SUPPLEMENTARY METHODS

Pool synthesis. An Expedite Nucleic Acid Synthesis System was used to synthesize partially randomized DNA oligonucleotides used to construct the four pools used in these experiments. The first of these oligonucleotides encoded a partially degenerate version of the parent ribozyme, and had the sequence CGGAAGCTTCTGC**TACATGCAATGG GCAGCGCAGACAAGTCGACCGATTTTATCGAGGTCTTGGATGGGACACG TGTA**GTATCCTCTCCCTATA (positions shown in bold type mutagenized at a rate of 10%; T7 promoter sequence underlined). The second had the same sequence as the first, but positions were mutagenized at a level of 20% rather than 10%. Following deprotection and gel purification, T7 promoters were added to these oligonucleotides by large scale PCR[1] using the primers AACTAATACGACTCACTATAGGGAGAGG and CGGAAGCTTCTGC. The resulting double-stranded DNA templates were transcribed using T7 RNA polymerase[2]. The third encoded a partially degenerate version of kinase ribozyme 5-16, flanked on each side by a 4-nt spacer and an 18-nt primer-binding site. This oligonucleotide had the sequence TTACGGGATTCGTCATTCTTTT**CGGAAGC TTCTGCTACATGAAGTGTGCAGCGCAAGCAAGACTACCGATTATGTCGT GATCTTGCATCCGACTCGGGTAGTATCCTCTCCC**TTTTACTGCAACCGCAT AGACC (positions shown in bold type mutagenized at a rate of 20%). Following deprotection and gel purification, a T7 promoter was added to this oligonucleotide by mutually primed synthesis using the primer AACTAATACGACTCACTATAGGTCT ATGCGGTTGCAGT, and the resulting double-stranded DNA template was transcribed using T7 RNA polymerase[2]. This oligonucleotide was amplified by RT-PCR using the primers AACTAATACGACTCACTATAGGTCTATGCGGTTGCAGT and TTACGGGATTCGTCATTC. The fourth encoded a partially degenerate version of

kinase ribozyme 7-16, flanked on each side by a 4-nt spacer and an 18-nt primer-binding site. This oligonucleotide had the sequence CAGCTAGCTTGCATAACGTTTT**CGGA AGCTTCTGCTACATCACTTCGCAGTTCACACAAGTCGACCGATTTTATCG ACGTCTTGCATTGGATACGTGTGGTATCCTCTCCC**TTTTTCTTACGCGATG CCGACC (positions shown in bold type mutagenized at a rate of 20%). Following deprotection and gel purification, a T7 promoter was added to this oligonucleotide by mutually primed synthesis using the primer AACTAATACGACTCACTATAGGTCGG CATCGCGTAAGA, and the resulting double-stranded DNA template was transcribed using T7 RNA polymerase[2]. This oligonucleotide was amplified by RT-PCR using the primers AACTAATACGACTCACTATAGGTCGGCATCGCGTAAGA and CAGCTAGCTTGCATAACG.

Isolation of kinase ribozymes by *in vitro* selection. To dephosphorylate pool RNA, it was incubated in the presence of calf alkaline phosphatase (Roche) at 50°C for 15 minutes (10 μM RNA and 0.2 units/μL phosphatase). After extracting with phenol and chloroform, RNA was purified by polyacrylamide gel electrophoresis, eluted, and precipitated with ethanol.

Pool RNA, including a trace amount of body-labeled pool RNA, was heated at 65°C for 5 minutes in water, and cooled at room temperature for 5 minutes. The reaction was initiated by adding an equal volume of a mix containing selection buffer and GTPγS (Sigma). Final concentrations were 1 μM RNA, either 1 mM (rounds 1–4) or 100 μM (rounds 5–6) GTPγS, 10 mM $MgCl_2$, 5 mM $CaCl_2$, 200 mM KCl, and 100 mM HEPES buffer, pH 7.25, in a volume of up to 15 mL. After incubating at room temperature for either 24 hours (rounds 1–4), 1 hour (round 5), or 20 minutes (round 6), the reaction was split into 2 tubes, and a 90-nt radiolabeled RNA marker with a 5' thiophosphate was

added to one of the tubes. Following ethanol precipitation, the contents of each tube were resuspended in gel loading buffer (8 M urea, 25 mM EDTA, xylene cyanol, and bromophenol blue).

RNA was then heated at 80°C for 5 minutes, and loaded on a 6% polyacrylamide, 20–40 μM APM stacking gel. Electrophoresis was at 22 watts for 45–90 minutes. RNA was visualized by autoradiography, and molecules that co-migrated with the marker oligonucleotide were excised from the gel with a sterile razor blade. RNA was eluted from the gel overnight, on a rotator, at 4°C, in 20 mM DTT and 300 mM NaCl. In some rounds of selection, two consecutive gel purifications were performed.

