 313-314:GA→UG mutations, while all five members of the latter group lack these changes.

We chose clones 29 and 23 for more detailed kinetic analysis, in comparison with the wild-type ribozyme. Initial rates were determined for the reaction with $(5'-^{32}P)$-labeled $d(GGCCCTCT-A_3(TA_3)_3)$ substrate at 37°C and 10 mM $MgCl_2$, with 1 μM ribozyme and excess substrate. An Eadie-Hofstee plot of v_0 as a function of $v_0/[S]$ was used to obtain V_{max} and K_m (Fig. 6). From this data, k_{cat} and k_{cat}/K_m were calculated. For the wild-type ribozyme, $K_m = 6.6$ μM and $k_{cat} = 0.0002$ min^{-1} ($k_{cat}/K_m = 36$ M^{-1} min^{-1}). This compares to $K_m = 30$ μM and $k_{cat} = 0.006$ min^{-1}, previously reported for the wild-type ribozyme in a related reaction at 50°C and 10 mM $MgCl_2$ (12). For clone 29, $K_m = 2.0$ μM and $k_{cat} = 0.007$ min^{-1} ($k_{cat}/K_m = 3600$ M^{-1} min^{-1}); for clone 23, $K_m = 1.9$ μM and $k_{cat} = 0.005$ min^{-1} ($k_{cat}/K_m = 2700$ M^{-1} min^{-1}) (data obtained at 37°C and 10 mM $MgCl_2$). Thus, the catalytic efficiency of the two evolved RNA's was increased and was about 100 times greater

Fig. 5. Catalytic activity of 14 individual ribozymes obtained at generation 9. Ribozymes were transcribed as described in the legend to Fig. 2 and assayed as described in the legend to Fig. 3. ●, catalytic behavior of individuals; - - -, average catalytic behavior of the 14 individuals (corresponds to catalytic behavior of the generation 9 population as a whole, $P > 0.99$); —, catalytic behavior of the wild type. Numbers in parentheses are correlation coefficients (r) relating activity with DNA substrate at "low" to activity with either DNA substrate at "high" or RNA substrate at "low." RNA substrate was prepared by in vitro transcription with a synthetic oligodeoxynucleotide template (41); reaction conditions were as described in the legend to Fig. 2, but include $[\alpha-^{32}P]$ATP at 0.003 μCi/pmol to label the 3' portion of the substrate.
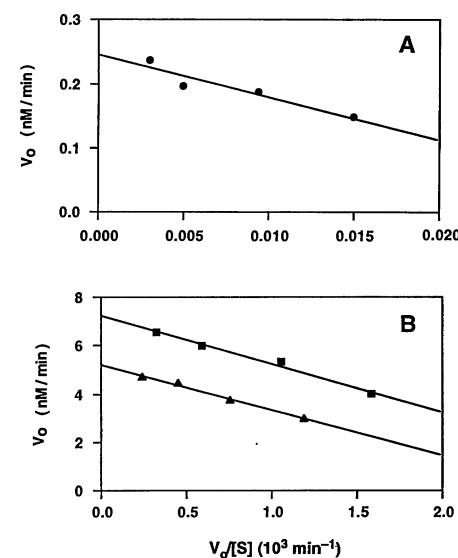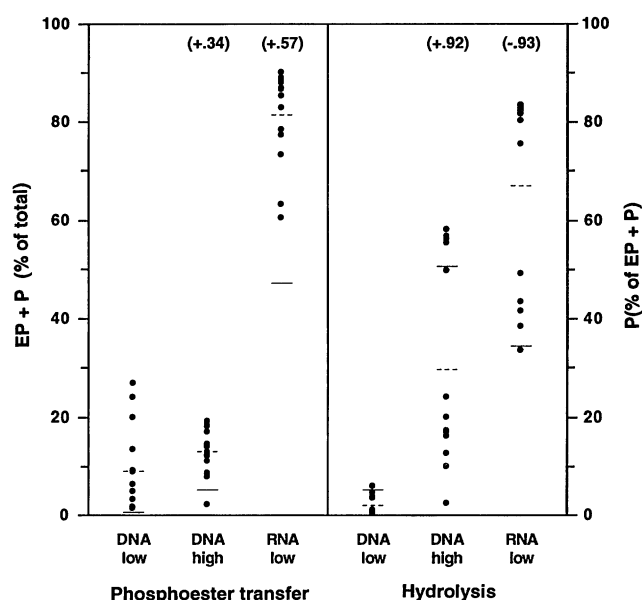
Fig. 6. Eadie-Hofstee plots (42) used to determine K_m (negative slope) and V_{max} (y-intercept) for cleavage of $(5'-^{32}P)$-labeled $d(GGCCCTCT-A_3(TA_3)_3)$ by (**A**) wild type; (**B**) clones 29 and 23 from generation 9. Reaction conditions: 1 μM ribozyme, 10 mM $MgCl_2$, 30 mM EPPS (pH 7.5), 37°C; for wild type, 10, 20, 40, and 80 μM substrate at 0.25, 30, 60, 120, 180, and 240 minutes; for clone 29, 2.5, 5, 10, and 20 μM substrate at 0.25, 5, 10, and 15 minutes; for clone 23, 2.5, 5, 10, and 20 μM substrate at 0.25, 5, 10, 20, and 30 minutes. Ribozyme and substrate were first incubated separately in 10 mM $MgCl_2$, 30 mM EPPS (pH 7.5), 37°C, for 15 minutes, then mixed to start the reaction. ●, wild type; ■, clone 29; ▲, clone 23; each data point is the average of three independent determinations of initial velocity. The extent of the reaction was linear over the chosen time interval ($r_{min} = 0.94$, $r_{avg} = 0.99$).

