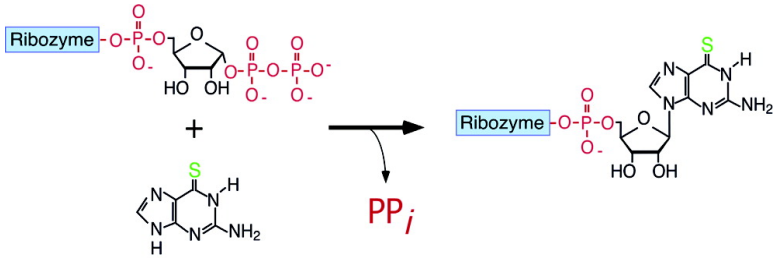
Article

Isolation of Fast Purine Nucleotide Synthase Ribozymes

Matthew W. L. Lau, Kelly E. C. Cadieux, and Peter J. Unrau

J. Am. Chem. Soc., **2004**, 126 (48), 15686-15693• DOI: 10.1021/ja045387a • Publication Date (Web): 12 November 2004



More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 1 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

J|A|C|S
ARTICLES
Published on Web 11/12/2004

Isolation of Fast Purine Nucleotide Synthase Ribozymes

Matthew W. L. Lau, Kelly E. C. Cadieux, and Peter J. Unrau*

Contribution from the Department of Molecular Biology and Biochemistry, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia V5A 1S6, Canada

Received July 30, 2004; E-mail: punrau@sfu.ca

Abstract: Here we report the in vitro selection of fast ribozymes capable of promoting the synthesis of a purine nucleotide (6-thioguanosine monophosphate) from tethered 5-phosphoribosyl 1-pyrophosphate (PRPP) and 6-thioguanine (^{6S}Gua). The two most proficient purine synthases have apparent efficiencies of 284 and 230 M^{-1} min^{-1} and are both significantly more efficient than pyrimidine nucleotide synthase ribozymes selected previously by a similar approach. Interestingly, while both ribozymes showed good substrate discrimination, one ribozyme had no detectable affinity for 6-thioguanine while the second had a K_m of ~80 μM, indicating that these ribozymes use considerably different modes of substrate recognition. The purine synthases were isolated after 10 rounds of selection from two high-diversity RNA pools. The first pool contained a long random sequence region. The second pool contained random sequence elements interspersed with the mutagenized helical elements of a previously characterized 4-thiouridine synthase ribozyme. While nearly all of the ribozymes isolated from this biased pool population appeared to have benefited from utilizing one of the progenitor's helical elements, little evidence for more complicated secondary structure preservation was evident. The discovery of purine synthases, in addition to pyrimidine synthases, demonstrates the potential for nucleotide synthesis in an 'RNA World' and provides a context from which to study small molecule RNA catalysis.

Introduction

The 'RNA World' hypothesis suggests that RNA predates protein in evolution.[1,2] While this parsimonious model is increasingly consistent with our detailed knowledge of metabolism, it is still unclear if RNA catalysts (ribozymes) can manipulate the small substrates required for an RNA-based metabolism. Specifically, RNA replication requires nucleotide monomers that are in turn synthesized from simpler compounds. Modern metabolism uses PRPP in at least 16 different pathways to synthesize pyrimidine nucleotides (OMP, UMP, Figure 1a), purine nucleotides (AMP, GMP, IMP, XMP, Figure 1b), several pyridine nucleotide cofactors, and the amino acids histidine and tryptophan.[3] As purine and pyrimidine bases are known to be synthesized by prebiotic processes,[4,5] a hypothetical RNA-based metabolism could have used purine and pyrimidine bases together with PRPP to synthesize the nucleotide building blocks required to replicate RNA.

The ability of a ribozyme to efficiently perform small molecular chemistry requires both substrate recognition and rate acceleration. At one extreme, natural and artificial RNA sequences that bind, but do not react with, small metabolically relevant substrates demonstrate that RNA can recognize small molecules with nanomolar affinities and high specificity.[6−8] At the other end of the spectrum, ribozymes that recognize their substrates through Watson−Crick base pairs demonstrate that RNA can perform efficient catalysis when substrate recognition is performed tens of angstroms away from the site of chemical modification.[7] Decreasing the substrate size forces a catalytic system to perform both recognition and rate acceleration within an increasingly small region of space. Observing the catalytic strategies employed by ribozymes performing similar chemistry on a succession of small substrates may therefore provide clues as to how RNA compromises between these two important catalytic parameters.

