nature structural & molecular biology

New catalytic structures from an existing ribozyme

Edward A Curtis & David P Bartel

Although protein enzymes with new catalytic activities can arise from existing scaffolds, less is known about the origin of ribozymes with new activities. Furthermore, mechanisms by which new macromolecular folds arise are not well characterized for either protein or RNA. Here we investigate how readily ribozymes with new catalytic activities and folds can arise from an existing ribozyme scaffold. Using *in vitro* selection, we isolated 23 distinct kinase ribozymes from a pool of sequence variants of an aminoacylase parent ribozyme. Analysis of these new kinases showed that ribozymes with new folds and biochemical activities can be found within a short mutational distance of a given ribozyme. However, the probability of finding such ribozymes increases considerably as the mutational distance from the parental ribozyme increases, indicating a need to escape the fold of the parent.

Since 1958, when the structure of myoglobin was determined at atomic level resolution[1], the folds of thousands of protein enzymes have been characterized, providing a wealth of information about protein structure, mechanism and evolution[2,3]. From an evolutionary standpoint, several of these studies suggest that protein enzymes with new biochemical activities can arise from existing protein scaffolds[4,5]. For instance, enzymes in the α/β hydrolase superfamily catalyze a wide range of reactions and include an acetylcholinesterase, a semialdehyde dehalogenase and a haloalkane dehalogenase[5,6]. Similarities in the global folds of these enzymes, local details of their structures and the positioning and identity of catalytic residues suggest that at least some members of this superfamily were derived from a common ancestor, perhaps after a gene-duplication event[7]. In some cases, such a transformation has been directly observed in the laboratory. For example, a single amino acid change is sufficient to change sheep blowfly carboxylesterase into an organophosphorus hydrolase[8]. It has also been noted that protein enzymes with particular activities sometimes catalyze unrelated reactions at low levels, suggesting a possible starting point for the evolution of new catalysts[9]. Furthermore, with few exceptions[10,11], proteins with new catalytic functions or binding properties generated using *in vitro* evolution have been isolated in the context of existing scaffolds (reviewed in ref. 12).

Comparatively less is known about the origin of ribozymes with new activities, although it has been suggested that starting with a ribozyme scaffold might provide an advantage when isolating new ribozymes by *in vitro* selection[13,14]. Furthermore, the origin of new macromolecular folds is not well understood for either protein[15] or RNA. Here, the technique of *in vitro* selection was used to investigate how readily ribozymes with new activities could arise from an previously isolated aminoacylase ribozyme[16,17] and whether the folds of such ribozymes would be new as well. Our results indicate that ribozymes with new biochemical activities and folds can be found within a short mutational distance of a given ribozyme, but that the probability of finding such ribozymes increases considerably as the mutational distance from the starting ribozyme increases.

RESULTS

Kinase ribozymes from an aminoacylase ribozyme

Our first objective was to determine whether ribozymes with new catalytic activities could be found in the sequence neighborhood of an existing parent ribozyme. As a starting point for our experiments we chose a previously isolated and well-characterized self-aminoacylating ribozyme called isolate 77 (refs. 16,17). This ribozyme aminoacylates its 3′ terminus using adenylated phenylalanine as a substrate (**Fig. 1a**) and, under optimized conditions, catalyzes this reaction with a k_{cat}/K_m of 6×10^4 M^{-1} min^{-1} and a second-order rate enhancement of 6×10^6-fold (relative to the nonenzymatic hydrolysis rate of adenylated phenylalanine[18]). To generate variants of the parent, RNA was transcribed from a partially degenerate DNA template in which 65 of the 90 positions encoding this ribozyme were partially randomized at an average rate of ~11% per position; in other words, at each partially randomized position, the parental base was present at a frequency of 0.89, and each of the other three bases was present at a frequency of ~0.04.

Several considerations influenced the design of this pool. First, we wanted most of the sequences in the pool to be similar to the parent ribozyme. About 95% of the unique sequences in a pool mutagenized at this level will be within 12 mutations of the parent ribozyme, and essentially every possible sequence within 8 mutations of the parent will be present at least once[19]. At the same time, we wanted our pool to be diverse enough to contain new ribozymes. At this level of mutagenesis, the pool of 2×10^{15} double-stranded DNA templates contained about 4×10^{14} unique sequences; this number would probably be sufficient if kinases were being selected from a randomized nucleic acid pool[20,21]. Finally, we wanted our pool to be enriched for ribozymes that are

Whitehead Institute for Biomedical Research and Department of Biology, Massachusetts Institute of Technology, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA. Correspondence should be addressed to D.P.B. (dbartel@wi.mit.edu).

Received 20 May; accepted 2 September; published online 9 October 2005; doi:10.1038/nsmb1003