Table 2. Genotype and phenotype of 14 individuals from generation 9. Genotype is represented as a binary matrix (shown in brackets). Phenotype is represented as a column vector b_i, with values normalized to wild type = 1.0. DNA cleavage and hydrolysis activity were determined with ($3'$-^{32}P)-labeled DNA substrate under physiologic conditions, as described in the legend to Fig. 3C. Accuracy was determined with ($5'$-^{32}P)-labeled DNA substrate under physiologic conditions, measuring the fraction of substrate cleavage that occurs at the target phosphodiester bond; reaction conditions were as described in the legend to Fig. 3B. Avg., the average of the 14 individuals. G9 genotype is the average of the 50 subclones obtained from the 9th generation; G9 phenotype is the behavior of the G9 population as a whole.

Clone	Errors (N)	94: A→Y	98: C→U	205: U→C	215: G→A	313-314 GA→UG	317: U→R	333: U→C	DNA cleavage b_1	Hydrolysis b_2	Accuracy b_3
29	6	1	0	0	1	1	0	0	65	0.1	0.7
23	7	1	0	0	1	1	0	0	57	0.1	0.6
30	8	1	0	0	1	1	1	0	48	0.1	0.8
43	8	1	0	0	1	1	0	1	32	0.0	0.4
5	7	1	0	0	0	1	0	0	22	0.1	0.4
37	4	0	0	0	1	1	0	0	21	0.1	0.6
28	5	0	1	0	0	1	0	0	15	0.0	0.7
2	5	0	0	0	0	1	0	0	11	0.0	0.8
42	6	1	0	0	1	0	0	0	7	0.7	0.8
8	8	1	0	1	0	1	0	0	3	0.2	1.1
40	1	0	0	0	1	0	0	0	3	0.9	0.6
12	6	0	1	1	0	0	1	0	3	0.7	0.8
27	2	1	0	0	0	0	0	0	3	0.8	0.6
11	6	1	1	0	0	0	0	1	3	1.2	0.8
wt	0	0	0	0	0	0	0	0	1	1.0	1.0
x_1:		10	2	-18	13	18	13	-9			
x_2:		0.4	0.7	0.2	0.4	-0.4	-0.2	-0.1			
x_3:		0.3	0.6	0.4	0.3	0.2	-0.1	-0.3			
avg.	5.6	0.6	0.2	0.1	0.5	0.6	0.1	0.1	21	0.3	0.7
G9	5.9	0.7	0.1	0.2	0.4	0.3	0.1	0.2	21	0.3	0.6

Fig. 7. DNA cleavage activity of individuals obtained by site-directed mutagenesis. Reaction conditions as described in the legend to Fig. 3C. —, no enzyme; G9, generation 9 population as a whole. Reaction products were separated in a 20 percent polyacrylamide–8 M urea gel, an autoradiogram of which is shown.

than that of the wild type, because of improvement in both K_m and k_{cat}.