Previously, we isolated three families of ribozymes able to perform pyrimidine nucleotide synthesis using tethered PRPP and free 4-thiouracil (^{4S}Ura).[9] These ribozymes displayed high specificity for their ^{4S}Ura substrate and had apparent efficiencies ($k_{cat\ app}/K_m$) in the 0.7−7 M^{-1} min^{-1} range. One family was optimized and shown to perform rate-limiting and highly dissociative chemistry likely to involve the charge-mediated stabilization of an oxocarbenium-ion intermediate.[10,11] As this small molecule chemistry differs considerably from other known ribozyme reactions in both mechanism and substrate size, an exploration of purine nucleotide synthesis will provide an important context to our initial nucleotide synthesis studies. Here we isolate purine nucleotide synthases with a range of binding

(1) Bartel, D. P.; Unrau, P. J. *Trends Biochem. Sci.* ***1999***, *24*, M9−M13.
(2) Joyce, G. F. *Nature* **2002**, *418*, 214−221.
(3) *Kyoto Encyclopedia of Genes and Genomes*, Release 31.0, 2004.
(4) Joyce, G. F. *Nature* **1989**, *338*, 217−224.
(5) Orgel, L. E. *Trends Biochem. Sci.* **1998**, *23*, 491−495.
(6) Winkler, W. C.; Nahvi, A.; Roth, A.; Collins, J. A.; Breaker, R. R. *Nature* **2004**, *428*, 281−286.
(7) Wilson, D. S.; Szostak, J. W. *Annu. Rev. Biochem.* **1999**, *68*, 611−647.
(8) Osborne, S. E.; Ellington, A. D. *Chem. Rev.* **1997**, *97*, 349−370.
(9) Unrau, P. J.; Bartel, D. P. *Nature* **1998**, *395*, 260−263.
(10) Unrau, P. J.; Bartel, D. P. *Proc. Natl. Acad. Sci., U.S.A.* **2003**, *100*, 15393−15397.
(11) Chapple, K. E.; Bartel, D. P.; Unrau, P. J. *RNA* **2003**, *9*, 1208−1220.

10.1021/ja045387a CCC: $27.50

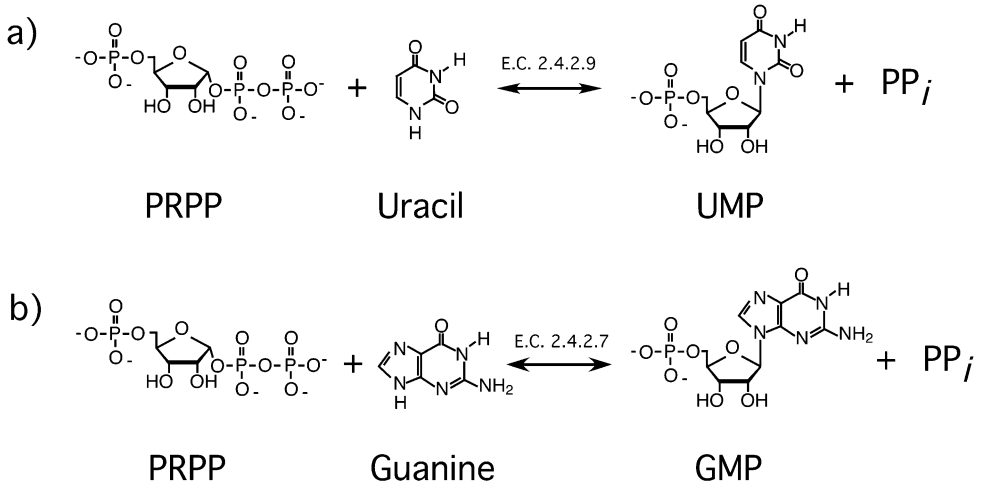


Figure 1. Reactions catalyzed by pyrimidine and purine nucleotide synthases. (a) Uracil phosphoribosyltransferase-catalyzed (EC 2.4.2.9) synthesis of UMP from PRPP and uracil. (b) Hypoxanthine guanine phosphoribosyltransferase-catalyzed (HGPRTase, EC 2.4.2.7) synthesis of GMP from PRPP and guanine.

affinities and having efficiencies dramatically higher than those observed for pyrimidine nucleotide synthase ribozymes.