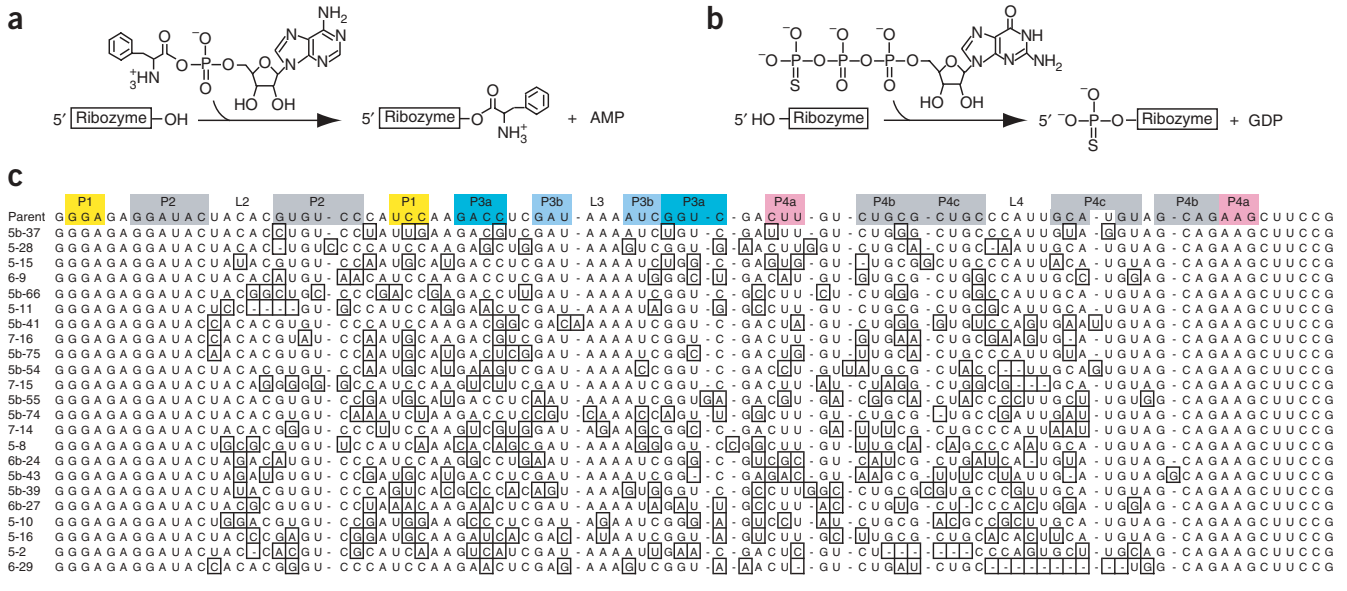

Figure 1 Aminoacylase and kinase ribozymes. (**a**) Reaction catalyzed by the parent aminoacylase ribozyme. (**b**) Reaction catalyzed by a 5′-kinase ribozyme. Note that our selection scheme did not require ribozymes to specifically thiophosphorylate their 5′-hydroxyl group, and about half of the ribozymes characterized modified an internal 2′-hydroxyl group. (**c**) Sequence alignment of 23 kinase ribozymes isolated in this selection. Boxes, positions that differ from the aminoacylase parent ribozyme; colored bars, parts of the sequence that form base pairs in the parent. Sequences are shown in order of mutational distance from the parent. (**d**) Ribozyme 6-9 is a 5′-kinase. Ribozyme 6-9 was reacted with GTPγS-γ^{35}S and digested with P1 nuclease. The resulting 5′ nucleotide monophosphates were analyzed by polyethyleneimine cellulose thin-layer chromatography (PEI-TLC). (**e**) Ribozyme 5-16 is an internal kinase. Reacted 5-16 was radiolabeled at its 5′ terminus, partially hydrolyzed with base and analyzed on an APM polyacrylamide sequencing gel. (**f**) Ribozyme 5-16 transfers a thiophosphate from GTPγS to itself. An optimized version of 5-16 (data not shown) was reacted with GTPγS-γ^{35}S, purified and incubated in the presence of different leaving groups that might be produced in the forward reaction of this ribozyme with GTPγS. Reaction aliquots were analyzed by PEI-TLC.

especially close to the parent. In a pool designed in this manner, sequences closer to the parent are present at higher initial copy numbers than are sequences farther away. For our pool, this was true for sequences within eight mutations of the parent, whereas sequences farther away were probably present as single copies, if at all. Thus, a ribozyme within eight mutations of the parent ribozyme was expected to be over-represented in this pool relative to a ribozyme farther from the parent.

An *in vitro* selection protocol was used to isolate rare individuals from this pool that thiophosphorylate themselves in the presence of the GTP analog GTPγS (**Fig. 1b**). We were interested in the transition from an aminoacylase ribozyme to kinase ribozymes for several reasons. First, phosphorylation is chemically distinct from aminoacylation, with a transition state that is trigonal bipyramidal rather than tetrahedral. Furthermore, GTPγS differs from the adenylated phenylalanine substrate used by the parent ribozyme in several important respects, including charge and hydrophobicity, and the nucleotide base of these substrates is also different. By selecting for variants of the parent ribozyme that catalyze a reaction with a different transition state and that recognize a different substrate, we hoped to obtain ribozymes whose active sites were different from that of the parent ribozyme. Selecting for kinase ribozymes seemed reasonable, because it had earlier been shown that RNA can readily catalyze this type of transformation and that kinase ribozymes can be selected using ATPγS as a substrate[20]. Finally, phosphorylation and aminoacylation are both of considerable biological importance.

To isolate kinase ribozymes, pool RNA was incubated with GTPγS, and molecules that became thiophosphorylated during the incubation were purified on *N*-acryloylaminophenylmercuric chloride (APM) polyacrylamide gels[22,23], amplified by RT-PCR and transcribed to generate RNA for the next round of selection. After four rounds of selection, pool activity was detected, and two additional rounds of selection were conducted using progressively shorter incubation times and a lower concentration of GTPγS. The 112 clones that were sequenced from rounds 4, 5 and 6 were grouped into classes based on several criteria. Initially, a phylogenetic tree was used to identify clusters of closely related or identical sequences. Sequences within four mutations of one another were assigned to the same class. Of the 52 classes identified in this way, 41 were represented by a single unique sequence, whereas 11 contained between 2 and 28 similar or identical sequences. At least one member of each class was then tested for catalytic activity. Active members were detected in 29 of 52 classes. Further examination revealed several sequences that were more than four mutations from one another but that nevertheless shared more mutations in common than would be expected by chance in a data set of this size. For example, the sequences of 6b-12 and 6b-27 differed at 13 positions, yet they had eight mutations in common. Five other sequence pairs with six or more mutations in common were assigned to the same class, producing a total of 23 unique classes with detectable activity (**Fig. 1c**).

At 1 mM GTPγS, initial rates of these ribozymes ranged between 8×10^{-6} and 6×10^{-4} min^{-1}, and the extent of the self-

thiophosphorylation reaction was typically between 1% and 60% in a 24-h incubation. k_{cat}/K_m values were determined for 12 of these ribozymes and ranged between 0.1 and 2 M^{-1} min^{-1}. Second-order rate enhancements were estimated by comparing ribozyme k_{cat}/K_m values with the nonenzymatic rate of GTPγS hydrolysis and ranged between 3×10^5-fold and 6×10^6-fold.

Ribozymes catalyzing at least three types of reactions were represented among the 23 kinase classes isolated in this selection. Of 12 ribozymes characterized, 5 (5-10, 5-15, 5-28, 5b-43 and 6-9) transferred a thiophosphate group from GTPγS to their 5′-hydroxyl group. These ribozymes could not be radiolabeled using T4 polynucleotide kinase after they reacted with GTPγS unless they were first treated with alkaline phosphatase, suggesting that their 5′-hydroxyl group becomes covalently modified during catalysis. P1 digests of GTPγS-γ^{35}S-labeled ribozymes generated a labeled product that comigrated with a guanosine-5′-monothiophosphate marker, indicating that this modification at the 5′-hydroxyl was a thiophosphate (**Fig. 1d**).

