Non-unity molecular heritability demonstrated by continuous evolution *in vitro*

Thomas Schmitt[1,2] and Niles Lehman[1]

Introduction: When catalytic RNA is evolved *in vitro*, the molecule's chemical reactivity is usually the desired selection target. Sometimes the phenotype of a particular RNA molecule cannot be unambiguously determined from its genotype, however. This can occur if a nucleotide sequence can adopt multiple folded states, an example of non-unity heritability (i.e. one genotype gives rise to more than one phenotype). In these cases, more rounds of selection are required to achieve a phenotypic shift. We tested the influence of non-unity heritability at the molecular level by selecting for variants of a ligase ribozyme via continuous evolution.

Results: During 20 bursts of continuous evolution of a 152-nucleotide ligase ribozyme in which the Mg^{2+} concentration was periodically lowered, a nine-error variant of the starting 'wild-type' molecule became dominant in the last eight bursts. This variant appears to be more active than the wild type. Kinetic analyses of the mutant suggest that it may not possess a higher first-order catalytic rate constant, however. Examination of the multiple RNA conformations present under the continuous evolution conditions suggests that the mutant is superior to the wild type because it is less likely to misfold into inactive conformers.

Conclusions: The evolution of genotypes that are more likely to exhibit a particular phenotype is an epiphenomenon usually ascribed only to complex living systems. We show that this can occur at the molecular level, demonstrating that *in vitro* systems may have more life-like characteristics than previously thought, and providing additional support for an RNA world.

Addresses: [1]Department of Biological Sciences, University at Albany, State University of New York, 1400 Washington Avenue, Albany, NY 12222, USA. [2]Theodor-Boveri-Institut für Biowissenschaften der Universität, Zoologie III, Am Hubland, D-97074 Würzburg, Germany.

Correspondence: Niles Lehman
E-mail: niles@cnsunix.albany.edu

Key words: folding, heritability, *in vitro* selection, ribozymes, RNA

Received: 12 July 1999
Revisions requested: 10 August 1999
Revisions received: 13 September 1999
Accepted: 21 September 1999

Published: 5 November 1999

Chemistry & Biology December 1999, 6:857–869

1074-5521/99/$ – see front matter

Introduction

A chemical system that is increasingly being characterized as biological is that of catalytic nucleic acids — RNA and DNA enzymes. These molecules, discovered both in natural biological systems and through the efforts of *in vitro* evolution in the laboratory, lend tremendous credence to scenarios of life based entirely on nucleic acids, such as the 'RNA world' [1,2]. These molecules have a distinct genotype, their primary nucleotide sequence, which directs the formation of the 'phenotype', a catalytically active, folded, secondary/tertiary structure. An implicit assumption often made [3] is that a single such folded structure exists for each nucleic-acid enzyme, such that there is a one-to-one correspondence between genotype and phenotype. Yet it has long been appreciated that even under physiological environmental conditions, nucleic acids can adopt multiple conformations, and these alternate conformations may have different catalytic activities. Walstrum and Uhlenbeck [4] demonstrated that gel purification of the *Tetrahymena* self-splicing RNA following *in vitro* transcription can trap this ribozyme into a less active conformation. A more reactive conformation can be regenerated by a temperature-dependent denaturation/renaturation protocol. Similarly, the hepatitis delta virus has been shown to adopt alternate folding conformations that affect its self-cleavage activity [5–7]. Burke and Willis [8] postulated that misfolding prevented recombined RNA motifs from being fully active. In these cases, as well as in several others, the specific environmental conditions under which the RNAs are allowed to fold are important determinants of their phenotype. Moreover, it is likely that many folded states exist whose equilibrium distribution is strongly influenced by the environment [3,5,9,10].

In biology, heritability is a measure of the fraction of phenotypic variation for a trait in a population that is attributable to genotypic variation. When a trait is surveyed in a population, there are several ways to estimate both broad-sense heritability, where total genetic variation is compared with total phenotypic variation, and narrow-sense heritability, where additive (nonepistatic) genetic variation is compared with total phenotypic variation [11]. And in biological systems, one usually assumes that an organism's phenotype is an emergent property. That is, higher-level phenomena that could not necessarily be predicted from patterns gleaned from lower-level phenomena are thought to be a large determinant of the ultimate phenotype. In particular, the environment intervenes in the

phenotypic manifestation of the genotype, such that, in a population, a range of phenotypes can exist even when genotypic variation is minimal. When broad-sense heritability (H^2) is less than 1 (i.e., non-unity H^2), the final phenotype cannot be specified simply by knowing the genotype of an organism. Heritability has an important evolutionary consequence because traits with low heritabilities respond to natural selection more slowly than do traits with high heritabilities (see, for example [11]). This relationship led Fisher (in [12]) to propose the 'fundamental theorem of natural selection', which essentially conveys the observation that the rate of adaptive evolution is proportional to the amount of heritable genetic variation that exists in a population at that time.

In this report, we explore the evolutionary consequences of phenotypic heterogeneity at the level of single molecules that arises not only from genotypic heterogeneity. During the *in vitro* evolution of nucleic acids, a population of molecules is subjected to defined selection pressures, and presumably those sequences with a greater ability to conform to the selection will be preferentially amplified (reviewed in [13–15]). The issue of non-unity heritability in such experiments has not been addressed, however. This is in part because synthetically prepared RNAs often can be trapped in 'alternate conformation hell' [5], such that researchers often take steps to ensure that only a single conformer is likely to undergo selection. Furthermore, in most *in vitro* selection protocols, purification steps are alternated with selection and amplification steps such that there is an opportunity to homogenize conformationally each RNA population prior to subsequent selection.

Motivated by the observation that a single-sequence, gel-purified ligase ribozyme, designated E_{100}(#3) and selected in the course of a continuous evolution protocol [16], exhibits extremely fast reaction kinetics but does not appear to react to completion, we set out to examine the evolutionary fate of such a sequence upon further selection. Continuous evolution differs from typical 'stepwise' evolution in that all phases of the *in vitro* selection process — catalysis, reverse transcription and forward transcription — occur concomitantly in the same reaction vessel [16,17]. No purification steps are added between these steps, and all reactions must proceed in the same chemical milieu. Because the environment is constantly changing during the course of this process, as a consequence of the depletion of reactants, the accumulation of cDNA and RNA products, and the exhaustion of protein enzymes, periodically (e.g. every 30 min) an aliquot must be removed from the reaction vessel and used to seed a new vessel containing fresh reactants. Each such serial transfer results in a pulse of RNA production, referred to here as a 'burst'. We selected for variants of the E_{100}(#3) sequence under increasingly stringent environments with regard to divalent metal-ion availability, to test the hypothesis that the tendency to adopt (or not to adopt) multiple folded conformations can be an important evolutionary parameter, perhaps as critical as the RNA's chemical reactivity. We obtained the unusual result that the ligase ribozyme selected in low Mg^{2+} concentrations is not a better catalyst but instead folds into reactive conformations with higher probability than the E_{100}(#3) sequence. This suggests that the assumption of a one-to-one correspondence between genotype and phenotype (i.e., perfect or unity H^2) is not supported in this case. These findings have direct bearing on the design and implementation of *in vitro* selection experiments, on the plausibility of the RNA world, and on the epigenetic characteristics of single molecules.

Results and discussion

Continuous evolution

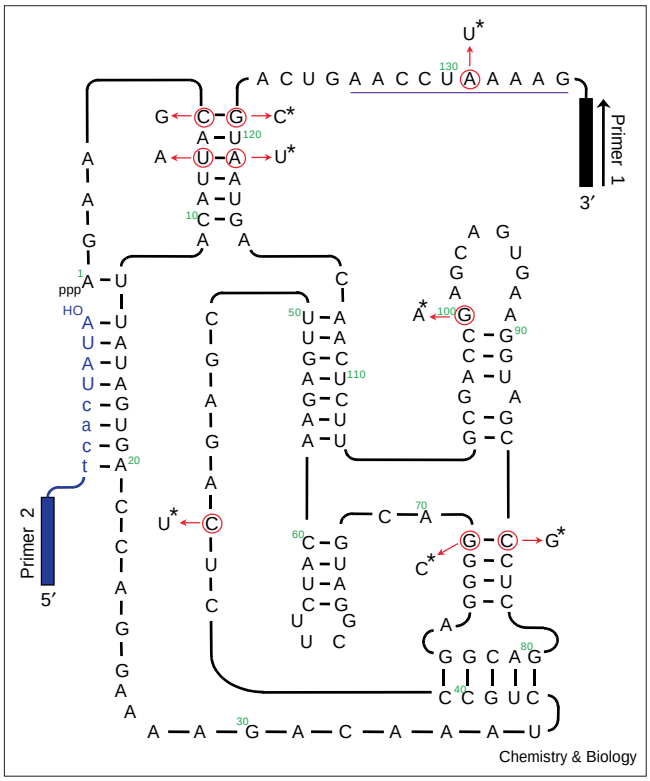
The endpoint of the continuous evolution experiment of Wright and Joyce [16] was a population dominated by the highly reactive E_{100}(#3) sequence (Figure 1). E_{100}(#3) is a 152 nucleotide (nt) RNA that can perform sequence-dependent phosphodiester-bond formation (ligation) to a short exogenous oligonucleotide substrate. The reaction is very rapid and efficient; when phenol-extracted and Sephadex-purified RNA transcripts were measured at pH 8.5 without any thermal pre-treatment, estimates of an apparent first-order rate constant (k_{cat}) for self-ligation greater than 20 min^{-1} and of a catalytic efficiency (k_{cat}/K_m) of 1×10^7 M^{-1} min^{-1} were reported [16]. This represents a 12,500-fold improvement in catalytic efficiency over the molecule used as a starting point for the *in vitro* selection, which was, in turn, based on the endpoint of a selection by Bartel and colleagues [18,19] that produced a ligase ribozyme from a pool of random RNAs. We performed assays of the E_{100}(#3) molecule by incubating it with an excess of substrate in 25 mM Mg^{2+} at 37°C and pH 8.5. For these assays, we isolated the ribozyme via run-off transcription from a double-stranded polymerase chain reaction (PCR) product and subsequently purified the RNA using denaturing-gel electrophoresis and rehydrated it in water at room temperature. Initially we did not subject the RNA to a thermal denaturation/renaturation pretreatment. On average, only 50% of input ribozyme performs the ligation in 30 min, and an overnight incubation allows only an additional 10% to react (Figure 2). Under these conditions we never observed 100% reaction; the assays were performed in at least fivefold molar excess of substrate over ribozymes such that the reverse reaction was expected to be heavily disfavored by Le Chatelier's principle. As noted before [16,20], the inability of these ligase ribozymes to achieve complete reaction is probably because the molecules present are not all in the correctly folded conformation.