Methods

AppRpp and App^{6S}G Synthesis. The synthesis of adenylated PRPP (AppRpp) was as previously described.[9] Adenylated 6-thioguanosine-5′-monophosphate (App^{6S}G) was synthesized by reacting 100 mM ^{6S}GMP with 50 mM adenosine 5′-phosphorimidazolide[12] in the presence of 100 mM $MgCl_2$ and 200 mM HEPES (pH 7.4) at 50 °C for 3 h followed by HPLC purification.[9] ^{6S}GMP was synthesized from 6-thioguanosine (Sigma) according to the method of Breter and Mertes[13] and purified on a A-25 DEAE Sephadex column (Sigma) using a 50−400 mM triethylammonium bicarbonate (TEAB) pH 7.5 gradient; product eluted at 250−300 mM TEAB.

Designing of Selection Pools. Two DNA pools were synthesized on an ABI 392 DNA synthesizer using 0.2 μmol of 2000 Å control pore glass (CPG) columns (Glen Research) using standard cyanoethyl phosphoramidite chemistry. The first random pool had the final sequence 5′-*TTCTAATACGACTCACTATA*GGAGACGCCATCAA-N_{95}-TCGCACCGCAGCAAGC (-N_{95}-, 95 random nucleotides; first 34 nt, 5′ primer; last 16 nt, 3′ primer; T7 promoter sequence in italics). The second biased pool was created by mixing together equal amounts of two mutagenized subpools each having the general sequence: 5′-*TTCTAATACGACTCACTATA*$(GGAG)_I$CGAAGTGCCC-N_{11}-$(atcc)_I$-$N_{(1-3)}$-(gcc tatt)$_{II}$-N_{18}-$(aacga)_{III}$-$N_{(4-6)}$-$(gcttgc)_{IV}$-$(tcgtt)_{III}$-N_{10}-$(aataggc)_{II}$-$N_{(7-9)}$-$(gcggtg)_V$-tT CG$(CACCGC)_V$A$(GCAAGC)_{IV}$ (first 34 nt, 5′ primer; last 16 nt, 3′ primer; N, random nucleotide; lower case, mutagenized either 10% or 20% for each subpool; brackets delimit the five helical components found in the secondary structure of the family A nucleotide synthase ribozyme and are named with Roman numeral subscripts[11]). Variable length random regions were synthesized using a split and pool strategy.[14] Random and mutagenized nucleotide couplings in both pools were obtained by mixing equiactive phosphoramidite (Applied Biosytems) stocks (dA:dC:dG:T in the molar ratios 0.28:0.27:0.23:0.22) as described.[14] Sequencing of random positions revealed the following relative nucleotide frequencies: 25.8% A (92/356), 21.1% G (75/356), 28.7% C (102/356), and 24.4% T (87/356). Large-scale PCR was performed as described[14] using the following primers: random pool 5′-TTCTAATACGACTCACTATAGGAGACGCCATCAA and for the biased pool 5′-TTCTAATACGACTCACTATAGGAGCGAAGTGCCC. The 3′ primer was 5′-GCTTGCTGCGGTGCGA and common to both pools. A total of ∼2 nmol of DNA from each of the random and mutagenized pools (∼3×10^{14} unique sequences, given that four copies of each unique sequence was expected to be present after PCR amplification) was transcribed into RNA using T7 RNA polymerase.[11]

RNA Ligation Protocol. RNA at 3.33 μM was ligated with 60 μM AppRpp for 4 h at 23 °C in ligation buffer (50 mM HEPES, 10 mM $MgCl_2$, 3.3 mM dithiothreitol, 10 μg/mL BSA, 8.3% v/v glycerol, and 15% polyethylene glycol 8000 at pH 8.0) using 0.5 U/μL T4 RNA ligase.[15] The reaction was terminated by addition of EDTA followed by a phenol−chloroform extraction and ethanol precipitation. Marker RNA 125-nt long and having the 3′ terminal nucleotide sequence of UCAGAAGACAUCACAUUGC-3′ was derivatized to contain a terminal p^{6S}G by performing a ligation using 33 μM App^{6S}G for 4 h. Marker RNA was gel purified through an *N*-acryloylaminophenylmercuric acetate (APM) gel[16] for use in rounds 1−3.