Of 12 ribozymes characterized, 5 used an internal 2′-hydroxyl group as a nucleophile. Analysis of reacted ribozymes on APM polyacrylamide sequencing gels after partial base hydrolysis indicated that these ribozymes modify themselves at internal sites, as exemplified by analysis of 5-16 (**Fig. 1e**). For this ribozyme, the ladders of the hydrolysis fragments from reacted and unreacted RNA comigrated until A32, but from U33 on they did not, indicating that U33 contained a thiol modification. Ribozymes 5-8, 5-11, 7-15 and 7-16 also modified themselves at internal sites (G82, A79, G17 and C62, respectively). In each case, the phosphodiester linkage at the modified position was resistant to base hydrolysis, suggesting that the corresponding 2′-hydroxyl group became modified during catalysis. Using an optimized version of one of these 2′-kinase ribozymes (data not shown), we further characterized the nature of this modification by incubating the reacted ribozyme with potential leaving groups in an attempt to regenerate the original substrate, presumably through the reverse reaction. Reacted ribozyme generated GTPγS in the presence of GDP but not in the presence of GMP, guanosine or buffer alone (**Fig. 1f**). This suggested that GDP was the leaving group in the forward reaction catalyzed by this ribozyme and, therefore, that a single thiophosphate group was transferred to the ribozyme during catalysis.

The two remaining ribozymes characterized (5-2 and 7-14) seemed to promote a two-step reaction in which both the 5′-hydroxyl group and an internal site become modified during catalysis. The mechanisms of these two ribozymes were not further characterized.

The kinase reactions differ from the reaction catalyzed by the parent ribozyme in several respects. First, the characterized kinase ribozymes used either 5′ or internal 2′-hydroxyl groups as nucleophiles, whereas the parent ribozyme uses a hydroxyl group at its 3′ terminus as a nucleophile[17]. Second, these ribozymes break and form bonds between phosphorus and oxygen, whereas the parent ribozyme breaks and forms bonds between carbon and oxygen. Third, these ribozymes probably promote a reaction involving a trigonal bipyramidal transition state, whereas the parent ribozyme probably promotes one involving a tetrahedral transition state. Thus, ribozymes with new catalytic activities could be found in the sequence neighborhood of an existing parent ribozyme.

Kinase density increases with mutational distance

Next, we set out to determine how readily such ribozymes arise from an existing ribozyme scaffold. We sequenced 67 clones from the starting pool used in this experiment and determined the mutational distance of each sequence from the parent ribozyme. Comparison to the distribution obtained from the 23 classes of kinase ribozymes indicated that kinases were not uniformly distributed in sequence space with respect to the parent (**Fig. 2a**). Instead, as the mutational distance from the parent ribozyme increased, the probability of finding a kinase ribozyme increased markedly, at least over the range of mutational distances examined. In the range between 10 and 16 mutations from the parent, the abundance of sequences in the starting pool decreased as the mutational distance from the parent increased (**Fig. 2a**). This was expected: for a mutagenesis rate of 11% per position and 65 mutagenized positions, the starting pool should contain ~60-fold more unique sequences located 10 mutations from the parent than 16 mutations from the parent. Despite this expectation, the number of kinases isolated was approximately constant over this range, implying that the density of kinase ribozymes (ribozyme density = number of ribozymes that are x point substitutions from the parent / total number of unique sequences that are x point substitutions from the parent) among sequences located 10 mutations from the parent was considerably lower than that among sequences 16 mutations from the parent.

A small difference between the mutational distances of evolved and unevolved sequences from the parent ribozyme might be expected owing to random mutation during the reverse transcription and PCR steps of the selection. To measure the magnitude of this effect, we determined the average distance between the parent ribozyme and

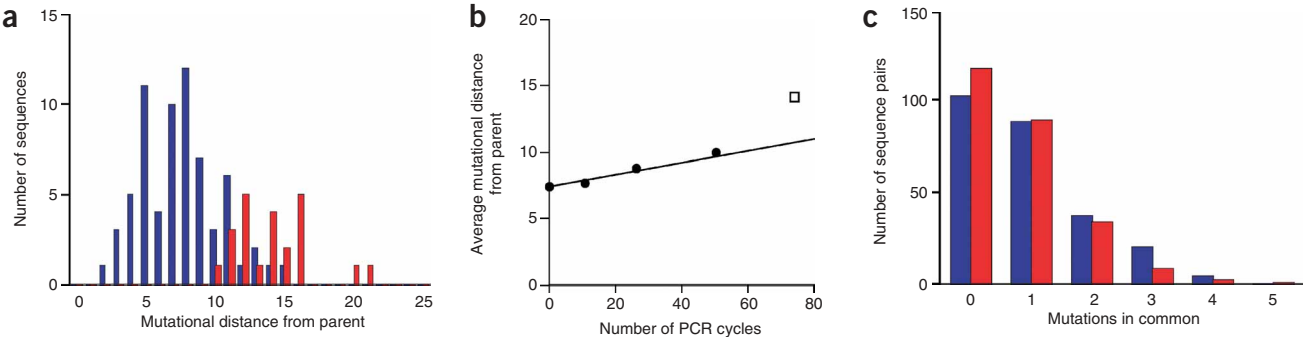

Figure 2 Distribution of kinase ribozymes with respect to the parent ribozyme. (**a**) The distances between 67 sequences from the starting pool and the parent ribozyme (blue bars; average distance = 7.5) compared to the distances between 23 kinase ribozymes and the parent ribozyme (red bars; average distance = 14). (**b**) The average mutational distance of inactive sequences from rounds 0, 1, 2 and 3 (closed circles) and kinase ribozymes (open square) plotted as a function of the number of PCR cycles by which these sequences were amplified. The slope of this line is 0.05 ± 0.01 (best fit ± s.e.). (**c**) Independence of kinase ribozyme classes. The number of mutations in common shared by each possible pair of kinase ribozymes (blue bars) is compared to the distribution that would be expected if these sequences were independent (red bars).

inactive clones from rounds 0, 1, 2 and 3 and from these data calculated an average mutation rate of 0.05 ± 0.01 (best fit ± s.e.) per molecule per PCR cycle during the selection (**Fig. 2b**). This mutation rate was too low to account for the high number of changes observed among the kinase ribozymes isolated in this selection.