During continuous evolution *in vitro*, RNAs can react and are subject to selection immediately upon transcription. We reasoned that the dynamic mixture of components extant during the continuous protocol would

present an evolving population of ribozymes with a complex set of selection pressures that would exacerbate any folding heterogeneities present in the evolving pool. Over many bursts of continuous evolution then, sequences that can best manifest all the components of fitness, including survival (catalysis), reproduction (transcription), and phenotypic expression (proper folding), should have a selective advantage. Accordingly, we designed a strategy to select for ribozymes that were catalytic in an Mg^{2+}-poor environment with the hope that conformations different from those exhibited by E_{100}(#3) would be favored. If phenotypic homogeneity were an important component of fitness during continuous evolution, then the heritability of this trait (folding) should increase above any subunity value present in a population composed solely of the E_{100}(#3) sequence. The ionic medium, especially that of divalent cations, is a critical determinant of ribozyme folding [21] and previous selections for altered, reduced or absent divalent cation compositions have engendered many novel ribozymes [22–25]. Lowering the concentration of Mg^{2+} has been shown to lower the extent of folding in the *Tetrahymena* ribozyme [21], and therefore we expected that a drop in Mg^{2+} availability during continuous evolution would grant a selective advantage to molecules with a greater tendency to adopt an active conformation.

Starting with an 8% per-position randomization of the E_{100}(#3) sequence, we selected variants that were active in an increasingly Mg^{2+}-poor environment (Figure 3). The total Mg^{2+} concentration was reduced by 2.5 mM every fourth burst, starting with 25 mM total Mg^{2+}. Because during the continuous protocol a high concentration of nucleotides (NTPs and dNTPs; present at a sum of 8.8 mM at the beginning of each burst) is also in solution with Mg^{2+} coordinating to the 5′-triphosphates of these in stoichiometric amounts, the free Mg^{2+} available to ribozymes is below the total Mg^{2+} added to the reaction buffer. Over 20 bursts of selection, therefore, the free Mg^{2+} was slowly weaned from ~25 mM in the first four bursts to ~3.7 mM in the last four bursts. Note that as a burst of continuous evolution proceeds and RNA is synthesized from free ribonucleotides in an exponential fashion, the free Mg^{2+} concentration actually increases as nucleotides are spent. Consequently the maximum Mg^{2+} concentration during this experiment approached 25 mM during the later minutes of the first bursts and the minimum Mg^{2+} concentration approached 3.7 mM during the first minute of the last burst (12.5–8.8 mM = 3.7 mM). The presence of ligating RNA molecules was monitored using gel electrophoresis of α^{32}P[ATP]-labeled RNA samples taken at 10 min intervals during each burst. As expected from a selection protocol in which both the RNA population and the environment are constantly changing throughout the experiment, each sample revealed a different amount of RNA and a different proportion of ligated

Figure 1



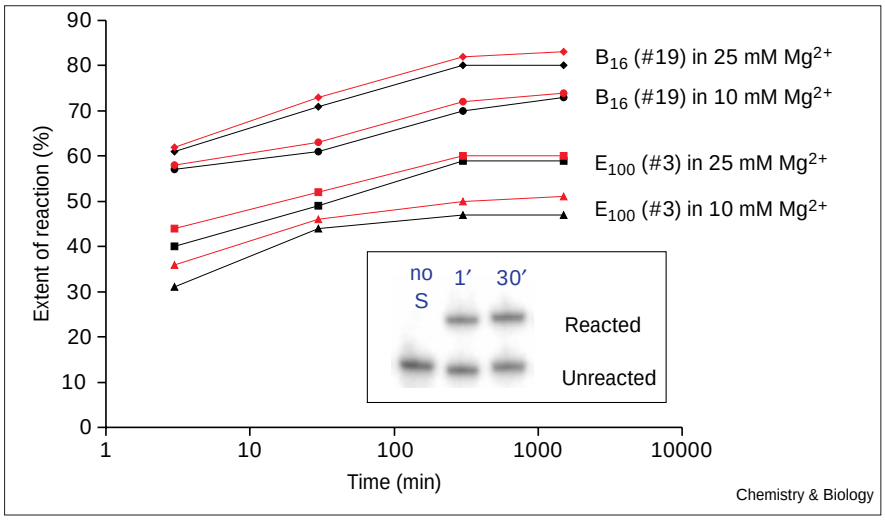
The presumed primary sequence and secondary structure of the E_{100}(#3) ribozyme selected by Wright and Joyce [16]. The secondary structure depicted is based on the covariance of nucleotides when the class I ligase ribozymes were originally isolated from a random pool of RNA sequences [18,20]. The exogenous substrate for ligation is depicted in blue with lowercase nucleotides representing deoxynucleotides; the 17 3′ nucleotides of this substrate form the promoter for T7 RNA polymerase when double stranded. The boxed regions indicate the primer-binding regions for cDNA synthesis during continuous evolution (primer 1) and for PCR amplification of cDNA (primer 1 and primer 2). Mutations in red were found to be fixed after 16 bursts of continuous evolution with selection for activity in a Mg^{2+}-poor environment (position #44 was mutated C→U in only 14 of 15 clones). Mutations with asterisks were found in one of 15 clones after seven bursts. Purple underline represents region where single mutations occur without patterns in all bursts analyzed.

(versus unligated) material. RNA was detected in all bursts, however, and from a qualitative standpoint, the later bursts appeared to generate an increase in the ligated proportion (data not shown).

Genotypic characterization

To characterize the genotypic and phenotypic composition of the evolving population, we obtained individual ribozyme sequences by cloning 25, 15 and 15 PCR amplicons derived from aliquots taken after bursts 7, 13 and 16 (i.e., B_7, B_{13} and B_{16}, respectively). After B_7 (20 mM total Mg^{2+}), the E_{100}(#3) sequence was still dominant, occurring completely unmutated in 28% of the clones (7/25). An additional 13 clones were single-error mutants in the

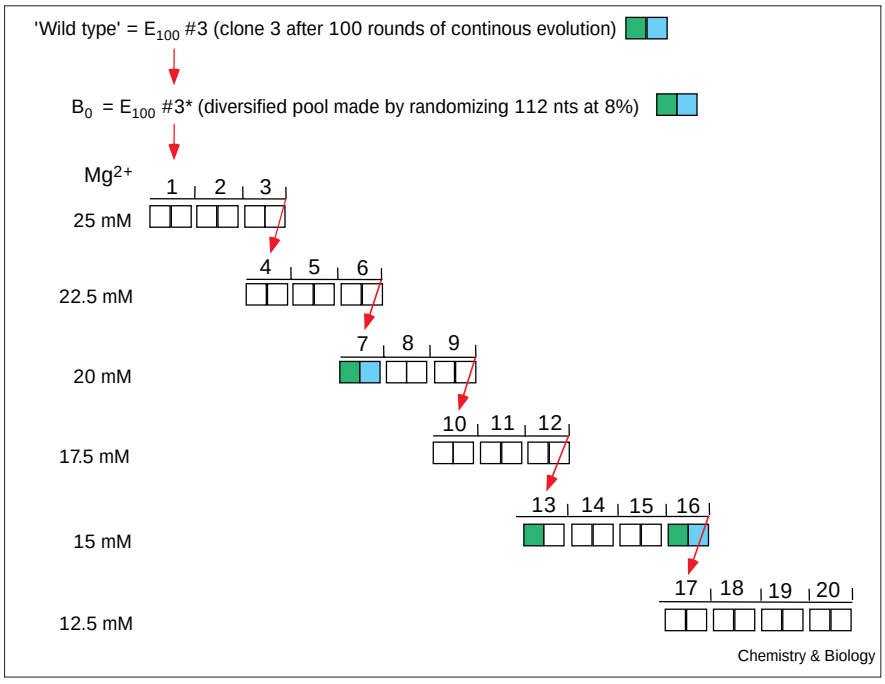
Figure 2



Time course of ligation reaction with gel-purified, phenol-extracted and Nensorb-desalted ribozymes. Reactions were performed with substrate in fivefold excess over ribozymes and reaction products were separated on denaturing 5% polyacrylamide gels containing 8 M urea (the inset shows an example gel; no S = reaction without substrate). Black lines, no thermal denaturation/renaturation of ribozyme; red lines, 2 min pre-incubation of ribozyme at 80°C and cooling to room temperature in reaction buffer prior to incubation with substrate.