Correlating genotype and phenotype. The relation between genotype and phenotype in the context of an RNA-based evolving system can now be formalized. Genotype can be represented as a matrix **A**, the rows corresponding to individuals in the population and the columns corresponding to functionally significant positions within the nucleotide sequence (Table 2). Phenotype can be represented as a column vector **b**, whose entries are some measure of fitness (catalytic behavior) of the various individuals. We then seek a row vector **x** that provides a best fit to the equation: **Ax** = **b**, that is, provides a best fit linear estimation of the relation between genotype and phenotype. The solution that minimizes the least-squares error is: $x = (A^{*}A)^{-1} A^{*}b$, where A^{*} is the transpose of **A**. In this way, we obtain a weighting vector **x** that provides an estimate of phenotype for any given genotype (Table 2).

The data obtained from 14 individuals is not sufficient to provide a meaningful solution to the relation of genotype to phenotype, even for those nucleotide positions that are known to be most significant based on their high frequency of accepted mutation. We use

the weighting vector **x** as a guide to help decide which mutations are sufficiently important to warrant individual study.

The following individual mutations were prepared by site-directed mutagenesis: 94:A→U, 94:A→C, 215:G→A, 313:G→U, 314:A→G, and 313-314:GA→UG (39). Catalytic activity was studied with d(GG-CCCTCT-A$_3$(TA$_3$)$_3$[$5'$-^{32}P]A) substrate. The individual mutations result in improved activity compared to the wild type, but they do not result in activity exceeding that of the generation 9 population as a whole (Fig. 7). Activity in the 94:A→U mutant is seven times greater and in the 94:A→C mutant it is two times greater than in the wild type. The 313-314:GA→UG double mutant is more active than either the 313:G→U or 314:A→G single mutant, explaining why the 313-314 mutations occur together among the evolved individuals that we have studied. As predicted from the analysis of 14 individuals at generation 9, the 313-314:GA→UG mutations result in diminished site-specific hydrolysis of the DNA substrate compared to the wild type. These mutations confer both enhanced phosphoester transfer activity and diminished site-specific hydrolysis activity, and thus are well
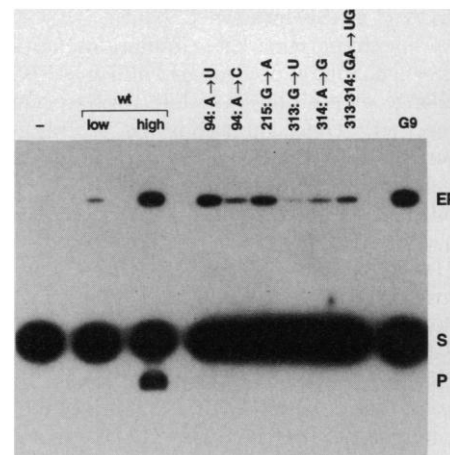
suited to meet the imposed selection constraint which depends on availability of the EP covalent intermediate.

The evolutionary frontier. As an in vitro model of Darwinian evolution, a population of macromolecular catalysts was directed toward the expression of novel catalytic function. In our study, we wanted to develop ribozymes that cleave DNA with improved efficiency under physiologic conditions. In a related study, we used these evolved RNA's to cleave a target DNA in vivo; ribozymes obtained from generation 9 were expressed in E. coli and shown to prevent infection by M13 single-stranded DNA bacteriophage (40).

We are continuing the present successful phylogeny beyond the tenth generation, but only after decreasing the concentration of DNA substrate in the target reaction. Through the first ten generations the substrate concentration was 10 μM, roughly matching the K_m for the wild type. Now that the evolved individuals have attained a K_m of about 2 μM, the substrate concentration must be reduced to subsaturating levels to promote further improvement in substrate binding. In addition, we are attempting to improve catalytic turnover in the DNA cleavage reaction by selecting for both phosphoester transfer activity, which generates the EP covalent intermediate, and subsequent RNA-catalyzed site-specific hydrolysis activity, which frees the ribozyme to act on another substrate molecule.

The selection scheme that we used could be applied to various substrates of the form: d(CCCTCΩ-A$_3$(TA$_3$)$_3$), where Ω refers to some nucleotide analogue and the ribozyme is selected for its ability to cleave the phosphodiester bond following the se-

quence CCCTCΩ. The substrate need not be a nucleotide or nucleotide analogue. The only requirement is that RNA's that react with the substrate become tagged in some way so that they can be distinguished from nonreactive molecules with respect to the amplification process. For example, reactive RNA's could become joined to a portion of the substrate that is attached to a solid support. Nonreactive RNA's would be washed away, leaving the bound RNA's to be selectively amplified.