In interpreting these distributions, another important consideration is the extent to which the kinase ribozymes analyzed belong to different classes (that is, have different sequence requirements and folds). If these kinase ribozymes were all members of the same class, their distribution with respect to the parent might reflect an idiosyncratic feature of this particular class. On the other hand, if each of these ribozymes was an example of a different kinase motif, then their distribution would reveal more general principles for mapping function within RNA sequence space. The diverse thiophosphorylation sites of these ribozymes suggested that many different motifs were represented among the kinases isolated in this selection. Sequence analysis of these ribozymes also suggested that most or all belong to different classes. For this analysis, we focused on positions at which the same mutation occurred in different kinase ribozymes. The expected number of mutations shared by two ribozymes in different classes (or two independent sequences) derived from the parent depends on the mutational distance between these ribozymes and can be calculated using the binomial distribution (see Methods). Two ribozyme sequences belonging to the same class, on the other hand, are likely to share more mutations in common than would be expected by chance, because ribozymes typically contain nucleotides whose identities are required for catalytic activity. For the 23 classes of kinase ribozymes we analyzed, the number of sequence pairs with different numbers of mutations in common almost perfectly matched that expected if each ribozyme belonged to a different class (**Fig. 2c**). Note that different isolates of the same kinase motif can only be identified by this approach if they arise in exactly the same part of the parental sequence. Nonetheless, this assumption seems reasonable given the structural complexity of the kinase motifs we identified (described later) and the low level (11%) at which the parental sequence was mutagenized.

Another important question was whether the nonuniform distribution of kinase ribozymes in the starting pool was related in some idiosyncratic way to the sequence of the parent ribozyme. The available evidence suggests that this was not the case. In a recent study, aptamers to GMP were isolated from variants of flavin aptamers[24], and in a second recent report, GTP aptamers were selected from a pool of variants of an aptamer to ATP[25]. In both of these studies, the aptamers isolated seem to be farther from the starting sequence than expected on the basis of the composition of the starting pools. This suggests that the distribution of kinase ribozymes we observed reflects something general about the way that RNA functions map onto RNA sequence space.

Our results suggest that ribozymes with new catalytic activities are not uniformly distributed in RNA sequence space with respect to existing ribozymes. Instead, the density of such ribozymes increases substantially as the mutational distance from the starting ribozyme increases. Because studies using RNA-folding algorithms have shown that the probability of finding new RNA secondary structures increases markedly as the mutational distance from a reference sequence increases[26], we suspected that the observed distribution of kinase ribozymes reflected a need to escape the fold of the parent ribozyme. In this view, the folds of most or all of the kinase ribozymes we isolated

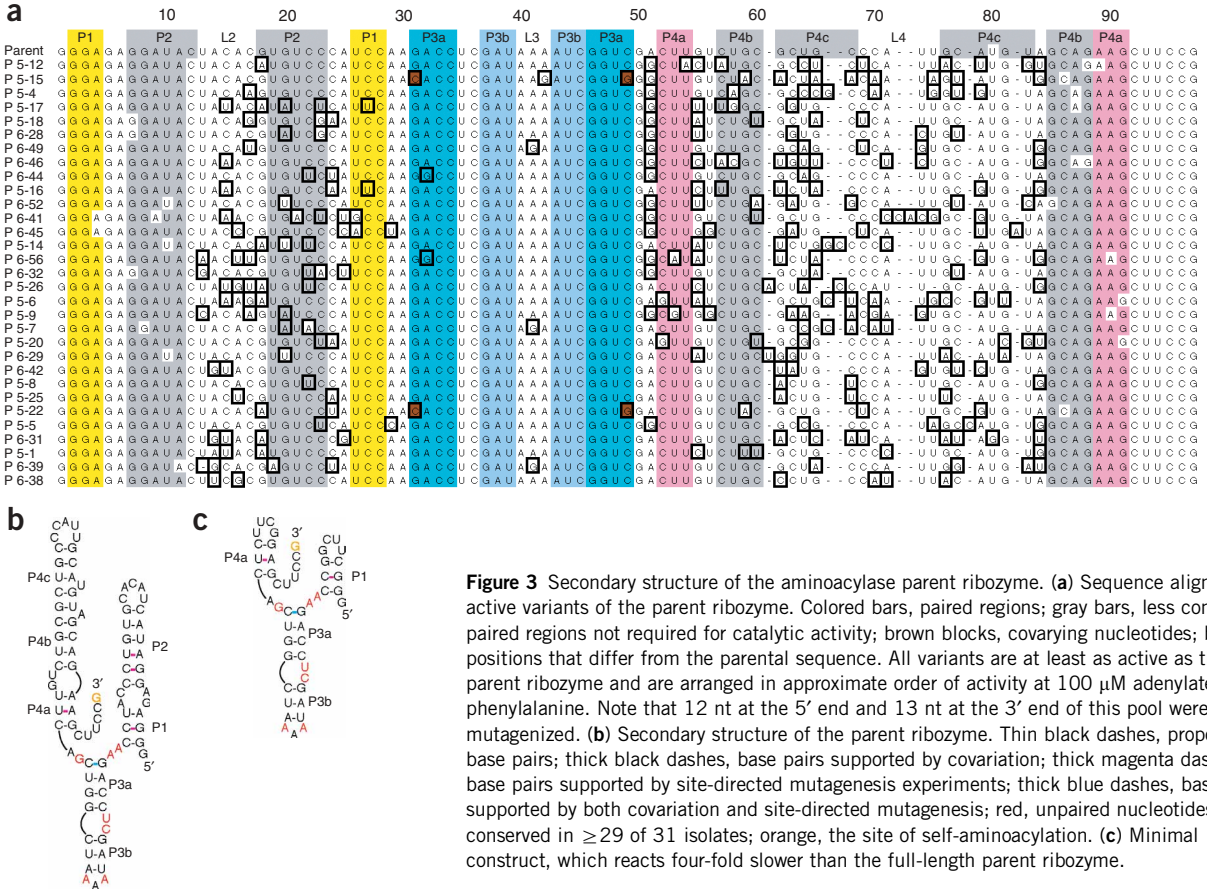

Figure 3 Secondary structure of the aminoacylase parent ribozyme. (**a**) Sequence alignment of active variants of the parent ribozyme. Colored bars, paired regions; gray bars, less conserved paired regions not required for catalytic activity; brown blocks, covarying nucleotides; boxes, positions that differ from the parental sequence. All variants are at least as active as the parent ribozyme and are arranged in approximate order of activity at 100 μM adenylated phenylalanine. Note that 12 nt at the 5′ end and 13 nt at the 3′ end of this pool were not mutagenized. (**b**) Secondary structure of the parent ribozyme. Thin black dashes, proposed base pairs; thick black dashes, base pairs supported by covariation; thick magenta dashes, base pairs supported by site-directed mutagenesis experiments; thick blue dashes, base pairs supported by both covariation and site-directed mutagenesis; red, unpaired nucleotides conserved in ≥29 of 31 isolates; orange, the site of self-aminoacylation. (**c**) Minimal construct, which reacts four-fold slower than the full-length parent ribozyme.

should differ from that of the parent ribozyme. To test this idea, we set out to characterize the folds of the parent and two kinase ribozymes.