last ten 3′ nucleotides (Figure 1), a region found to vary essentially inconsequentially in the ligase ribozyme [16]. Five clones exhibited mutations in the core of the ribozyme, however. One of these clones contained six mutations, two of which (positions #71 and #84) swap a G•C pair in an internal stem, on the basis of the proposed secondary structure of the ligase ribozyme [20]. Provocatively, two other mutations (#119 and #121) in this clone disrupt one-half of another putative stem, without compensatory mutations on the other side of the stem. The other four of these clones had single-nucleotide mutations not found after further rounds of selection. After B_{13} (total Mg^{2+}, 15 mM), the population had clearly shifted its genotypic focal point, as eight mutations were completely fixed and a ninth was 60% fixed (position #44 was mutated in 9 of 15 clones). Two of these fixed mutations were those found in the lone B_7 clone that putatively disrupted a stem, but in B_{13} they were always found with the compensatory mutations (#5 and #7) that restore the proposed stem (Figure 1). After B_{16} the eight fixed mutations

Figure 3



The 'slow wean' strategy for selecting ligase ribozymes in an increasingly Mg^{2+}-poor environment. The E_{100}(#3) ribozyme was mutagenized at the 8% level and then subjected to selection for three or four bursts at decreasing levels of total Mg^{2+} concentration. One picomole of mutagenized RNA were used to seed the first burst. Each subsequent burst involved a 1000-fold dilution of the resultant RNA that was generated by transcription. Only those ribozymes that could successfully ligate substrate under the conditions of continuous evolution will be amplified by *in vitro* transcription by T7 RNA polymerase [16]. Dark green squares indicate when ribozymes were cloned for sequencing and light blue squares indicate when activity assays were performed.

in B_{13} remained fixed, and position #44 was now mutated in 14 of 15 clones. The dominant nine-error sequence in B_{13} and B_{16}, exemplified as clone #19 from B_{16} and designated B_{16}(#19), represents the endpoint of our genotypic evolution in this experiment.

Phenotypic characterization

To begin to study the phenotypic characteristics of the starting (E_{100}(#3)) and evolved (B_{16}(#19)) sequences, we transcribed these molecules from DNA templates, purified them using denaturing-gel electrophoresis in 8 M urea, desalted them on Nensorb columns, rehydrated them in water, and quantitated them using UV spectrometry. When we assayed this material for ligation activity at 37°C in the buffers used for continuous evolution without any thermal pretreatment, the B_{16}(#19) ribozyme demonstrated an apparent rate enhancement. In both 25 mM and 10 mM Mg^{2+}, B_{16}(#19) generated a greater proportion of ligated product than the E_{100}(#3) sequence (Figure 2). Because, under these assay conditions, as during continuous evolution, substrate was always present in vast excess such that binding could not be rate-limiting, it was tempting to ascribe the improvement of the B_{16}(#19) sequence over E_{100}(#3) to a chemical-catalytic enhancement of the former that endowed it with a greater fitness during continuous evolution. When the ribozymes were allowed to denature for 2 min at 80°C in water, and then equilibrate for 5–10 min to room temperature in 1× reaction buffer prior to reacting with substrate, both sequences improved their 30 min reactivities, but this effect was more pronounced for E_{100}(#3) than for B_{16}(#19) (Figure 2). The biphasic reaction kinetics shown in Figure 2 suggest that there are kinetically distinct conformations of the E_{100}(#3) and B_{16}(#19) ribozymes. This mirrors the situation for the hairpin ribozyme, in which active ribozymes bind substrate and react quickly, whereas inactive ribozymes bind substrate, undergo a slow thermal refolding into the active state, and then react [10].

To isolate more precisely the qualities that distinguish these two ribozymes, we obtained estimates of the apparent first-order self-ligation rates (k_{cat}) under various conditions. These experiments were designed to separate the folding and chemical stages of catalysis. In our studies with the ligase ribozyme, kinetic evaluation is difficult because the reaction proceeds extremely rapidly under the cationic and temperature conditions of continuous evolution [16]. We were able to estimate k_{cat} values for denaturing-gel-purified E_{100}(#3) and B_{16}(#19) ribozymes to at least one significant digit, however. Under single-turnover conditions, where substrate was limiting and ribozyme was present in at least tenfold excess, we projected k_{cat} for both sequences to be greater than 20 min^{-1} in either 25 mM or 10 mM Mg^{2+}. These estimates were made by following the appearance of product from 5–120 s with unlabeled ribozyme and 5′-[γ^{32}P]-labeled substrate in the buffers used for continuous evolution (pH 8.5). Because the half-times of these reactions were very fast, often less than 2 s, when derived from manual pipetting methods, these k_{cat} estimates have significant error and as such provide little information regarding the relative catalytic rates of the two ribozymes.

We therefore obtained k_{cat} estimates using a modified continuous-evolution buffer in which the pH was lowered to 6.0. The rate of the chemical step of ribozyme-catalyzed phosphoryl-transferase reactions, a nucleophilic attack of a hydroxyl moiety on a phosphorus, in the case of the class I ligase, may be pH dependent ([26]; D. Bartel, personal communication) to the extent that lowering the pH could retard the reaction enough to highlight chemical-catalytic differences between E_{100}(#3) and B_{16}(#19). At pH 6 and 25 mM Mg^{2+}, our k_{cat} estimates were 7.7 min^{-1} and 18 min^{-1} for E_{100}(#3) and B_{16}(#19), respectively (Table 1). We did not thermally pretreat these ribozymes before they were mixed with substrate to initiate reaction. When we did first denature these ribozymes for 2 min at 80°C, their k_{cat} estimates essentially reversed, with E_{100}(#3) improving to 18 min^{-1} and B_{16}(#19) dropping to 8.1 min^{-1}. The half-times for these reactions ranged from 3–30 s, allowing greater accuracy than at pH 8.5. Even allowing for up to 40% error (a 2 s deviation from a 5 s time point), two trends are clear from these data. First, the chemical-cleavage rate of neither ribozyme is markedly better than the other, and second, a thermal denaturation/renaturation step enhances the k_{cat} estimate for the E_{100}(#3), but lowers that of B_{16}(#19). In addition, a thermal denaturation/renaturation has no effect on the projected maximum extent of reaction for B_{16}(#19) at 25 mM Mg^{2+}, whereas it increases the value for E_{100}(#3) from 75–80% to match that for B_{16}(#19) (Table 1). We also observed these trends to a lesser degree at 10 mM Mg^{2+}, where we noted that the B_{16}(#19) ribozyme was hindered less by the drop in Mg^{2+} concentration than the E_{100}(#3) ribozyme (Table 1). The selection of B_{16}(#19) during *in vitro* evolution may not, therefore, have been based solely on catalytic ability, as measured by k_{cat}, at least to the extent that these estimates would translate to the conditions under which we performed continuous evolution. Instead the tendency to fold correctly may have had a substantial effect on fitness; during transcription at 37°C the B_{16}(#19) sequence may be prone to folding into an active conformation, whereas a complete denaturation and renaturation may be necessary to bring more E_{100}(#3) molecules into this state. Within our measurement error, these ribozymes both generated a k_{cat}/K_m efficiency of 1×10^7 M^{-1} min^{-1} in this range of Mg^{2+} values, confirming the earlier report [16].

The inversion of relative performances of the 'wild-type' and 'evolved' sequences under differing environmental conditions — thermal denaturation/renaturation versus

Table 1

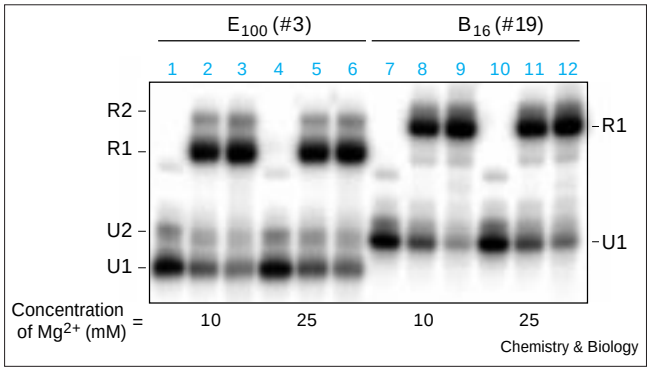
Estimates of kinetic parameters for ribozymes.