The step from an RNA-cleaving ribozyme to a DNA-cleaving ribozyme is a modest one. It seems reasonable that RNA is capable of catalyzing a broader range of phosphoester transfer reactions, but it is not clear which of these activities are accessible from existing ribozymes via directed evolution. In some cases, it may be necessary to evolve a succession of ribozymes that lead progressively toward the desired catalytic behavior. An important milestone will be the evolution of a ribozyme that performs novel chemistry, that is, catalyzes some reaction other than phosphoester cleavage or ligation. Considering the functional groups that exist within RNA, there are a number of plausible avenues to be explored.

REFERENCES AND NOTES

1. J. R. Knowles, Science 236, 1252 (1987); K. A. Johnson and S. J. Benkovic, in The Enzymes, D. S. Sigman and P. D. Boyer, Eds. (Academic Press, San Diego, CA, 1990), vol. 19, pp. 159–211.
2. D. R. Mills, R. L. Peterson, S. Spiegelman, Proc. Natl. Acad. Sci. U.S.A. 58, 217 (1967); R. Green, A. D. Ellington, J. W. Szostak, Nature 347, 406 (1990); B. M. Chowrira, A. Berzal-Herranz, J. M. Burke, ibid. 354, 320 (1991).
3. G. F. Joyce, Gene 82, 83 (1989).
4. D. L. Robertson and G. F. Joyce, Nature 344, 467 (1990).
5. C. Tuerk and L. Gold, Science 249, 505 (1990); A. D. Ellington and J. W. Szostak, Nature 346, 818 (1990).
6. K. S. Lam et al., Nature 354, 82 (1991).
7. J. K. Scott and G. P. Smith, Science 249, 386 (1990); J. J. Devlin, L. C. Panganiban, P. E. Devlin, ibid., p. 249; S. E. Cwirla, E. A. Peters, R. W. Barrett, W. J. Dower, Proc. Natl. Acad. Sci. U.S.A. 87, 6378 (1990).
8. M. Eigen, J. McCaskill, P. Schuster, J. Phys. Chem. 92, 6881 (1988).
9. K. Kruger et al., Cell 31, 147 (1982).
10. A. J. Zaug and T. R. Cech, Science 231, 470 (1986); P. S. Kay and T. Inoue, Nature 327, 343 (1987).
11. A. J. Zaug, M. D. Been, T. R. Cech, Nature 324, 429 (1986).
12. D. Herschlag and T. R. Cech, ibid. 344, 405 (1990).
13. C.-T. Chu, D. S. Parris, R. A. F. Dixon, F. E. Farber, P. A. Schaffer, Virology 98, 168 (1979); D. Shortle and D. Botstein, Methods Enzymol. 100, 457 (1983); R. M. Myers, L. S. Lerman, T. Maniatis, Science 229, 242 (1985).
14. M. D. Matteucci and H. L. Heyneker, Nucleic Acids Res. 11, 3113 (1983); J. A. Wells, M. Vasser, D. B. Powers, Gene 34, 315 (1985).
15. J. B. McNeil and M. Smith, Mol. Cell. Biol. 5, 3545 (1985); C. A. Hutchison, S. K. Nordeen, K. Vogt, M. H. Edgell, Proc. Natl. Acad. Sci. U.S.A. 83, 710 (1986); K. M. Derbyshire, J. J. Salvo, N. D. F. Grindley, Gene 46, 145 (1986).