An escape to new ribozyme folds

A combination of *in vitro* selection and comparative sequence analysis[27,28] was used to characterize the secondary structures of the parent ribozyme and two kinase ribozymes (called 5-16 and 7-16). For each ribozyme, a pool of 10^{14}–10^{15} variants was synthesized as described above, except that nucleotides were mutagenized at an average rate of 20% per position. To isolate active variants of the parent, pool RNA was incubated with adenylated phenylalanine in selection buffer, and reacted molecules were selectively derivatized with a biotin group using *N*-hydroxysulfosuccinimide biotin (sulfo-NHS-biotin)[29]. Biotinylated RNA molecules were bound to soluble streptavidin, isolated on a polyacrylamide gel[30,31], amplified by RT-PCR and transcribed to generate RNA for the next round of selection. Active variants of kinase 5-16 and kinase 7-16 were isolated as described above.

Once activity was detected in each of these pools, molecules were cloned, sequenced and individually assayed for catalytic activity. Several approaches were then used to derive secondary-structure models for each of these ribozymes. First, each sequence alignment was manually searched for paired regions more highly conserved than would be expected given the level at which the pool was mutagenized. Because this type of analysis does not provide direct evidence for pairing, we also searched for covarying nucleotides in each of the paired regions in the sequence alignment. To better support proposed helices, especially those in which covariation was not observed, mutant ribozymes were synthesized in which proposed base pairs were either disrupted or restored in a compensatory fashion. Finally, to provide additional support for the proposed secondary structures, minimized versions of each ribozyme were synthesized and tested for catalytic activity.

Our secondary-structure model of the parent ribozyme (**Fig. 3a–c**) was similar in some respects to that proposed to be the most populated structure in solution[17], but with several important differences. First, P2 and the distal portion of P4 were only weakly conserved among the sequences isolated in this selection (**Fig. 3a**) and could be deleted from the ribozyme with only a five-fold reduction in catalytic rate (**Supplementary Fig. 1** online). Second, the catalytically active structure of P3 was different from that previously proposed for this part of the sequence. Though not consistent with lead and S1 nuclease probing data[17], this alternative version of P3 was well supported by both comparative sequence analysis (**Fig. 3a**) and site-directed mutagenesis experiments (**Supplementary Fig. 1**). The initial rate of a 50-nucleotide (nt) version of the parent ribozyme in which P2 and P4b were deleted, and in which P1 and P4a were capped with CUUCGG tetraloops[32], was four-fold slower than that of the full-length ribozyme (**Fig. 3c** and **Supplementary Fig. 1**). An even smaller 38-nt variant of the parent, in which P1 was also deleted, was 30-fold slower than the full-length ribozyme (**Supplementary Fig. 1**), indicating that the elements most essential for catalysis are located in P3 and P4a.

The secondary structure model of 5-16 (**Fig. 4a–c**) consisted of a bulged hairpin (P2 and P3) folded back on itself in a pseudoknot (P4), flanked by two less conserved helices (P1 and P5). Comparative sequence analysis also revealed 11 highly conserved unpaired nucleotides, all within the P2-P4 domain of the ribozyme, whose identities were probably important for catalytic activity (**Fig. 4a**). P1 and P5 were generally variable in sequence and length, and they did not always form in exactly the same place in different isolates, unlike helices in the catalytic core of the ribozyme, which were in a constant position (**Fig. 4a**). P1 and P5 could be deleted from the ribozyme without any reduction in catalytic rate (**Fig. 4c** and **Supplementary Fig. 2** online). Disrupting pairings in these helices, however, reduced the catalytic rate of this ribozyme as much as ten-fold, and activity could be restored with compensatory mutations (**Supplementary Fig. 2**). These helices seemed to prevent nucleotides in flanking regions from interfering with those in the P2-P4 domain of the ribozyme. Deleting them completely was tolerated, but disrupting pairings within them was not (**Supplementary Fig. 2**).