Selection. Pool RNA ligated to PRPP was incubated at 0.24 μM in incubation buffer (50 mM Tris-HCl, 150 mM KCl, 25 mM $MgCl_2$, pH 7.5) supplemented with 0.26 mM ^{6S}Gua for 15 h (rounds 1−5). Five nanomoles of RNA were used in the first round of selection for each pool (∼10 copies of each pool sequence). Ribozyme reactions were stopped by adding one volume of gel-loading buffer (90% formamide and 50 mM EDTA). Reactive RNA was separated from nonreactive species using denaturing 6% PAGE gels containing 3.75 μM APM. During rounds 1−5, the RNA for each incubation was divided in two. One-half contained radiolabeled RNA pool to allow detection of reactive ribozymes. The second half contained unlabeled RNA pool mixed with the synthetic marker RNA, which served as an internal control to locate the position of the reactive species in the APM gel. Gel fragments at the position of the marker were excised and eluted (300 mM NaCl, 1 mM dithiothreitol) overnight and recovered by ethanol precipitation. The resulting RNA was reverse transcribed using Superscript II (50 mM Tris-HCl, 75 mM KCl, 3 mM $MgCl_2$, 10 mM dithiothreitol, 560 μM of each dNTP, pH 7.5, 5 μM 3′ primer, 10 U/μL enzyme) at 48 °C for 1 h. RNA was hydrolyzed with 100 mM KOH at 90 °C for 10 min. cDNA was neutralized with HCl and PCR amplified before re-entering the next round of selection. From rounds 6 to 10, the RNA pools were subjected to increasing time pressure by lowering of incubation time in each round (4 h, 1 h, 6 min, 1 min, and 15 s respectively).

TLC Analysis. Gel-purified ribozyme isolates were ligated with PRPP and then reacted to completion with ^{6S}Gua (15 h incubation). Reacted RNA was derivatized with radiolabeled cytidine 5′-[^{32}P], 3′

(12) Lohrmann, R.; Orgel, L. E. *Tetrahedron* **1978**, *34*, 853−855.
(13) Breter, H. J.; Mertes, H. *Biochim. Biophys. Acta* **1990**, *1033*, 124−132.
(14) Zaher, H. S.; Unrau, P. J. *Methods Mol. Biol.* **2004**; *288*, 359−378.
(15) Wang, Q. S.; Unrau, P. J. *Biotechniques* **2002**, *33*, 1256−1260.
(16) Igloi, G. L. *Biochemistry* **1988**, *27*, 3842−3849.

bisphosphate (pCp) using the ligation protocol described above in the presence of 2 μM ATP. pCp was synthesized by phosphorylating 3′-cytidine monophosphate (Sigma) with γ-[^{32}P] ATP using T4 polynucleotide kinase (NEB). The RNA radiolabeled at its 3′ end was gel purified using a 6% APM gel in order to isolate the thiol-containing material. The recovered RNA was digested into 3′-mononucleotides using T2 ribonuclease (25 mM sodium citrate, 4 mM dithiothreitol, pH 4.5, 0.26 U/μL T2 (Sigma), for 2 h at 37 °C). A 6-thioguanosine 3′-mononucleotide (^{6S}Gp) control was synthesized by ligation of the marker RNA (derivatized with 3′ p^{6S}G) with radiolabeled pCp. This sample was APM gel purified and digested with T2 ribonuclease as described above. Digested mononucleotides were separated using two-dimensional thin-layer chromatography on 10 cm × 10 cm cellulose TLC plates (J. T. Baker) presoaked in 1:10 saturated $(NH_4)_2SO_4$:H_2O. The first dimension was developed with 80% ethanol, second with 40:1 saturated $(NH_4)_2SO_4$:2-propanol; both solvents were supplemented with 100 mM β-mercaptoethanol.[9,17] Samples were spotted 1 cm in from the corner of the TLC plates.