Ribozyme	Conditions	*A*	k_{cat} (min^{-1})
E_{100}(#3)	pH 8.5, no heat/cool, 10 and 25 mM Mg^{2+}	0.75–0.86	> 20 min^{-1}
E_{100}(#3)	pH 6.0, no heat/cool, 25 mM Mg^{2+}	0.75	7.7 min^{-1}
E_{100}(#3)	pH 6.0, 2′ heat/cool at 80°C, 25 mM Mg^{2+}	0.80	18 min^{-1}
E_{100}(#3)	pH 6.0, no heat/cool, 10 mM Mg^{2+}	0.69	5.9 min^{-1}
E_{100}(#3)	pH 6.0, 2′ heat/cool at 80°C, 10 mM Mg^{2+}	0.76	6.8 min^{-1}
B_{16}(#19)	pH 8.5, no heat/cool, 10 and 25 mM Mg^{2+}	0.86	> 20 min^{-1}
B_{16}(#19)	pH 6.0, no heat/cool, 25 mM Mg^{2+}	0.81	18 min^{-1}
B_{16}(#19)	pH 6.0, 2′ heat/cool at 80°C, 25 mM Mg^{2+}	0.81	8.1 min^{-1}
B_{16}(#19)	pH 6.0, no heat/cool, 10 mM Mg^{2+}	0.75	16 min^{-1}
B_{16}(#19)	pH 6.0, 2′ heat/cool at 80°C, 10 mM Mg^{2+}	0.82	9.7 min^{-1}

Ribozymes were incubated with substrate either without or with a 2 min pre-incubation in water at 80°C followed by an equilibration to room temperature in reaction buffer prior to mixing with substrate. *A* is the projected maximum extent of reaction; the ribozyme concentration was 0.5 mM, and the substrate concentration 5.0 nM; k_{cat} is the estimated apparent first-order self-ligation rate of ribozyme.

constant temperature — bolsters the notion that the environment dramatically affects the phenotypic expression of these molecules, and that there may be subtle alterations in structure–function relationships that are not evident when phenotypes are measured from ribozymes purified using denaturing gel electrophoresis. Consequently, we characterized these ribozymes using native-gel electrophoresis to assess their heterogeneities in folding conformation. First, to explore the properties of the gel-purified ribozymes, we separated folded ribozymes on nondenaturing gels in various buffers and with and without substrate present (Figure 4). For the E_{100}(#3) and the B_{16}(#19) ribozymes, at least two predominant bands, or conformers, were present. These bands suggest conformational heterogeneity at the secondary and/or tertiary levels; multiple conformers may even exist in each band, as is the case with the hairpin ribozyme [10]. To control for the possibility that the conformers were the products of contaminating primary sequences in the DNA stocks, we reverse-transcribed all four major conformers, amplified them using PCR and determined their sequences. Both E_{100}(#3) conformers generated the E_{100}(#3) sequence, whereas both B_{16}(#19) conformers generated the B_{16}(#19) sequence (data not shown).

Figure 4



Native-gel electrophoresis of gel-purified ligase ribozymes. RNA conformers were separated on native 5% polyacrylamide gels without urea after incubation with substrate at 37°C. Major unreacted (U1 and U2) and reacted (R1 and R2) conformers are indicated. Lanes 1–3 and 7–9, incubation in continuous-evolution buffer containing 10 mM $MgCl_2$. Lanes 4–6 and 10–12, incubation in continuous-evolution buffer containing 25 mM $MgCl_2$; lanes 1, 4, 7 and 10, no substrate present in reaction; lanes 2, 5, 8 and 11, 1 min incubation with substrate; lanes 3, 6, 9 and 12, 30 min incubation with substrate.

With gel-purified material, neither the relative mobilities nor the relative proportions of each conformer are equivalent between the E_{100}(#3) and the B_{16}(#19) ribozymes (Figure 4). The lower conformer of E_{100}(#3), which we denote U1, ligates to form the lower product (R1), and the upper band, which we denote as U2, ligates to form the upper product (R2; Figure 4, lanes 3, 6, 9 and 12). The U1 conformer comprises ~75% of (unligated) E100(#3) RNA gel-purified material, although this value varies from preparation to preparation and can be as low as 50%. In contrast, more than 80% of the B_{16}(#19) RNA is reproducibly contained in the lower conformer, which migrates slower than E_{100}(#3) U1, but is also designated U1 to indicate it is the fastest-migrating conformer of B_{16}(#19). The B_{16}(#19) gel-purified material sometimes contains small amounts of upper conformers, designated U3 and U4, that ligate to form upper products (R3 and R4). All conformers are thermally stable: in 30 min reaction times at 37°C, we detect no interconversion among conformers, and an overnight incubation generates less than 5% interconversion of any conformer into another. Moreover, when excised from the gel, electroeluted and re-run on a native gel, the conformers remigrate to their original positions. In Figure 5, lane 1 shows both unligated and both ligated conformers of E_{100}(#3), as in Figure 4, lane 6. Lanes 2–4 show the results of the excision and treatment of just the U2 conformer, and lanes 5–7 show the results of the excision and treatment of just the U1 conformer. Lanes 8–10

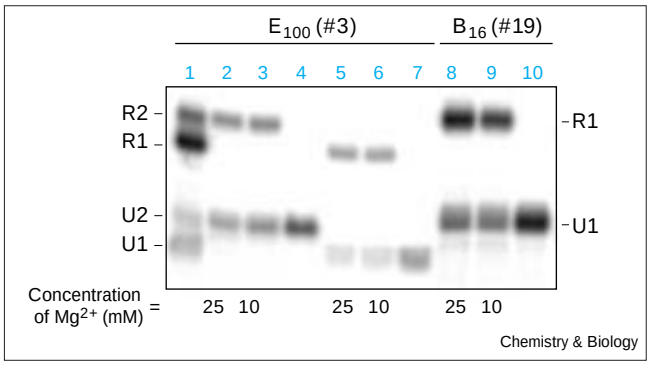
demonstrate that the U1 conformer of B_{16}(#19) will remigrate to its original location as well. Of note is the fact that these bands can differ in their reactivities. Under these assay conditions — gel-purified, electroeluted and ethanol-precipitated ribozymes incubated without thermal pretreatment for 30 min at 37°C in fivefold substrate excess at pH 8.5 — the lower E_{100}(#3) conformer reacts to a greater percentage extent than the upper conformer, whereas both conformers react to greater extents in higher Mg^{2+} concentrations (Table 2). Alternate folding conformers therefore appear to exhibit different reactivities, suggesting that a single RNA sequence can present multiple phenotypes.

Simulated continuous evolution

An analysis of the conformational heterogeneity of gel-purified material is not entirely reflective of the RNA structures that are formed during *in vitro* transcription in the course of continuous evolution, however. We therefore performed native-gel electrophoresis on the products of *in vitro* transcription under the exact conditions of continuous evolution. The E_{100}(#3) and B_{16}(#19) sequences were transcribed from DNA templates at 37°C in the presence of substrate, rNTPs, dNTPs, reverse transcription primer and buffer such that the conditions of continuous evolution were mimicked as precisely as possible. Only reverse transcriptase was omitted relative to the continuous evolution protocol. After one 30 min burst of transcription, RNAs were immediately quenched on ice and loaded on nondenaturing gels, which were then run as quickly as possible at ≤ 37°C. We performed transcriptions in both high and low Mg^{2+} by using continuous-evolution buffers containing 25 mM and 15 mM total Mg^{2+}; the free Mg^{2+} in these reactions thus centered around 20 mM and 10 mM, respectively. The resulting RNAs exhibited conformational heterogeneity but in a different manner than with the gel-purified material (Figure 6; compare lanes 1 and 3 or lanes 12 and 14). Multiple bands were observable for both molecules; in both higher and lower Mg^{2+} we delineated five major, distinct, unligated bands for E_{100}(#3) and four for B_{16}(#19). In addition to the U1 and U2 conformers that dominated the gel-purified ribozymes, a slow-migrating U3 conformer and an even slower migrating U4 conformer appeared for both the E_{100}(#3) and B_{16}(#19) ribozymes. The E_{100}(#3) ribozyme also exhibited a conformer just above the dominant U1 conformer that appears not to be reactive (see below), and was therefore designated U0. We measured transcription-to-transcription variation by performing four replicate transcriptions, spanning different days and different reagent and template preparations. As is the case with gel-purified conformers, we observed more variation in the relative frequencies of the E_{100}(#3) conformers than of the B_{16}(#19) conformers (Table 2).

When the accumulation of ligated conformers during *in vitro* transcription conditions was followed, either by observing a gel-shift of ribozymes body-labeled with [α^{32}P]ATP or by gel-shift of substrates previously 5′-end-labeled with [γ^{32}P]ATP, we could distinguish roughly the same numbers of ligated as unligated conformers. Four ligated products were clearly visible for both ribozymes (Figure 6). From gel-purification and religation studies such as those shown (Figure 5), we were able to assign reacted conformers to unreacted conformers from which they formed (i.e., U3 reacts to form only R3). Unfortunately, we were not able to isolate enough material from each conformer to perform independent kinetic analyses on each. And, although quantitation of the relative proportions of ligated conformers is tainted by the fact that this is a dynamic system and new reactants are generated during the course of the experiments, it is again evident that some conformers react to a greater extent in 30 min than others (Table 2). For example, the U4 conformer is the always the most reactive, whereas the U0 conformer of E_{100}(#3) does not generate any detectable product and may be a salt-front artifact. After a 30 min incubation with substrate, at least 50% of U4 conformers of both ribozymes converted to their ligated forms R4, which migrate just above the corresponding U4 forms. Of particular significance is the observation that the U2 conformer of E_{100}(#3) is more reactive than the more abundant U1 conformer under these simulated continuous-evolution conditions (Figure 6; lane 3), in contrast to the assay of gel-purified material in which U1 was more reactive than U2 (Figure 5; lanes 2 and 5). These data indicate that either U1 is in a different, more reactive conformation following gel purification or that it reacts differently under different reaction

Figure 5



Isolation by gel purification and subsequent reaction of individual conformers. Reaction products were separated on a native 5% polyacrylamide gel without urea after 30 min incubations of ribozyme with excess substrate in continuous-evolution buffers containing either 25 mM $MgCl_2$ (lanes 1, 2, 5 and 8) or 10 mM $MgCl_2$ (lanes 3, 6 and 9). Lanes 4, 7, and 10, gel-purified, electroeluted conformers re-run on native gel without incubation with substrate; lanes 2, 3, 5 and 6, conformers of E_{100}(#3) cut out of gel, electroeluted, incubated with substrate, and re-run on native gel; lanes 8 and 9, U1 conformer of B_{16}(#19) cut out of gel, electroeluted, incubated with substrate, and re-run on native gel; lane 1, gel-purified total E_{100}(#3) incubated with substrate and run as a location control.