16. R. A. Zakour and L. A. Loeb, Nature 295, 708 (1982); P. M. Lehtovaara, A. K. Koivula, J. Bamford, J. K. C. Knowles, Protein Eng. 2, 63 (1988).
17. D. W. Leung, E. Chen, D. V. Goeddel, Technique 1, 11 (1989); Y. Zhou, X. Zhang, R. H. Ebright, Nucleic Acids Res. 19, 6052 (1991).
18. R. K. Saiki et al., Science 230, 1350 (1985); R. K. Saiki et al., ibid. 239, 487 (1988).
19. D. Y. Kwoh et al., Proc. Natl. Acad. Sci. U.S.A. 86, 1173 (1989); G. F. Joyce, in Molecular Biology of RNA: UCLA Symposia on Molecular and Cellular Biology, T. R. Cech, Ed. (Liss, New York, 1989), pp. 361–371.
20. M. Chamberlin and T. Ryan, in The Enzymes, P. D. Boyer, Ed. (Academic Press, New York, 1982), vol. 15, pp. 87–108.
21. J. C. Guatelli et al., Proc. Natl. Acad. Sci. U.S.A. 87, 1874 (1990); G. F. Joyce, in Antisense RNA and DNA, J. A. H. Murray, Ed. (Wiley-Liss, New York, 1992), pp. 353–372.
22. In our study, the PCR was performed under our standard reaction conditions (see legend to Fig. 1), resulting in an error rate of approximately 0.1 percent per position per generation. We have developed a mutagenic PCR procedure that provides an error rate of 0.66 ± 0.13 percent per position (95 percent confidence level) (R. C. Cadwell and G. F. Joyce, PCR Methods Appl. 2, 28 (1992).
23. In practice, there is always the danger of developing a "parasite" that circumvents the selection criterion and is amplified more efficiently than the most reactive species (21). In our study, for example, a sequence arose that allowed false hybridization of one of the amplification primers at an internal site, generating a species with a 53-nucleotide deletion that was amplified more efficiently than the full-length ribozyme.
24. M. D. Been and T. R. Cech, Science 239, 1412 (1988); S. A. Woodson and T. R. Cech, Cell 57, 335 (1989); J. A. Doudna and J. W. Szostak, Nature 339, 519 (1989).
25. F. L. Murphy and T. R. Cech, Proc. Natl. Acad. Sci. U.S.A. 86, 9218 (1989).
26. D. Herschlag and T. R. Cech, Biochemistry 29, 10159 (1990).
27. J. M. Burke et al., Nucleic Acids Res. 15, 7217 (1987).
28. S.-H. Kim and T. R. Cech, Proc. Natl. Acad. Sci. U.S.A. 84, 8788 (1987); D. W. Celander and T. R. Cech, Science 251, 401 (1991).
29. F. Michel and E. Westhof, J. Mol. Biol. 216, 585 (1990).
30. R. W. Davies, R. B. Waring, J. A. Ray, T. A. Brown, C. Scazzocchio, Nature 300, 719 (1982); F. Michel, A. Jacquier, B. Dujon, Biochimie 64, 867 (1982); F. Michel and B. Dujon, EMBO J. 2, 33 (1983); T. R. Cech, Gene 73, 259 (1988).
31. J. V. Price, G. L. Kieft, J. R. Kent, E. L. Sievers, T. R. Cech, Nucleic Acids Res. 13, 1871 (1985); J. W. Szostak, Nature 322, 83 (1986); G. F. Joyce and T. Inoue, Nucleic Acids Res. 15, 9825 (1987); E. T. Barfod and T. R. Cech, Genes Dev. 2, 652 (1988); G. F. Joyce, G. van der Horst, T. Inoue, Nucleic Acids Res. 17, 7879 (1989); S. Couture et al., J. Mol. Biol. 215, 345 (1990); A. A. Beaudry and G. F. Joyce, Biochemistry 29, 6534 (1990).
32. G. F. Joyce and T. Inoue, Nucleic Acids Res. 17, 711 (1989).
33. The (3′-32P)-labeled DNA substrate was prepared with terminal deoxynucleotide transferase. Reaction conditions: 4 μM d(GGCCCTCT-A₃(TA₃)₃), 1 μM [α-32P]dATP (3 μCi/pmol), 1 mM CoCl₂, 1 mM DTT, 50 mM potassium cacodylate (pH 7.2), and terminal transferase (BRL) at 2.7 U/μl, incubated at 37°C for 30 minutes. The product corresponding to addition of a single dA residue was purified by electrophoresis in a 20 percent polyacrylamide–8 M urea gel and subsequent affinity chromatography on Nensorb (DuPont).
34. T. Inoue, F. X. Sullivan, T. R. Cech, J. Mol. Biol. 189, 143 (1986).
35. DNA's used to transcribe the population of RNA's at each generation (legend to Fig. 1) were amplified in a second PCR reaction with primers 5′-CCAAGCTTGATCTCGAGTACTCCAAAACTAA-