Kinetic Analysis. Kinetic studies were performed in incubation buffer supplemented with $MgCl_2$ to a final concentration of 75 mM. Time points (6, 30, 60, 150, and 240 min) were taken and stopped by addition of an equal volume of gel-loading buffer. The reaction rate at a given ^{6S}Gua concentration was determined by simultaneously fitting the fraction reacted for at least five independent time courses (determined by phosphorimager analysis on a Molecular Dynamics Storm 820 of the resulting gels) to the equation $F = \beta(1 - e^{-k_{obs}t})$ using the program KaleidaGraph (Synergy Software), F being the fraction reacted at time t, k_{obs} the first-order apparent rate constant, and β the total fraction able to react. The resulting rates were fit to the Michaelis−Menten equation $k_{obs} = k_{cat\ app}[^{6S}Gua]/(K_m+[^{6S}Gua])$. A weighted error analysis was performed to obtain $k_{cat\ app}$ and K_m.

Analogues. Purine analogues 6-thioguanine ϵ(347 nm, pH 1) = 20 900 M^{-1} cm^{-1},[18] 6-thiopurine ϵ(325 nm, pH 1) = 20 500 M^{-1} cm^{-1},[19] 2-methyl-6-thiopurine assumed ϵ(330 nm, pH 1) = 20 000 M^{-1} cm^{-1}, 6,8-dithiopurine ϵ(358 nm, pH 1) = 27 800 M^{-1} cm^{-1};[20] 2,6-dithiopurine, 6-thio-9-methylpurine, 6-hydroxy-2-thiopurine, 2-hydroxy-6-thiopurine, 2-thiopurine, 2-amino-9-butyl-6-thiopurine, 2-methylthio-6-thiopurine, 8-methyl-6-thiopurine, 2,6-dithio-7-methylpurine, 6,8-dithio-2-hydroxypurine, 2,6,8-trithiopurine, as well as 4-thiouracil were obtained from Sigma-Aldrich, and saturated solutions were prepared in 1.05× incubation buffer.

Results

Selection. We were curious to understand the effect a secondary structure bias might have on the outcome of a selection for purine nucleotide synthesis. Could a secondary structure motif previously selected for its ability to perform pyrimidine nucleotide synthesis be beneficial to purine nucleotide synthesis? To address this question we constructed two pools which were subjected to in vitro selection in identical fashion. The first consisted of a 95 nt long random sequence pool having a diversity of ~3 × 10^{14} different sequences spread uniformly throughout sequence space. Our second pool, having the same number of different sequences, approximate length (92−98 nt of variable sequence), and 3′ primer binding sequence as the first, also contained significant amounts of random sequence (50−56 nt) but was on average only 4−8 mutations away from being able to form the complete secondary structure of a structurally complex pyrimidine nucleotide synthase ribozyme.[11]

RNA sequences from both random and structurally biased pools were selected for their ability to promote glycosidic bond formation between the PRPP at their 3′ ends and a free ^{6S}Gua substrate using an APM gel shift strategy[9] (Figure 2). The isolation of RNA containing a thiol group was possible due to the slowing of sulfur-containing material in a mercury-containing gel.[16] After 6 rounds of in vitro selection, ^{6S}Gua-dependent ribozymes having the same mobility as a RNA-p^{6S}G marker on an APM gel were observed in both pools after 15 h of incubation. At this point selection pressure was increased for both pools by lowering the incubation time allowed with the ^{6S}Gua substrate (see methods). By round 10, nearly 1% of each pool had reacted after 15 s of incubation. Both pools of ^{6S}G synthases had reaction rates (given by the ribozyme pools reactivity per unit substrate concentration) that were 50−100 times higher than the efficiency observed for an equivalent pool of pyrimidine nucleotide synthases (Figure 3). The ^{6S}G synthases of both pools seemed to have reached maximum reaction rates by round 9, slightly faster than the previously selected ^{4S}U synthases that required at least one more round of selection to plateau.