Table 2

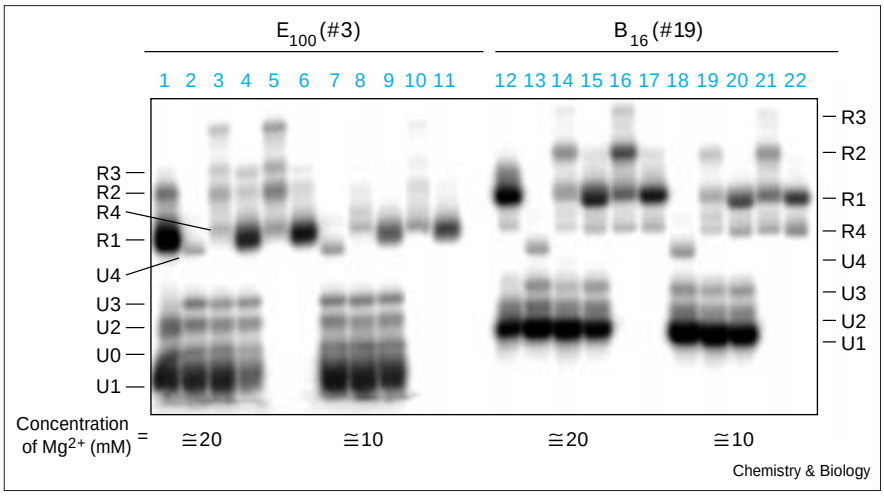
Frequencies (*f*) and estimated reactivities (*z*) of ribozymes and their folding conformers.

Ribozyme or conformer	*f*, gel-purified	*z*, gel-purified (25 mM Mg^{2+})	*z*, gel-purified (10 mM Mg^{2+})	*f*, CE in higher Mg^{2+}	*z*, CE in higher Mg^{2+}	*f*, CE in lower Mg^{2+}	*z*, CE in lower Mg^{2+}
E_{100}(#3)	–	49%	44%	–	11%	–	2.1%
E_{100}(#3)-U0	0%	–	–	9.0% (0.29%)	0% (0.0%)	9.6% (0.41%)	0% (0.0%)
E_{100}(#3)-U1	76%	70%	67%	48% (6.7%)	0% (0.0%)	58% (5.4%)	0% (0.0%)
E_{100}(#3)-U2	22%	55%	48%	21% (3.4%)	21% (4.8%)	17% (2.9%)	7.7% (0.42%)
E_{100}(#3)-U3	0%	–	–	16% (2.5%)	22% (1.7%)	11% (1.8%)	0% (0.0%)
E_{100}(#3)-U4	1.0%	ND	ND	6.0% (1.5%)	58% (9.8%)	4.5% (0.56%)	99% (0.81%)
B_{16}(#19)	–	71%	61%	–	11%	–	5.9%
B_{16}(#19)-U1	80%	66%	64%	80% (1.4%)	5.5% (0.43%)	82% (0.48%)	4.5% (0.08%)
B_{16}(#19)-U2	18%	69%	61%	11% (1.1%)	25% (5.1%)	10% (0.029%)	13% (2.1%)
B_{16}(#19)-U3	0%	–	–	5.7% (0.36%)	12% (1.6%)	5.3% (0.079%)	6.7% (0.09%)
B_{16}(#19)-U4	1.0%	ND	ND	2.7% (0.19%)	100% (0.0%)	2.9% (0.37%)	100% (0.0%)

Frequencies were determined as the fraction of conformers present in a native polyacrylamide gels run at 37°C or lower. Reactivities were estimated either from total percentage of ribozyme in the ligated form after 30 min at 37°C with excess substrate as determined from aliquots simultaneously run on denaturing-gel electrophoresis (ribozymes) or from percentages of nascent transcripts in the ligated form on native gels (conformers). Concentrations of Mg^{2+} are 25 mM initially for the 'higher' columns and 15 mM for the 'lower' columns; these values represent the total amount of $MgCl_2$ added to the reaction; the free Mg^{2+} will vary over the course of the reaction beginning at a value of ~9 mM lower than the total amounts added. Percentage values may not sum to exactly 100% because of round-off error; all values are given to two significant digits. Values for conformers under the conditions of continuous evolution (CE) are the averages of four replicate transcriptions; standard errors (s.d./√n) are given in parentheses. ND, not determined.

conditions. In fact, after 30 min with substrate, no detectable R1 product above gel background can be observed for the E_{100}(#3) ribozyme; all of the product in the R1 region of the gel is attributable to R4 (Figure 6; lane 3). In higher Mg^{2+}, R2 and R3 products are also visible (Figure 6; lanes 3 and 4). In contrast to the E_{100}(#3) ribozyme, the dominant U1 form of the B_{16}(#19) ribozyme does generate a ligated form (Figure 6; lane 14). For this ribozyme, the U4 conformer appears to react to completion, and R2 and R3 products are also observed.

Figure 6



Native-gel electrophoresis of ribozymes under the conditions of continuous evolution. Lanes 1 and 12, gel-purified and ligated ribozymes run as position controls. All other lanes, ribozymes transcribed from double-stranded DNA templates in continuous-evolution buffer containing either 25 mM total Mg^{2+} (lanes 1–6 and 13–17) or 15 mM total Mg^{2+} (lanes 7–11 and 18–22); free Mg^{2+} available to ribozyme folding in these experiments will be roughly 5 mM below these values (see text) as indicated at the bottom of the figure. Lanes 2, 7, 13 and 18, transcriptions performed for 30 min with [α^{32}P]ATP but without substrate present; lanes 3, 8, 14 and 19, transcriptions performed for 30 min with [α^{32}P]ATP in the presence of vast substrate excess; lanes 5, 10, 16 and 21, transcriptions performed for 30 min without [α^{32}P]ATP but with 5′-[γ^{32}P]ATP-labeled substrate present in vast excess (visible bands in these lanes represent only ribozymes that have bound to or ligated to a substrate molecule); lanes 4, 9, 15 and 20, transcriptions performed for 29 min with [α^{32}P]ATP but without substrate present, and then an additional one minute after the addition of 64 pmol substrate; lanes 6, 11, 17 and 22, transcriptions performed for 29 min without [α^{32}P]ATP or substrate present, and then an additional one minute after the addition of 64 pmol 5′-[γ^{32}P]ATP-labeled substrate (visible bands in these lanes represent only ribozymes that have bound to or ligated to a substrate molecule).

When we ran aliquots of the products from the experiments shown in Figure 6 on a denaturing polyacrylamide gel and quantitated the proportions of ligated RNA (now in a single band) by phosphorimaging, there was, in fact, little difference between the E_{100}(#3) and B_{16}(#19) ribozymes (10.7% and 11.2%, respectively); these results are reproducible over multiple transcriptions and are not significantly different (*t*-test, $P > 0.1$). Although this would generally suggest that the two ribozymes have similar phenotypes, our data from native gels show that conformational heterogeneity exists with single genotypes, even under the conditions of continuous evolution, and that the phenotypes may differ in their relative fitnesses.

When we allowed transcription to proceed in the lower concentration of Mg^{2+} (total Mg^{2+} was 15 mM, the same conditions under which the B_{16}(#19) sequence appeared during the original continuous evolution experiment), RNA production was still high; the optimal total Mg^{2+} concentration is reported as 8–15 mM for commercial T7 RNA polymerase. Under these conditions nearly all conformers reacted to a lesser percentage extent than under the higher Mg^{2+} concentration, although the U4s appeared to retain the most activity. The U1 conformer of E_{100}(#3) still lacked activity, whereas the U1 conformer of B_{16}(#19) retained activity. Also, although the U2 conformer of E_{100}(#3) failed to produce ligated products above background under these conditions, the U2 conformer of B_{16}(#19) did retain activity. In large part because the most frequent conformer for the B_{16} (#19) sequence (U1 = 82% of total unligated material) is reactive, whereas the most frequent conformer for the E_{100}(#3) sequence (U1 = 58% of total unligated material) is not, the B_{16}(#19) exhibits almost three times the proportion of ligated product as does E_{100}(#3) at low Mg^{2+} (5.9% versus 2.1% as measured from quantitation of a denaturing gel). Hence when the Mg^{2+} concentration was weaned over the course of the bursts of continuous evolution, we infer that ribozymes with some or all of the mutations found in B_{16}(#19) were presented with a selective advantage, not because they were intrinsically better catalysts with faster chemical-step kinetic rates, but because a greater proportion of them were in a more active folded conformation.