Following ethanol precipitation in the presence of carrier (either tRNA or reverse transcription primer), RNA was resuspended in water, heated at 70°C for 3 minutes, and cooled at room temperature for 3 minutes. After adding buffer, and heating at 48°C for 2 minutes, reverse transcriptase was added to the reaction (1900 units SuperScript II RNase H^- reverse transcriptase from Invitrogen and 2.5 μM reverse transcription primer in a volume of 100 μL). After a 1 hour incubation at 48°C, RNA was hydrolyzed at 90°C for 15 minutes in the presence of 20 mM Tris base, 4 mM EDTA, and 0.3 M KOH. The pH of the solution was then adjusted to 8.5 by adding HCl. DNA was amplified by PCR without further purification in a volume at least 15 times greater than that of the reverse transcription. At various times, aliquots were removed and analyzed by agarose gel electrophoresis. Once a PCR product could be detected, DNA was precipitated with ethanol and transcribed using T7 RNA polymerase.

In subsequent selections to characterize 5-16 and 7-16, pool RNA was not dephosphorylated, because both the phosphorylated and dephosphorylated forms of these ribozymes were active. In addition, to prevent potential pairing between primer binding sites and the core of each ribozyme, blocking oligonucleotides (each at a final

concentration of 1 μM) complementary to the primer binding sites of these pools were present during incubation with GTPγS. The oligonucleotides ACTGCAACCGCATAGA CC and TTACGGGATTCGTCATTC were used to block the primer binding sites in the 5-16 pool, and the oligonucleotides TCTTACGCGATGCCGACC and CAGCTAGCTTG CATAACG were used to block the primer binding sites in the 7-16 pool.

Isolation of aminoacylase ribozymes by *in vitro* selection. Adenylated phenylalanine was synthesized as described[3]. Following methanol elution from a Waters Sep-Pak C18 cartridge, aliquots (typically 25 μL) were concentrated to dryness in a Savant speed vacuum using a low drying rate. Evaporation typically took 45–90 minutes, and analysis by reversed-phase HPLC[3] indicated that preparations were typically 75% pure (contaminated with phenylalanine and 5' AMP that presumably formed during purification). Dry aliquots were stored at –80°C, and typically used within 2–3 weeks.

Following transcription and gel purification, pool RNA, including a trace amount of body-labeled pool RNA, was heated at 65°C for 5 minutes in water, and cooled at room temperature for 5 minutes. The reaction was initiated by first adding selection buffer, and then adding a solution of adenylated phenylalanine. Because adenylated phenylalanine is extremely susceptible to hydrolysis[4], aliquots were resuspended in water immediately before use. Final conditions were 1 μM RNA, either 1 mM (rounds 1–4) or 100 μM (round 5) phenylalanine-AMP, 10 mM $MgCl_2$, 5 mM $CaCl_2$, 200 mM KCl, and 100 mM HEPES buffer, pH 7.25, in a volume of up to 5 mL. Note that this buffer differs from that used in the original selection to isolate this ribozyme in several ways: it contains 200 mM KCl rather than 20 mM KOAc, it does not contain NaCl, and it does not contain trace amounts of $CuSO_4$, $FeCl_3$, $MnCl_2$, or $ZnCl_2$[3]. After incubating at room temperature for either 15 minutes (round 1), or 1–3 minutes (rounds 2–5), the reaction

was split into 2 tubes, and a radiolabeled positive control (a 10-nt RNA oligonucleotide from Dharmacon Research containing a primary amine at its 3' terminus) was added to one of the tubes.

Following ethanol precipitation in the presence of 300 mM sodium acetate (pH 5), aminoacylated RNA was biotinylated (280 mM HEPES buffer, pH 7.6, and 50 mM sulfo-NHS-biotin from Pierce in a volume of 100 μL). After a 30 minute incubation at room temperature, RNA was ethanol precipitated in the presence of 300 mM sodium acetate (pH 5), and the contents of each tube were resuspended in 90 μL gel-loading buffer (8 M urea, 25 mM EDTA, xylene cyanol, and bromophenol blue). 30 μL of a 50 μM streptavidin solution was then added to each tube.

RNA was loaded without heating on a 6%, 0.75 mm thick polyacrylamide gel, and electrophoresis was at 22 watts for 35 minutes. Note that because aminoacylated RNA is considerably more stable following derivatization with sulfo-NHS-biotin, it was not necessary to perform electrophoresis at low pH, as is typical when working with aminoacylated RNA. RNA was visualized by autoradiography, and molecules that co-migrated with the positive control were excised from the gel with a sterile razor blade. RNA was eluted from the gel overnight, on a rotator, at 4°C, in 50 mM Tris-HCl, pH 8.5 and 5 mM EDTA. Reverse transcription, PCR, and transcription were as described for the kinase selection.