Sequence Analysis. Sequencing round 10 of the structurally biased pool revealed at least 20 distinct (as judged by primary sequence alignment) families of which six contained multiple isolates. One family was isolated eight times; this family was also found as a single isolate in round 6. One family was repeated three times, and four families were isolated twice each. These families are henceforth called MA to MF, where "M" stands for "mutagenized". The majority of families (12/20) shared a short conserved UCUUU sequence motif (we accepted in this count one C to U containing isolate) that was not found in the family A motif, and 6 of these 12 families appeared to extend this motif by another six residues to AGGCGUUCUUU (refer to Supporting Information Figure 1a). The short motif was found in a random sequence region of the biased pool immediately 5′ to a well-conserved hairpin loop, which was found in 18/20 families. This hairpin, which forms stem V in the original family A RNA motif (see Methods for sequence information), contains one arm that was not mutagenized in order to allow the efficient binding of a reverse transcription primer and thus would be expected to be conserved by chance 46% of the time. Only one isolate (MF) was found that could hypothetically contain all five helices of the family A motif, although this isolate had a mismatch in the middle of stem III of the family A motif. This would be expected to occur ~2% of the time by chance, roughly consistent with the number of isolates found (1/20). Moreover, folding with the PKNOTs algorithm[21] indicated that this sequence was likely to adopt a fold considerably different from that of the family A nucleotide synthase.

The round 10 random sequence pool was found to contain at least 33 distinct families. In contrast to the biased pool, none of the families appeared to form a terminal hairpin loop and none contained the UCUUU motif near their 3′ end, even though both pools contained exactly the same 3′ terminal sequence (refer to Supporting Information Figure 1b). Three of the random families, named RA, RB, and RC, occurred more than once (two were repeated three times and one was repeated twice out of 38 sequences total). One sequence from each of the three

(17) Gray, M. W. *Biochemistry* **1974**, *13*, 5453−5463.
(18) Elion, G. B.; Hitchings, G. H. *J. Am. Chem. Soc.* **1954**, *77*, 1676−1676.
(19) Elion, E. B. *J. Org. Chem.* **1962**, *27*, 2478−2491.
(20) Robins, R. K. *J. Am. Chem. Soc.* **1958**, *80*, 6671−6679.
(21) Rivas, E.; Eddy, S. R. *J. Mol. Biol.* **1999**, *285*, 2053−2068.