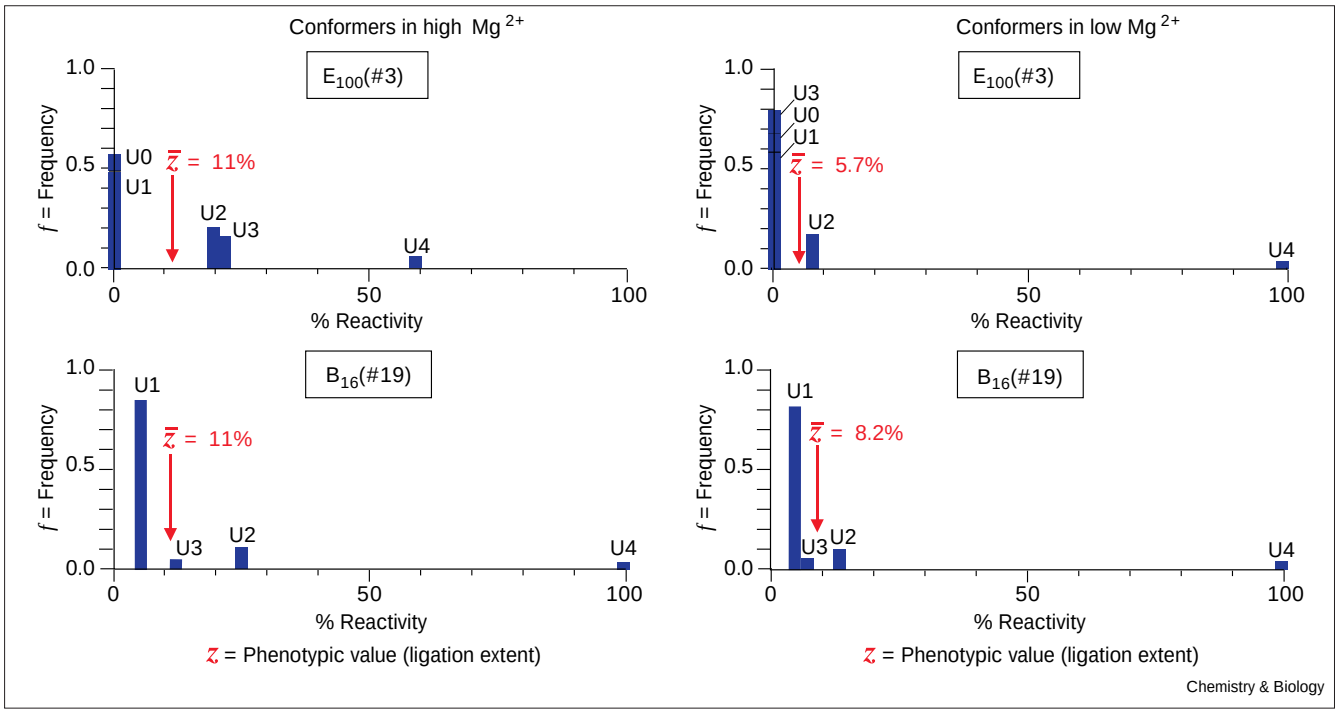
Model of conformational heterogeneity

It is likely that the differences in reactivities of U1–U4 of E_{100}(#3) and B_{16}(#19) are a consequence of distinct conformations. Four of the nine mutations that separate these two sequences are found in the first putative stem structure downstream of the 5′ end of the molecule (Figure 1). These mutations appear merely to swap bases in base pairs, while having little effect on the overall structure. However, because two of these nucleotides are transcribed very early (positions #5 and #7), and the other two are transcribed very late (positions #119 and #121), the possibility exists that an inactive conformation could be more likely to form during transcription of a particular sequence. For instance, at low Mg^{2+} E_{100}(#3)-like sequences may be more prone to being trapped in an alternative conformation formed after positions #119 and #121 are transcribed but before they can fold to form the aforementioned stem. The appearance of a mutant in B_7 (20 mM total Mg^{2+}) with nucleotides #119 and #121 mutated but not #5 and #7 is consistent with this hypothesis. Furthermore, Sabeti *et al.* [27] predicted that longer ligase ribozymes would be more likely to misfold into less active conformers than shorter ribozymes.

To explore further the effects on the relative proportions of conformers by the exact conditions of transcription, we slowed the transcription rate by lowering the concentration of ATP from 2 to 1 mM (data not shown). This should allow ribozymes a greater opportunity to fold into a more thermodynamically favorable conformation. In 25 mM Mg^{2+}, no effect was observed on E_{100}(#3); the same proportion of unligated and ligated conformers was seen and overall ligation improved only from 13–15%. A dramatic effect was observed on B_{16}(#19), however; the proportion of R2 increased and the overall extent of ligation improved almost tenfold to 73%. Similarly, in 10 mM Mg^{2+}, no effect on E_{100}(#3) ligation extent was observed at 1 mM ATP, whereas that of B_{16}(#19) improved by 20%. We interpret these data as additional evidence for the heightened propensity of the B_{16}(#19) sequence to fold into active conformations. When transcription is slowed, B_{16}(#19) has even better access to rarer, but more chemically reactive, conformations, whereas E_{100}(#3) is apparently still trapped in less reactive conformations.

It is also likely that substrate binding plays a role in the folding of the ribozyme during transcription. Substrate is always present in vast excess (at least tenfold stoichiometrically) during continuous evolution and can potentially bind to the internal guide sequence (IGS) of the ribozyme as soon as the first 20 nucleotides are transcribed (Figure 1). When we allowed transcription and ligation to proceed concurrently for the full 30 min in 25 mM Mg^{2+}, both the E_{100}(#3) and the B_{16}(#19) ribozymes exhibit a lower total accumulation of ligated product than when we allowed transcription to proceed for 29 min in the absence of substrate followed by only 1 min after the addition of substrate (Figure 6; lanes 4, 6, 9, 11, 15, 17, 20 and 22). The enhancement in the latter case is almost entirely contained within the R1 conformer of E_{100}(#3) and the R1 and R4 conformers of B_{16}(#19). The appearance of a strong R1 conformer band for E_{100}(#3) after only 1 min with substrate after 29 min without, compared with a lack of an R1 conformer after 30 min of transcription in the presence of substrate, strongly suggests that the substrate prevents the E_{100}(#3) transcript from folding into an active conformation. We surmise that, as the nascent RNA is being transcribed, the substrate binds and induces a folded state that does not catalyze ligation. Because adding substrate after 29 min increases the extent of reaction for both ribozymes, they are

Figure 7



Frequencies and reactivities of individual conformers under simulated conditions of continuous evolution. Values were taken from experiments such as that shown in Figure 6. The mean genotype ($\bar{z}$) for each ribozyme under a given environmental condition is given by Σ(conformer frequency)(conformer activity) over all conformers.

each subject to substrate-dependent misfolding. But B_{16}(#19) is much less so, as evidenced by its generation of R1 products even when substrate is always present during transcription. Therefore, although the E_{100}(#3) conformers could inherently possess the most catalytic power, they cannot fold into an active conformation in the presence of substrate, placing them at a selective disadvantage during the continuous evolution experiment. The above phenomenon is similar to the formation of metastable RNA during viroid transcription [28], and underscores the selective advantage for nucleic acids that can buffer themselves against less active conformations. These results further emphasize the possibility that the outcome of an *in vitro* evolution experiment may not necessarily be the ribozymes that are the most catalytically active.

To investigate further the effects of substrate, we repeated the experiments of Figure 6 but with a modified substrate that replaces the 5′ half of the T7 RNA polymerase promoter and all of the PCR primer-binding region with riboadenosine nucleosides. This modification leaves only the total length and the IGS for ribozyme–substrate binding intact, while preventing any spurious binding of the 5′ end of the substrate to the ribozyme during transcription, other than at the IGS. Because of the truncated promoter sequence this substrate could not be used in continuous evolution. Yet in high Mg^{2+} concentrations the E_{100}(#3) sequence more than doubles its total percent ligated product, and now the U1 conformer can ligate to form to R1 when the poly-A-tailed substrate is used, in contrast to the normal substrate. At low Mg^{2+}, however, this conformer fails to ligate even with the poly-A substrate (data not shown). The B_{16}(#19) sequence is negatively affected by the poly-A substrate in both Mg^{2+} concentrations, exhibiting a fourfold reduction in fraction ligated in high Mg^{2+} and a 35% reduction in low Mg^{2+}. These results further confirm that the substrate is partially responsible for the misfolding of the E_{100}(#3) sequence during continuous evolution and that the inhibition intensifies as the Mg^{2+} concentration is lowered.

On the basis of our results, we can describe the phenotypic expression during continuous evolution (Figure 7). Under the prevailing environmental conditions, the evolving pool of ribozymes forms multiple conformers with varying degrees of reactivity. Some conformers may be trapped in inactive states, and substrate binding can influence this. From the relative proportions of bands observed in experiments such as Figure 6, the reactivities of each conformer under the conditions of continuous evolution can be estimated. Here, reactivity refers to the extent of reaction — the percentage of a conformer population that

has ligated to form product. The mean phenotype, $\bar{z}$, of a population of sequences, or of a single sequence that can form multiple conformations, is calculated as:

$$\bar{z} = \sum_i f(C_i) z(C_i) \quad (1)$$

where $f(C_i)$ is the frequency of conformer i and $z(C_i)$ is the phenotypic value (reactivity) of conformer i. For the ligation under the conditions of continuous evolution in 25 mM total Mg^{2+}, the phenotype of the E_{100}(#3) sequence is estimated from the frequencies and reactivities given in Table 2 (i = 0, 1, 2, 3, 4) as $\bar{z}$ = (0.09)(0) + (0.48)(0) + (0.21)(0.21) + (0.16)(0.22) + (0.06)(0.58) = 0.11. This value agrees exactly with the 11% observed ligated value derived from the denaturing-gel quantitation (Table 2). Similarly, $\bar{z}$ for B_{16}(#19) in 25 mM total Mg^{2+} is estimated (i = 1, 2, 3, 4) as $\bar{z}$ = (0.80)(0.055) + (0.11)(0.25) + (0.12)(0.053) + (0.029)(1) = 0.11, an equivalent phenotype to E_{100}(#3) under these conditions. In 15 mM total Mg^{2+}, however, the ligation phenotypes are estimated in the same manner as 0.057 and 0.082 for E_{100}(#3) and B_{16}(#19), respectively, reflecting the greater relative fitnesses of the B_{16}(#19) sequence in low Mg^{2+}. Using errors obtained from replicate experiments, the superiority of B_{16}(#19) in low Mg^{2+} is significant (t-test, $P < 0.05$, n = 4). This difference stems from the inactivity of E_{100}(#3) conformers U0, U1, and U2 in low Mg^{2+}, contrasted with the retention of activity of B_{16}(#19) conformers U1–U3. The result of our selection of RNA ligases that function in progressively lower Mg^{2+} was not the advent of a ribozyme that is necessarily faster in a kinetic sense, but the advent of one that is better able to survive by overcoming a tendency to form inactive conformers.