TC-3′ and 5′-CTGCAGAATTCTAATACGACTCA-CTATAGGAGGGAAAAGTTATCAGGC-3′, producing a 435-bp (base pair) fragment with unique restriction sites at its ends. The fragment was digested with Eco RI and Hind III and ligated into a pUC18 vector that had been linearized with Eco RI and Hind III and purified in a 1.5 percent agarose gel. The resulting plasmid DNA's were used to transform competent DH5α-F′ cells [D. Hanahan, in DNA Cloning: A Practical Approach, D. M. Glover, Ed. (IRL Press, Oxford, 1985), pp. 109–135], which were then grown on ampicillin-containing plates. Individual colonies were chosen at random and grown overnight in liquid media. DNA was prepared by the boiling, lysis method [D. S. Holmes and M. Quigley, Anal. Biochem. 114, 193 (1981)] and screened for the insert by restriction digestion.
36. Cloned individuals were sequenced by dideoxy chain-termination [F. Sanger, S. Nicklen, A. R. Coulson, Proc. Natl. Acad. Sci. U.S.A. 74, 5463 (1977); R. Zagursky, K. Baumeister, K. N. Lomax, M. L. Berman, Gene Anal. Tech. 2, 89 (1985)] with reciprocal primers 5′-GTAAAACGACGGCCAGT-3′ and 5′-CATGATTACGAATTCTA-3′, which are compatible with the pUC plasmid. Sequencing reactions utilized modified T7 DNA polymerase (Sequenase, USB) and [35S](α-thiosphate) dATP and were analyzed by electrophoresis in a 6 percent polyacrylamide–8 M urea gel. Nucleotide sequences of individual subclones are available on request.
37. B. Young, D. Herschlag, T. R. Cech, Cell 67, 1007 (1991).
38. F. Michel, M. Hanna, R. Green, D. P. Bartel, J. W. Szostak, Nature 342, 391 (1989); F. Michel, P. Netter, M.-Q. Xu, D. A. Shub, Genes Dev. 4, 777 (1990).
39. Site-directed mutagenesis was carried out as described [Y. Morinaga, T. Franceschini, S. Inouye, M. Inouye, Biotechnology 2, 636 (1984)]. Plasmid pT7L-21 [A. J. Zaug, C. A. Grosshans, T. R. Cech, Biochemistry 27, 8924 (1988)] was digested with either (i) Eco RI and Hind III to remove the ribozyme coding region, or (ii) Bsa I and Xmn I to remove the ampicillin-resistance gene. The resulting fragments were purified in a 1 percent agarose gel and cross-hybridized in the presence of a 5′-phosphorylated synthetic oligodeoxynucleotide that introduces the desired mutation. The annealing mixture contained 0.06 pmol of pT7L-21(ΔEco-Hind), 0.06 pmol pT7L-21(ΔBsa-Xmn), 15 pmol of mutagenic oligodeoxynucleotide, 40 mM tris-HCl (pH 7.2), and 8 mM MgSO₄ in 12-μl volume, which was heated to 100°C for 3 minutes, then incubated at 30°C for 30 minutes, 25°C for 30 minutes, 4°C for 30 minutes, and 0°C for 10 minutes. The annealing product was made fully double-stranded with the Klenow fragment of E. coli DNA polymerase I (Boehringer) and T4 DNA ligase (USB) and then used to transform competent DH5α-F′ cells, which were grown on ampicillin-containing plates. Colonies were screened by the colony hybridization method with (5′-32P)-labeled mutagenic oligodeoxynucleotide as a probe [M. Grunstein and D. S. Hogness, Proc. Natl. Acad. Sci. U.S.A. 72, 3961 (1975)]. DNA was prepared from positive colonies (35) and sequenced throughout the ribozyme gene (36). RNA was prepared by in vitro transcription as described in the legend to Fig. 2.
40. D. L. Robertson and G. F. Joyce, unpublished results.
41. J. F. Milligan, D. R. Groebe, G. W. Witherell, O. C. Uhlenbeck, Nucleic Acids Res. 15, 8783 (1987).
42. G. S. Eadie, J. Biol. Chem. 146, 85 (1942); B. H. J. Hofstee, ibid. 199, 357 (1952). Because of limited product yield, we were forced to work at a substrate concentration [S] ≥ K_m; thus K_m should be regarded as only an estimate of the true value.
43. We thank D. Decker and D. Rumschitzki for advice and N. Lehman and L. Orgel for comments on the manuscript. Supported by NASA grant NAGW-1671 and NIH grant AI30882.