Ribozyme activity assays. In a typical assay, a trace amount of body labeled RNA was heated and cooled, and selection buffer and substrate was added to initiate the reaction. Kinase reactions were monitored on APM polyacrylamide gels. Aminoacylase reactions were analyzed using either a streptavidin gel shift assay or a TLC assay similar to that previously described[5]. Initial rates were estimated from the earliest extent of reaction that

could be measured using the formula $R = 1-e^{-kt}$, where R represents the fraction of ribozyme reacted, t represents time, and k represents the ribozyme rate. For aminoacylase reactions, time points were typically taken within the first minute of the reaction, so that hydrolysis of adenylated phenylalanine was not significant. k_{cat}/K_m values were determined by plotting initial rate as a function of substrate concentration and determining the slope of the linear part of this curve. Full-length ribozymes were not separated from n+1 transcripts, which were presumably inactive. Inclusion of these inactive transcripts is estimated to have decreased the observed reaction rates by 2- to 3-fold.

Characterization of kinase ribozyme activities. To determine whether the 5' hydroxyl group of a reacted ribozyme was modified, it was incubated at 37°C for 30 minutes in the presence of 10 units of T4 polynucleotide kinase (New England BioLabs) and ATP-γ^{32}P and analyzed by polyacrylamide gel electrophoresis. To further characterize the ribozymes that could not be 5' labeled with T4 polynucleotide kinase, they were reacted with GTPγS-γ^{35}S, digested with P1 nuclease, and analyzed by PEI-cellulose thin layer chromatography as previously described[6]. To map the reaction sites of the ribozymes that modify themselves at internal sites, reacted RNA was 5' labeled, and following partial base hydrolysis in 100 mM sodium citrate (80°C for 75 seconds), aliquots were neutralized with HCl and analyzed on either polyacrylamide or APM polyacrylamide sequencing gels. To observe the reverse reaction of 5-16, an optimized version of this ribozyme (data not shown) was labeled with GTPγS-γ^{35}S and purified by APM polyacrylamide gel electrophoresis. Reacted ribozyme was incubated in the presence of selection buffer and 1 mM leaving group, and time points were analyzed by TLC as described[6]. For the experiment shown in **Fig. 1f**, a 92-hour time point was analyzed.

Calculations. Standard formulas were used to calculate the number of unique sequences in our starting pool[7], the number of sequences at specified mutational distances from the parent ribozyme[7], and the probability of a base pair being conserved in a certain number of isolates from a degenerate pool[8]. The probability that two kinase ribozymes in different classes derived from the parent will contain the same mutation at a particular position by chance (P) is approximately equal to the probability that the ribozymes will contain the same point mutation at this position ($P_m=M^2/3$, where M is the frequency of point mutations in the kinase ribozymes isolated in this selection), plus the probability that the ribozymes will both contain a deletion at this position ($P_d=D^2$, where D is the frequency of deletions in the kinase ribozymes isolated in this selection). Because the frequency of insertions in these kinase ribozymes was low (F=0.007), they make a negligible contribution to P, and therefore were not considered in this calculation. For the 23 kinase ribozymes analyzed, M=0.18 and D=0.024, so that P=0.011. For 65 mutagenized positions, then, the probability that two unrelated kinase ribozymes will contain x mutations in common can be calculated using the binomial distribution.

SUPPLEMENTARY REFERENCES

1. Bartel, D.P. & Szostak, J.W. Isolation of new ribozymes from a large pool of random sequences. *Science* **261**, 1411-1418 (1993).
2. Milligan, J.F., Groebe, D.R., Witherell, G.W. & Uhlenbeck, O.C. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res.* **15**, 8783-8798 (1987).
3. Illangasekare, M., Sanchez, G., Nickles, T. & Yarus, M. Aminoacyl-RNA synthesis catalyzed by an RNA. *Science* **267**, 643-647 (1995).
4. 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).
5. Illangasekare, M. & Yarus, M. Specific, rapid synthesis of Phe-RNA by RNA. *Proc. Natl. Acad. Sci. USA* **96**, 5470-5475 (1999).
6. Lorsch, J.R. & Szostak, J.W. *In vitro* evolution of new ribozymes with polynucleotide kinase activity. *Nature* **371**, 31-36 (1994).

7. Knight, R. & Yarus, M. Analyzing partially randomized nucleic acid pools: straight dope on doping. *Nucleic Acids Res.* **31**, e30 (2003).
8. Ekland, E.H. & Bartel, D.P. The secondary structure and sequence optimization of an RNA ligase ribozyme. *Nucleic Acids Res.* **23**, 3231-3238 (1995).