The phenotypic variation (number of conformers) exhibited by the ligase ribozyme during continuous evolution is therefore much greater than the available genotypic variation (number of different RNA sequences), even when only single sequences are considered. To estimate the heritability of this molecular trait (ligation) we can calculate the realized heritability, which is the degree to which the population's mean phenotype responds to selection in one generation: $H^2 = R/S$, where R is the response to selection and S is the selection differential [11]. In our case, the response to selection can be estimated as the differences in the mean phenotypes of E_{100}(#3) and B_{16}(#19) in low Mg^{2+} (denaturing-gel values: R = 5.9% – 2.1% = 3.8% = 0.038), a genotypic shift that was completed essentially between bursts #7 and #13. The selection differential is the difference between the population's mean before selection and the threshold value for surviving selection. If we assume that the mean before selection is 2.1% and the threshold value is 100% (a ribozyme must ligate to survive to the next burst), then S = 0.979 and H^2 = 0.038/0.979 = 0.04 over the six-burst period in the middle of the continuous-evolution experiment. This is a very low value that explains in part why little shift in overall ligation phenotype was observed after 16 bursts. One should keep in mind, however, that this estimate was based on the shift that occurred during the evolution of a population with very little genetic variation (B_7–B_{13}), and that heritability was probably much higher at the beginning of the experiment. Consequently, it must be stressed that this is an estimate of the heritability of the evolving population, not of a specific sequence. Heritability is a trait- and environment-dependent parameter and cannot be extrapolated from one system to another. Nevertheless, it should be relatively straightforward to estimate heritabilities in other *in vitro* selection experiments; H^2 for traits under selection in stepwise protocols such as SELEX are expected to be much greater than 0.04, as evidenced by the large phenotypic gains realized in small numbers of generations. In summary, we can treat a nucleic acid's phenotype as a quantitative trait under the influence of many 'genes', which here are individual nucleotides that interact in an epigenetic-like fashion.

Significance

We have studied the phenotypic variation associated with single RNA sequences *in vitro* and have described how this variation confounds the progress of *in vitro* selection. The ribozyme derived from the class I ligase of Bartel and Szostak [18] can adopt multiple conformations that each have different ligation activities under the conditions of continuous evolution. Like organismal phenotypes, each RNA sequence exhibits its own particular norm of reaction such that the fitnesses of sequences will depend on the environmental conditions under which selection occurs. Using the continuous system, we have demonstrated that selection can operate to favor sequences that are more likely to form chemically reactive conformers. During evolution, our RNA population exhibits heritability values estimated at less than 10%, indicative of a proportion of molecules folding into weakly reactive or inactive conformations —phenotypes that are not 'heritable' from one generation to the next.

Besides the consequences that non-unity heritability can have for *in vitro* selection systems, this phenomenon could affect molecular evolution in organismal systems. Stable RNA molecules, or even proteins, that adopt multiple folded conformations under physiological conditions may have a selective disadvantage, further reducing the rate at which favorable mutations arise or are driven to fixation in a population. A recent review of protein misfolding stressed that cellular mechanisms have evolved to ensure that a single well-defined polypeptide fold exists in solution, indicating that strong selection may exist in natural systems against sequences that are less likely to fold correctly [29]. Conceivably, genotypic polymorphisms could persist in a population for longer if alleles were not perfectly heritable. Conversely, in certain circumstances this phenomenon could actually result in a

selective advantage, by allowing gene products the ability to perform multiple functions in alternate environments. In fact, the origin of life on Earth has been postulated to have proceeded through a dual-functional RNA replicase that must be both a RNA-dependent RNA polymerase and a template for replication [18]. Alternating between folding conformations may have allowed primordial ribozymes such duality.

Materials and methods

Materials

The E_{100}(#3) ribozyme, was a generous gift from M. Wright. Salts and pH buffers were purchased from Sigma in the highest purity grade available. Urea, acrylamide (acrylamide:bis-acrylamide::19:1) and ribonucleotide triphosphates were purchased from Boeringher Mannheim. M-MLV reverse transcriptase, deoxynucleotide triphosphates, [α^{32}P]ATP, [γ^{32}P]ATP, [^{35}S]dATP and Sequenase 2.0 sequencing kits were purchased from Amersham Pharmacia Biotech. T7 RNA polymerase was purchased from Ambion at 200 U/µl. PCR primers (primer 1 was also the primer for reverse transcription, 5′-GCTGAGCCTGCGATTGG-3′; primer 2, 5′-CTTGACGTCAGCCTGGA-3′) were purchased from Operon. Substrate chimeric DNA/RNAs (standard = 5′-CTTGACGTCAGCCTGGAC*TAATACGACTCA**UAUA***-3′; with the T7 promoter sequence italicized and the ribonucleotides shown in bold; poly-A-tailed = 5′-AAAAAAAAAAAAAAAAAAAAAAAAAA***UCAUAUA***-3′) were purchased from Oligos, Etc. Substrates were gel-purified on 15% polyacrylamide/8 M urea gels and desalted with Nensorb 50 (NEN Life Sciences) columns prior to use.

Continuous evolution

A random pool of RNA sequences was generated by an 8% per-position randomization of 112 nucleotides of the E_{100}(#3) sequence using the mutator oligonucleotide approach of Tsang and Joyce [30]. A single round of 'rapid' stepwise selection for catalysis in 25 mM $MgCl_2$ for 30 min was performed with 1 pmol of randomized RNA (6×10^{11} molecules) as described previously [16] to purge the population of completely inactive sequences. The resulting RNA was reverse transcribed, amplified by PCR using primers 1 and 2, which regenerate the T7 promoter sequence, and forward transcribed to generate ~100 pmol of B_0 material that could be used to seed continuous evolution. Continuous evolution was carried out using the protocols described in [16]. For the first burst, 1 pmol of RNA was incubated with 64 pmol standard substrate, 50 pmol primer 1, 250 U M-MLV reverse transcriptase, 50 U T7 RNA polymerase, 5 nmol each dNTP, 50 nmol each rNTP, and 3.75 µCi [α^{32}P]ATP, and 25 mM $MgCl_2$ in reaction buffer (50 mM KCl, 30 mM 4-(2-hydroxyethyl)piperazine-1-propanesulfonic acid = EPPS, pH 8.5) in a total volume of 25 µl for 40 min at 37°C. The reaction was performed in a 37°C constant-temperature room to ensure that all materials were pre-warmed before reactions were initiated. At the end of the incubation period, 3 µl were removed, quenched in acrylamide gel-loading buffer (0.05% bromphenol blue, 40% sucrose) and saved on ice for subsequent electrophoresis. Also, an additional 3 µl were removed, and diluted into 981 µl of water. A 8.2 µl aliquot of this dilution was used to seed the next 25 µl burst, resulting in an overall 1000-fold dilution from one burst to the next. In the second, and all subsequent bursts, the diluted mixture from the previous burst was incubated with fresh amounts of substrate, primer, protein enzymes, nucleotides and buffer in the quantities described above. The $MgCl_2$ concentrations were lowered by 2.5 mM after every third burst as diagrammed in Figure 3. The incubation times were shortened during the second and third bursts at each $MgCl_2$ concentration as follows (burst #, time in minutes): (2, 30), (3, 25), (4, 30), (5 and 6, 25), (7, 30), (8 and 9, 25), (10, 30), (11 and 12, 25), (13, 40), (14-16, 30), (17 and 18, 30) and (19 and 20, 25). The quenched aliquots resulting from all 20 bursts were electrophoresed through 5% polyacrylamide/8 M urea gels in 1× TBE to monitor the presence of body-labeled RNA and the overall fraction ligated. Gels were visualized by phosphorimaging (Storm 860, Molecular Dynamics). 1 µl of the water-diluted end products after bursts #7, 14 and 16 were amplified via the PCR in 100 µl reaction volumes using primer 1 and primer 2 (30 cycles, 55°C annealing). The resulting single-sized PCR products were immediately ligated into pCR 2.1 vector (Invitrogen) and cloned into *E. coli* strain INV-αF′. Single bacterial colonies were chosen for 'miniprep' plasmid preparation by the boiling method. The nucleotide sequences of ribozyme inserts were determined by manual bi-directional sequencing with [^{35}S]dATP and Sequenase 2.0.