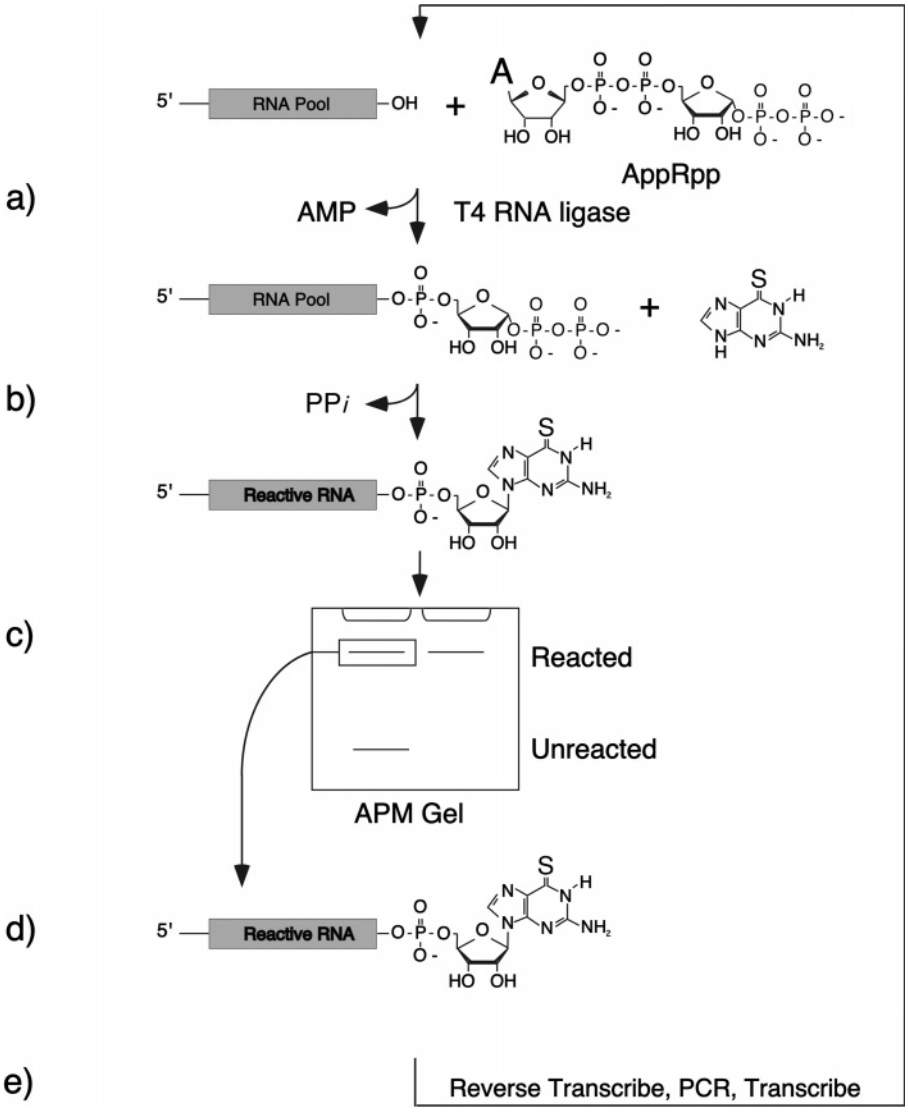


Figure 2. Overall in vitro selection scheme for purine nucleotide synthase ribozymes. (a) An RNA pool is derivatized with PRPP using AppRpp and T4 RNA ligase. (b) RNA pools tethered to PRPP are then incubated with ^{6S}Gua. Ribozymes catalyzing the synthesis of ^{6S}G from ^{6S}Gua release pyrophosphate (PP_i). (c) An APM gel was used to isolate active from nonreactive RNA (left lane). A radiolabeled marker RNA derivatized with a terminal ^{6S}G nucleotide was mixed with pool RNA and used to indicate the position of active ribozymes (right lane). (d) Reactive ribozymes are excised from the gel and eluted. (e) The enriched RNA pool is then reverse transcribed into DNA, PCR amplified, and transcribed back into RNA ready to enter the next round of selection.

repeating families along with three sequences from the remaining 30 'orphaned' families (called RD, RE, and RF) was selected for further analysis. These named sequences together with the functional families resulting from the structurally biased pool have been submitted to Genebank and have accession numbers AY701990−AY702000.

Magnesium Dependence. Two families from each pool (MA and ME, RA and RE) were evaluated for their difference in rate under varying magnesium-level conditions. The reactivity of the isolates appeared to increase linearly with magnesium concentration, with activity saturating or even decreasing above 75 mM $MgCl_2$. This optimal magnesium concentration was therefore used for all further kinetic assays on the remaining named families. The most reactive sequence (highest reaction rate and fraction reacted) from the random pool was a sequence from family RA, and for the mutagenized pool, it was from family MA (of all the tested isolates only the MD isolate was found to be unreactive). The initial reaction velocities for the random pool families (RB−RF) were, on average, 1.2−3 times slower than RA, while the mutagenized pool families (MB−MF, excluding MD) displayed a 5−11-fold decrease relative to MA. Since the fastest families had the highest frequency of occurrence, the selection appears, as desired, to have isolated ribozymes based on their catalytic prowess.

Product Characterization. To characterize further the utilization of ^{6S}Gua by the purine synthase ribozymes, RA and MA along with one of the less common isolates from each pool (RE and ME) were analyzed using thin-layer chromatography. Ribozymes reacted with ^{6S}Gua were labeled with 5′-[^{32}P] pCp so as to specifically tag the 3′ most nucleotide (Figure 4a). Ribozymes containing a thiol modification were purified using an APM gel and digested into mononucleotides that have a 3′ phosphate using T2 RNase (Figure 4b,c). A marker RNA derivatized with ^{6S}GMP at its 3′ end was also labeled with pCp and used to generate a radiolabeled ^{6S}Gp standard. Because the T2 digested standard, when treated with calf intestinal phosphatase, resulted in the production of radiolabeled inorganic phosphate and the disappearance of a shifted band on a high