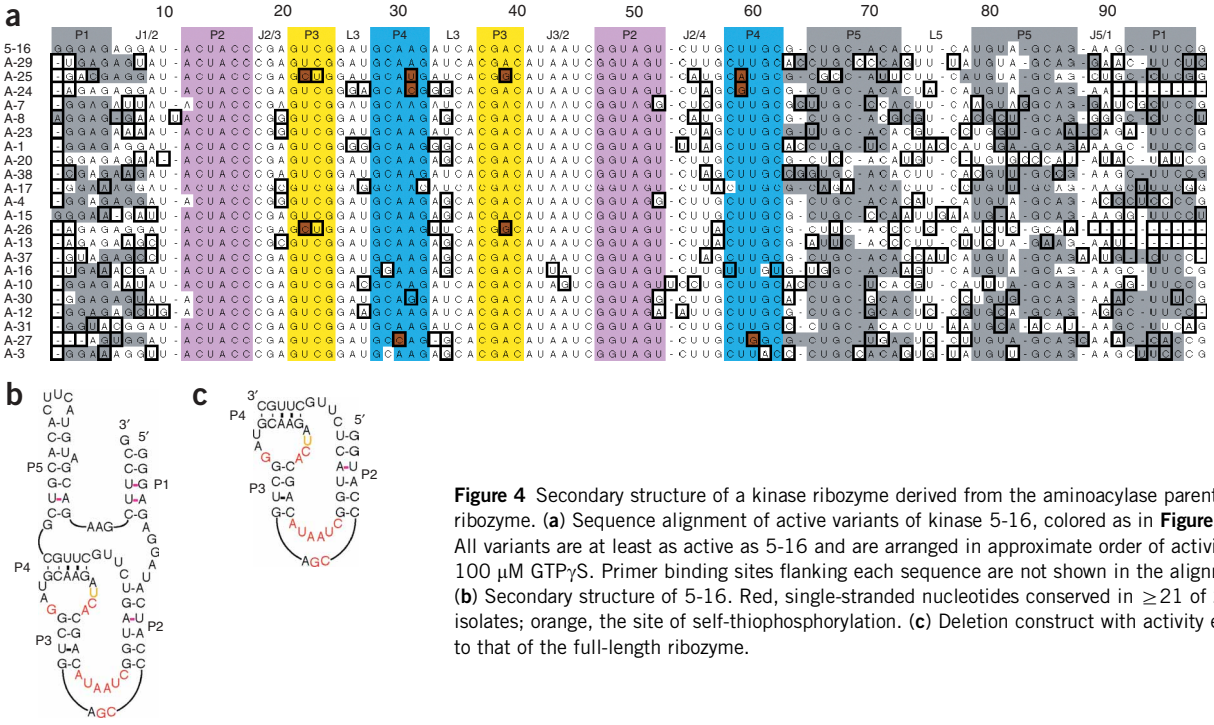

Figure 4 Secondary structure of a kinase ribozyme derived from the aminoacylase parent ribozyme. (**a**) Sequence alignment of active variants of kinase 5-16, colored as in **Figure 3**. All variants are at least as active as 5-16 and are arranged in approximate order of activity at 100 μM GTPγS. Primer binding sites flanking each sequence are not shown in the alignment. (**b**) Secondary structure of 5-16. Red, single-stranded nucleotides conserved in ≥21 of 22 isolates; orange, the site of self-thiophosphorylation. (**c**) Deletion construct with activity equal to that of the full-length ribozyme.

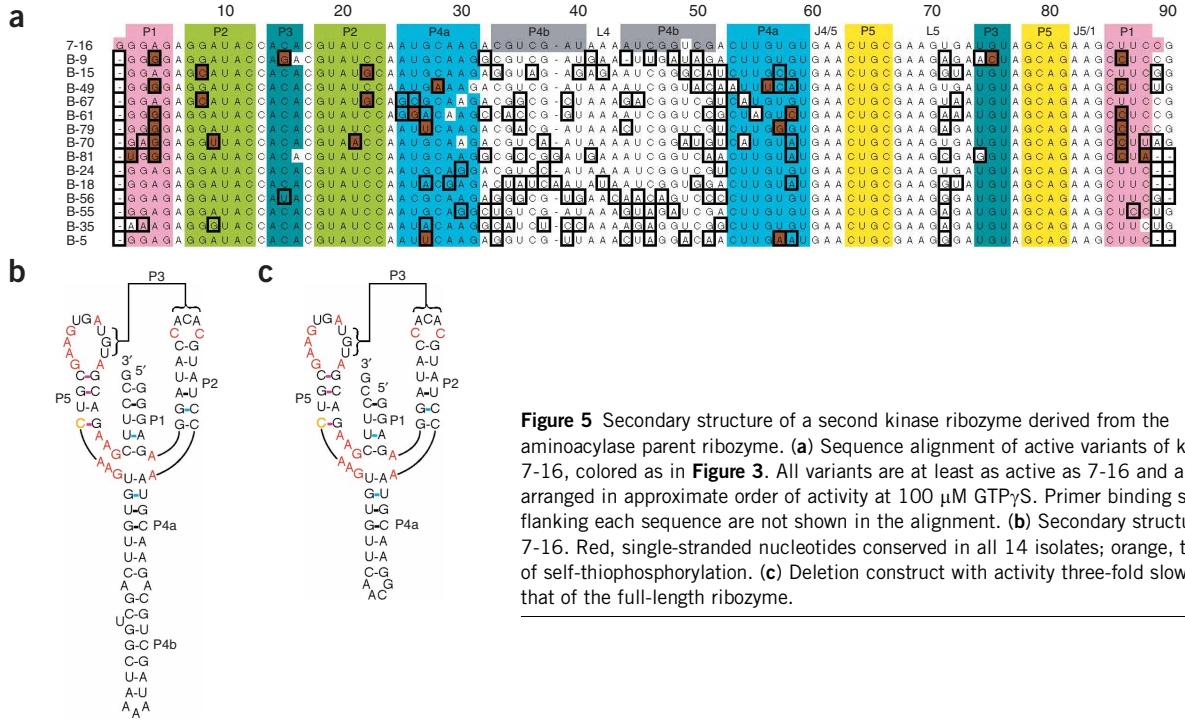

Figure 5 Secondary structure of a second kinase ribozyme derived from the aminoacylase parent ribozyme. (**a**) Sequence alignment of active variants of kinase 7-16, colored as in **Figure 3**. All variants are at least as active as 7-16 and are arranged in approximate order of activity at 100 μM GTPγS. Primer binding sites flanking each sequence are not shown in the alignment. (**b**) Secondary structure of 7-16. Red, single-stranded nucleotides conserved in all 14 isolates; orange, the site of self-thiophosphorylation. (**c**) Deletion construct with activity three-fold slower than that of the full-length ribozyme.

The secondary structure model of kinase 7-16 (**Fig. 5a–c**) was more complex than that of either the parent or 5-16, and consisted of a stem (P1) and three hairpins (P2, P4 and P5), two of which were linked by a loop-loop interaction (P3). Except for the distal portion of P4, each of the proposed helices in this structure was highly conserved. Interspersed between these helical elements were 16 additional invariant nucleotides, 10 of which were adenosines and several of which were clustered, hinting at possible A-minor interactions in the tertiary structure of this ribozyme[33]. In each of these isolates, P4 could be extended (extension shown only for 7-16 in **Fig. 5a**), although individual base pairs were generally not conserved. Consistent with this lack of conservation, the catalytic rate of a minimized version of 7-16 in which P4 was truncated was only three-fold lower than that of the full-length ribozyme (**Fig. 5c** and **Supplementary Fig. 3** online). Comparative analysis of 7-16 variants was problematic, because only about a third of the isolates had the same fold. Such an outcome has been observed previously[34,35], and its occurrence depends on several variables, including the degree to which the pool is mutagenized, the complexity of the parental ribozyme and the relative catalytic rates of variants with the parental fold and those with other folds. Because of this complication, site-directed mutagenesis was used to further test the model. The results provided strong support for each of the helices in the ribozyme (**Supplementary Fig. 3**).