Kinetic assays

Homogeneous RNAs were transcribed *in vitro* from PCR-amplified miniprep DNA, whereas heterogeneous pool RNAs were transcribed from PCR products generated at the end of continuous evolution bursts. Transcription was performed by standard methods [31] in the presence (simple percent-ligated assays) or absence (formal kinetic assays) of [α^{32}P]ATP; the resultant RNAs were electrophoresed through a 5% polyacrylamide/8 M urea gel, the full-length band was excised with a sterile razor blade, and the gel slice was chopped into 1–2 mm^3 cubes and shaken 4–12 h in elution solution (200 mM NaCl, 10 mM Tris, pH 7.5, 0.5 mM EDTA) at room temperature. For simple percent-ligated assays, the eluted RNA was desalted with Nensorb 50 columns and rehydrated in 0.1 mM EDTA. For a subset of these reactions, the RNA was heated to 80°C for 2 min, and allowed to cool to room temperature in the presence of reaction buffer. These assays were performed by incubating radiolabeled RNA (100 nM) with unlabeled substrate (500 nM) in reaction buffer containing 10–25 mM $MgCl_2$ for 30 min at 37°C. 3 µl aliquots were taken at various time points, quenched with gel-loading buffer and electrophoresed through 5% polyacrylamide/8 M urea gels. Care was taken to load the same sample volume on each lane of the gels. Products were visualized by phosphorimaging and extent of reaction was quantitated as percent of total RNA in the ligated form. Reactions were performed at least three times and the values reported are arithmetic averages. For analysis of single conformers, [α^{32}P]ATP-labeled ribozymes were preparatively transcribed and the products were run on native 5% polyacrylamide gels lacking urea. The bands corresponding to specific conformers (e.g. U1) were excised, electroeluted for 1 h at 1250 V-cm, and ethanol precipitated before assays as described above. For formal kinetic assays, the eluted RNA (unlabeled) from the gel cubes was desalted by ethanol precipitation [32] and rehydrated in 0.1 mM EDTA. The k_{cat} parameter was estimated for both the E_{100}(#3) and B_{16}(#19) sequences by the *y*-intercept of modified Eadie–Hofstee plots of the observed rate of reaction (k_{obs}) as a function of k_{obs}/[ribozyme]. The k_{obs} values of the ligation reaction were determined under ribozyme excess conditions of 5 nM 5′-[γ^{32}P]-labeled substrate and 0.05–5.0 µM ribozyme at 37°C. Six timepoints ranging between 5 s and 2 min were taken when the ribozymes were incubated in continuous-evolution buffer containing either 10 mM or 25 mM $MgCl_2$ and 30 mM EPPS (pH 8.5 or 6.0). The extent of reaction was quantitated as fraction of total RNA in the ligated form in 5% denaturing polyacrylamide gels as in the simple assays above. Single-turnover self-ligation k_{obs} values were obtained for each ribozyme concentration by an exponential curve fit to the equation $f = A(1\text{-}\exp(-k_{obs}*t))$, where f is the fraction of substrate reacted at time t, and A is the asymptote (projected maximum f). Only k_{obs} values from individual time series where the half-time of the reaction was 9 s or greater were utilized.

Transcription assays

Transcription products were examined under simulated continuous evolution conditions by incubating approximately 10 ng of double-stranded DNA containing the template for either the E_{100}(#3) or B_{16}(#19) ribozymes with 64 pmol standard or poly-A-tailed substrate, 50 pmol primer 1, 50 U T7 RNA polymerase, 5 nmol each dNTP, 50 nmol each rNTP and 3.75 µCi [α^{32}P]ATP, and 25 mM $MgCl_2$ (high Mg^{2+} concentration) or 15 mM $MgCl_2$ (low Mg^{2+} concentration) in reaction buffer (50 mM KCl, 30 mM EPPS, pH 8.5) in a total volume of 25 µl. The 30 regime involved incubation as such for 30 min at 37°C. The 29-1 regime involved incubation as such but without substrate for 29 min, at which

time 64 pmol substrate was added and the reaction was allowed to incubate for an additional 60 s. For assays with radiolabeled substrate, the [α^{32}P]ATP was left out of the reaction and instead 63 pmol of unlabeled substrate doped with 1 pmol of 5′-[γ^{32}P]ATP-labeled substrate was used. After 30 min, all reactions were quenched with EDTA in stoichiometric amounts to the Mg^{2+} concentration used, chilled on ice and aliquots immediately loaded onto both native and denaturing 5% polyacrylamide gels. Reaction products and percent ligated fractions were quantitated by phosphorimaging.

Acknowledgements

We are extremely grateful to M. Wright and G. Joyce for providing details of the continuous evolution scheme and for providing the E_{100}(#3) molecule. We also appreciate the efforts of P. Ordoukhanian and D. Bartel for advice on kinetic methodology, G. Joyce for suggestions on the manuscript, and D. Decker, X.-C. Dai, and D. Edgell for valuable discussions. This work was supported in part by the University at Albany, SUNY.

References

1. Joyce, G.F. (1989). RNA evolution and the origins of life. *Nature* **338**, 217-224.
2. Joyce, G.F. (1998). Nucleic acid enzymes: playing with a fuller deck. *Proc. Natl Acad. Sci. USA* **95**, 5845-5847.
3. Uhlenbeck, O.C. (1995). Keeping RNA happy. *RNA* **1**, 1-4.
4. Walstrum, S.A. & Uhlenbeck, O.C. (1990). The self-splicing RNA of *Tetrahymena* is trapped in a less active conformation by gel purification. *Biochemistry* **29**, 10573-10576.
5. Been, M.D., Perrotta, A.T. & Rosenstein S. (1992). Secondary structure of the self-cleaving RNA of hepatitis delta virus: application to catalytic RNA design. *Biochemistry* **31**, 11843-11852.
6. Smith, J.B., Gottlieb, P.A. & Dinter-Gottlieb, G. (1992). A sequence element necessary for self-cleavage of the antigenomic hepatitis delta RNA in 20 M formamide. *Biochemistry* **31**, 9629-9635.
7. Gottlieb, P.A., Prasad, Y., Smith, J.B., Williams, A.P. & Dinter-Gottlieb, G. (1994). Evidence that alternate foldings of the hepatitis δ RNA confer varying rates of self-cleavage. *Biochemistry* **33**, 802-2808.
8. Burke, D.H. & Willis, J.H. (1998). Recombination, RNA evolution, and bifunctional RNA molecules isolated through chimeric SELEX. *RNA* **4**, 1165-1175.
9. Bonhoeffer, S., McCaskill, J.S., Stadler, P.F. & Schuster, P. (1993). RNA multistructure landscapes. *Eur. Biophys. J.* **22**, 13-24.
10. Esteban, J.A., Raj Banerjee, A. & Burke, J.M. (1997). Kinetic mechanism of the hairpin ribozyme. *J. Biol. Chem.* **272**, 13629-13639.
11. Lynch, M. & Walsh, B. (1998). *Genetics and Analysis of Quantitative Traits*. Sinauer, Sunderland, MA.
12. Fisher, R.A. (1930). *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford.
13. Joyce, G.F. (1992). Directed molecular evolution. *Sci. Amer.* **267**, 90-97.
14. Breaker, R.R. & Joyce, G.F. (1994). Inventing and improving ribozyme function: rational design versus iterative selection methods. *Trends Biotechnol.* **12**, 268-275.
15. Kumar, P.K.R. & Ellington, A.D. (1995). Artificial evolution and natural ribozymes. *FASEB J.* **9**, 1183-1195.
16. Wright, M.C. & Joyce, G.F. (1997). Continuous *in vitro* evolution of catalytic function. *Science* **276**, 614-617.
17. Breaker, R.R., Banerji, A. & Joyce, G.F. (1994). Continuous *in vitro* evolution of bacteriophage RNA polymerase promotors. *Biochemistry* **33**, 11980-11986.
18. Bartel, D.P. & Szostak, J.W. (1993). Isolation of new ribozymes from a large pool of random sequences. *Science* **261**, 1411-1418.
19. Ekland, E.H., Szostak, J.W. & Bartel, D.P. (1995). Structurally complex and highly active RNA ligases derived from random RNA sequences. *Science* **269**, 364-370.
20. Ekland, E.H. & Bartel, D.P. (1995). The secondary structure and sequence optimization of an RNA ligase ribozyme. *Nucleic Acids Res.* **23**, 3231-3238.
21. Celander, D.W. & Cech, T.R. (1991). Visualizing the higher order folding of a catalytic RNA molecule. *Science* **251**, 401-407.
22. Lehman, N. & Joyce, G.F. (1993). Evolution *in vitro* of an RNA enzyme with altered metal dependence. *Nature* **361**, 182-185.
23. Lehman, N. & Joyce, G.F. (1993). *In vitro* evolution: analysis of a lineage of ribozymes. *Curr. Biol.* **3**, 723-734.
24. Geyer, C.R. & Sen, D. (1997). Evidence for the metal-cofactor independence of an RNA phosphodiester-cleaving DNA enzyme. *Chem. Biol.* **4**, 579-593.
25. Faulhammer, D. & Famulok, M. (1997). Characterization and divalent metal-ion dependence of *in vitro* selected deoxyribozymes which cleave DNA/RNA chimeric oligonucleotides. *J. Mol. Biol.* **269**, 188-202.
26. Weeks, K.M. & Cech, T.R. (1995). Efficient protein-facilitated splicing of the yeast mitochondrial bI5 intron. *Biochemistry* **34**, 7728-7738.
27. Sabeti, P.C., Unrau, P.J. & Bartel, D.P. (1997). Accessing rare activities from random RNA sequences: the importance of the length of molecules in the starting pool. *Chem. Biol.* **4**, 767-774.
28. Repsilber, D., Wiese, S., Rachen, M., Schroder, A.W., Reisner, D. & Steger, G. (1999). Formation of metastable RNA structures by sequential folding during transcription: time-resolved structural analysis of potato spindle tuber viroid (–)-stranded RNA by temperature-gradient electrophoresis. *RNA* **5**, 574-584.
29. Dobson, C.M. (1999). Protein misfolding, evolution and disease. *Trends Biochem. Sci.* **24**, 329-332.
30. Tsang, J. & Joyce, G.F. (1996). *In vitro* evolution of randomized ribozymes. *Methods Enzymol.* **267**, 410-426.
31. Beaudry, A.A. & Joyce, G.F. (1992). Directed evolution of an RNA enzyme. *Science* **257**, 635-641.
32. Herschlag, D. & Cech, T.R. (1990). Catalysis of RNA cleavage by the *Tetrahymena thermophila* ribozyme. 1. Kinetic description of the reaction of an RNA substrate complementary to the active site. *Biochemistry* **29**, 10159-10171.