The secondary structure model of kinase 5-16 was completely different from that of the parent ribozyme, with no base pairs in common (**Fig. 6**). The secondary structure of kinase 7-16 was mostly different from that of the parent as well, with 23 new base pairs and only 9 retained ones (**Fig. 6**). Furthermore, none of the 24 base pairs in the minimized version of 7-16 (**Fig. 5c**) were present in the minimized version of the parent (**Fig. 3c**). Although the folds of the other 21 classes of kinase ribozymes isolated in this experiment are not known, none of these ribozymes have the potential to form all of the helices in the secondary structure of the parent, and many contain disruptions in each of the parental helices. Thus, the folds of the 23 kinase ribozymes were largely different from that of the parental aminoacylase ribozyme.

DISCUSSION

Our results indicate that ribozymes with new catalytic activities can be found within a short mutational distance of a given ribozyme but that the probability of finding such ribozymes increases markedly as the mutational distance from the starting ribozyme increases. This seems to reflect a need to escape the fold of the starting ribozyme, in that the folds of the kinase ribozymes we isolated were distinct from that of the aminoacylase parent ribozyme. These results give a better sense of how readily ribozymes with new biochemical activities can arise

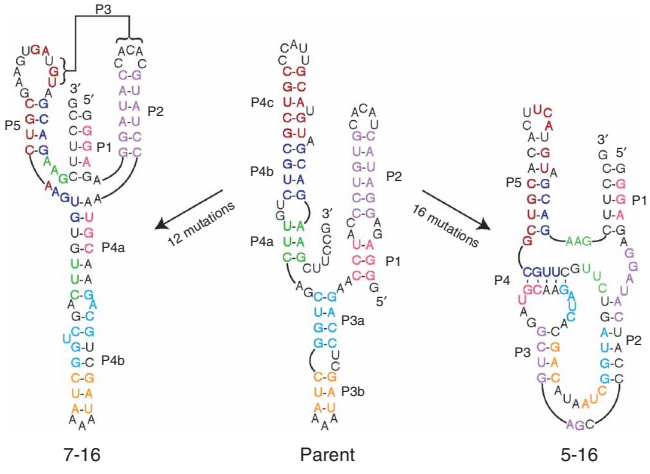

Figure 6 Comparison of the secondary structures of the parent ribozyme and two kinase ribozymes isolated in this selection. Colored segments represent paired regions in the parental structure and are mapped onto the structures of 5-16 and 7-16.

from existing ribozymes. On the average, 14 mutational changes were needed to convert a 90-nt aminoacylase ribozyme into a kinase ribozyme. They also suggest that, for a ribozyme with a new catalytic activity to arise, it is likely that a new RNA fold must also arise. This also seems to be true for aptamers[24,25,36], except in some cases in which variants of an existing aptamer are selected to bind a very similar target molecule[37,38]. In contrast, protein enzymes with very different substrates and activities are sometimes found in the context of the same scaffold, such as the α/β hydrolase fold[5,6]. Finally, these results provide experimental support for the emerging idea that RNA evolution is not continuous[39,40]. Instead, relatively few mutational changes in an existing ribozyme can generate new ribozymes with completely different catalytic activities and folds, thereby providing a mechanism for the emergence of new folds and activities from pre-existing ribozymes.

METHODS

Isolation of kinase ribozymes by *in vitro* selection. Pool RNA was dephosphorylated using calf alkaline phosphatase and gel purified. After heating at 65 °C for 5 min and cooling at room temperature for 5 min, pool RNA was incubated with 0.1–1 mM GTPγS in the presence of selection buffer (10 mM $MgCl_2$, 5 mM $CaCl_2$, 200 mM KCl and 100 mM HEPES buffer (pH 7.25)) for 20 min to 24 h. Reacted molecules were purified on APM polyacrylamide gels, eluted, amplified by RT-PCR and transcribed to generate RNA for the next round of selection. Additional details are provided in **Supplementary Methods** online.

Isolation of aminoacylase ribozymes by *in vitro* selection. Adenylated phenylalanine was synthesized as described[16]. After heating at 65 °C for 5 min and cooling at room temperature for 5 min, pool RNA was mixed with selection buffer and then added to a freshly dissolved solution of adenylated phenylalanine (0.1–1 mM final concentration). After a 1- to 15-min incubation, RNA was precipitated in 0.3 M sodium acetate (pH 5) and then biotinylated (280 mM HEPES buffer (pH 7.6) and 50 mM sulfo-NHS-biotin for 30 min at room temperature). RNA was then precipitated in 0.3 M sodium acetate (pH 5) and resuspended in a solution containing 4 M urea, 12.5 mM EDTA and 25 μM streptavidin. RNA was loaded without heating on a polyacrylamide gel and electrophoresed at 22 W for 35 min. The streptavidin-shifted band was then excised and eluted. Additional details are provided in **Supplementary Methods** online.

Pool synthesis, pool characterization and ribozyme characterization. See **Supplementary Methods** online.

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

ACKNOWLEDGMENTS

We thank M. Lawrence for comments on this manuscript, and members of the lab for helpful discussions. This work was supported by a grant from the National Institutes of Health.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Published online at http://www.nature.com/nsmb/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Kendrew, J.C. *et al.* A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* **181**, 662–666 (1958).
2. Patthy, L. *Protein Evolution* (Blackwell Science, Oxford, 1999).
3. Branden, C. & Tooze, J. *Introduction to Protein Structure* (Garland Publishing, New York, USA, 1999).
4. Babbitt, P.C. & Gerlt, J.A. Understanding enzyme superfamilies. Chemistry as the fundamental determinant in the evolution of new catalytic activities. *J. Biol. Chem.* **272**, 30591–30594 (1997).
5. Ollis, D.L. *et al.* The α/β-hydrolase fold. *Protein Eng.* **5**, 197–211 (1992).
6. Holmquist, M. α/β-hydrolase fold enzymes: structures, functions and mechanisms. *Curr. Protein Pept. Sci.* **1**, 209–235 (2000).
7. Ohno, S. *Evolution by Gene Duplication* (Springer-Verlag, New York, USA, 1970).
8. Newcomb, R.D. *et al.* A single amino acid substitution converts a carboxylesterase to an organophosphorus hydrolase and confers insecticide resistance on a blowfly. *Proc. Natl. Acad. Sci. USA* **94**, 7464–7468 (1997).
9. O'Brien, P.J. & Herschlag, D. Catalytic promiscuity and the evolution of new enzymatic activities. *Chem. Biol.* **6**, R91–R105 (1999).
10. Keefe, A.D. & Szostak, J.W. Functional proteins from a random-sequence library. *Nature* **410**, 715–718 (2001).
11. Raffler, N.A., Schneider-Mergener, J. & Famulok, M. A novel class of small functional peptides that bind and inhibit human α-thrombin isolated by mRNA display. *Chem. Biol.* **10**, 69–79 (2003).
12. Skerra, A. Engineered protein scaffolds for molecular recognition. *J. Mol. Recognit.* **13**, 167–187 (2000).
13. Jaeger, L., Wright, M.C. & Joyce, G.F. A complex ligase ribozyme evolved *in vitro* from a group I ribozyme domain. *Proc. Natl. Acad. Sci. USA* **96**, 14712–14717 (1999).
14. Yoshioka, W., Ikawa, Y., Jaeger, L., Shiraishi, H. & Inoue, T. Generation of a catalytic module on a self-folding RNA. *RNA* **10**, 1900–1906 (2004).
15. Grishin, N.V. Fold change in evolution of protein structures. *J. Struct. Biol.* **134**, 167–185 (2001).
16. Illangasekare, M., Sanchez, G., Nickles, T. & Yarus, M. Aminoacyl-RNA synthesis catalyzed by an RNA. *Science* **267**, 643–647 (1995).
17. Illangasekare, M. & Yarus, M. Specific, rapid synthesis of Phe-RNA by RNA. *Proc. Natl. Acad. Sci. USA* **96**, 5470–5475 (1999).
18. Lacey, J.C., Jr., Senaratne, N. & Mullins, D.W., Jr. Hydrolytic properties of phenylalanyl- and *N*-acetylphenylalanyl adenylate anhydrides. *Orig. Life Evol. Biosph.* **15**, 45–54 (1984).
19. Knight, R. & Yarus, M. Analyzing partially randomized nucleic acid pools: straight dope on doping. *Nucleic Acids Res.* **31**, e30 (2003).
20. Lorsch, J.R. & Szostak, J.W. *In vitro* evolution of new ribozymes with polynucleotide kinase activity. *Nature* **371**, 31–36 (1994).
21. Li, Y. & Breaker, R.R. Phosphorylating DNA with DNA. *Proc. Natl. Acad. Sci. USA* **96**, 2746–2751 (1999).
22. Igloi, G.L. Interaction of tRNAs and of phosphorothioate-substituted nucleic acids with an organomercurial. Probing the chemical environment of thiolated residues by affinity electrophoresis. *Biochemistry* **27**, 3842–3849 (1988).
23. Unrau, P.J. & Bartel, D.P. RNA-catalysed nucleotide synthesis. *Nature* **395**, 260–263 (1998).
24. Held, D.M., Greathouse, S.T., Agrawal, A. & Burke, D.H. Evolutionary landscapes for the acquisition of new ligand recognition by RNA aptamers. *J. Mol. Evol.* **57**, 299–308 (2003).
25. Huang, Z. & Szostak, J.W. Evolution of aptamers with a new specificity and new secondary structures from an ATP aptamer. *RNA* **9**, 1456–1463 (2003).
26. Schuster, P., Fontana, W., Stadler, P.F. & Hofacker, I.L. From sequences to shapes and back: a case study in RNA secondary structures. *Proc. R. Soc. Lond. B* **255**, 279–284 (1994).
27. Ellington, A.D. & Szostak, J.W. *In vitro* selection of RNA molecules that bind specific ligands. *Nature* **346**, 818–822 (1990).
28. Ekland, E.H. & Bartel, D.P. The secondary structure and sequence optimization of an RNA ligase ribozyme. *Nucleic Acids Res.* **23**, 3231–3238 (1995).
29. Putz, J. *et al.* Rapid selection of aminoacyl-tRNAs based on biotinylation of α-NH_2 group of charged amino acids. *Nucleic Acids Res.* **25**, 1862–1863 (1997).
30. Pagratis, N.C. Rapid preparation of single-stranded DNA from PCR products by streptavidin-induced electrophoretic mobility shift. *Nucleic Acids Res.* **24**, 3645–3646 (1996).
31. Baskerville, S. & Bartel, D.P. A ribozyme that ligates RNA to protein. *Proc. Natl. Acad. Sci. USA* **99**, 9154–9159 (2002).
32. Tuerk, C. *et al.* CUUCGG hairpins: extraordinarily stable RNA secondary structures associated with various biochemical processes. *Proc. Natl. Acad. Sci. USA* **85**, 1364–1368 (1988).
33. Nissen, P., Ippolito, J.A., Ban, N., Moore, P.B. & Steitz, T.A. RNA tertiary interactions in the large ribosomal subunit: the A-minor motif. *Proc. Natl. Acad. Sci. USA* **98**, 4899–4903 (2001).
34. Burke, D.H. *et al.* RNA aptamers to the peptidyl transferase inhibitor chloramphenicol. *Chem. Biol.* **4**, 833–843 (1997).
35. Tuschl, T., Sharp, P.A. & Bartel, D.P. A ribozyme selected from variants of U6 snRNA promotes 2′,5′-branch formation. *RNA* **7**, 29–43 (2001).
36. Mannironi, C., Scerch, C., Fruscoloni, P. & Tocchini-Valentini, G.P. Molecular recognition of amino acids by RNA aptamers: the evolution into an L-tyrosine binder of a dopamine-binding RNA motif. *RNA* **6**, 520–527 (2000).
37. Famulok, M. Molecular recognition of amino acids by RNA-aptamers: an L-citrulline binding RNA motif and its evolution into an L-arginine binder. *J. Am. Chem. Soc.* **116**, 1698–1706 (1994).
38. Yang, Y., Kochoyan, M., Burgstaller, P., Westhof, E. & Famulok, M. Structural basis of ligand discrimination by two related RNA aptamers resolved by NMR spectroscopy. *Science* **272**, 1343–1347 (1996).
39. Fontana, W. & Schuster, P. Continuity in evolution: on the nature of transitions. *Science* **280**, 1451–1455 (1998).
40. Schultes, E.A. & Bartel, D.P. One sequence, two ribozymes: implications for the emergence of new ribozyme folds. *Science* **289**, 448–452 (2000).