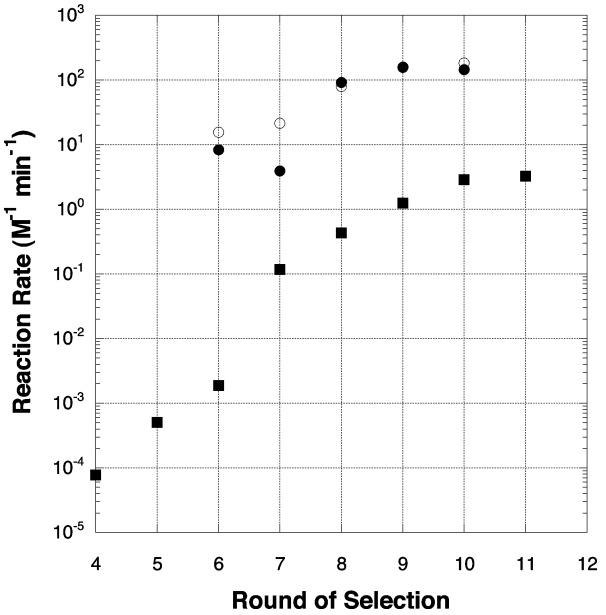


Figure 3. Purine ribozyme activity as a function of selection round. The random sequence pool (empty circles) and the biased pool resulting from mutagenizing the family A pyrimidine nucleotide synthase ribozyme (filled circles) have similar purine synthase activities during the course of the selection. Incubation times were initially 15 h, and by round 10 pools were incubated for only 15 s. In contrast to the purine nucleotide synthase pools, a previous selection for pyrimidine nucleotide synthases (solid squares) resulted in ribozyme populations that were 50−100 times less active.[9] Reaction rates were calculated by dividing the observed first-order reaction rates by the substrate concentration.

percentage APM polyacrylamide gel, we concluded that the nuclease was able to cleave the phosphodiester linkage joining the terminal ^{6S}G and the pCp label. Radiolabeled standard ^{6S}Gp was mixed with T2-digested RNA radiolabeled during transcription with α-[^{32}P] UTP. The resulting two-dimensional TLC standard revealed five spots, corresponding to Ap, Cp, Gp, Up, and ^{6S}Gp (Figure 4d). The ^{6S}Gp spot was not apparent in digestion of unrelated radiolabeled RNA containing no 3′ ^{6S}G. Two-dimensional TLC showed a single radiolabeled spot for all four ribozymes treated with the same pCp ligation procedure as used to generate the ^{6S}Gp standard (Figure 4e). Mixing the ribozyme sample with the five nucleotide references indicated that the ribozyme-dependent spot comigrated precisely with the ^{6S}Gp standard and resulted in a spot ∼2 times more radioactive than the initial reference marker. As this TLC system has previously been shown to resolve quite similar nucleotides[17] (for example, pseudouridine from uridine), we are reasonably confident that an N-9 linkage is in fact being produced (note also 6-thio-9-methylpurine did not react detectably, see below).

Kinetics. Both RA and MA were subjected to more detailed kinetic analysis since both isolates had the highest frequency of occurrence and initial velocity compared to other sequences found in both pools. ^{6S}Gua, which is sparingly soluble in water, was titrated over its solubility range. First-order rate constants were extracted as a function of ^{6S}Gua concentration by fitting simultaneously at least five independent time courses at a particular substrate concentration (see methods). The resulting rates were then fit to the Michaelis−Menten equation, revealing an apparent K_m of 78 ± 11 μM and a $k_{cat\ app}$ of 0.018 ± 0.007 min^{-1} for the RA isolate (Figure 5) with an apparent efficiency ($k_{cat\ app}/K_m$) of 230 M^{-1} min^{-1}. In contrast, the reaction rate of MA was directly proportional to ^{6S}Gua concentration, giving an apparent efficiency of 284 M^{-1} min^{-1}. The linear rate dependence with substrate concentration observed for MA makes it unlikely that the rate plateau observed for RA at high ^{6S}Gua concentrations was due to nonspecific ribozyme inhibition such as aggregation. The uncatalyzed rate of purine nucleotide synthesis was undetectable when a short radiolabeled RNA (sequence 5′-AAC) derivatized with PRPP was incubated for as long as 8 days with ^{6S}Gua as judged by APM gel shift (less than five parts in 10^4 detected, uncatalyzed rate $< 2 \times 10^{-4}$ M^{-1} min^{-1}).

Substrate Specificity. The substrate preference of both MA and RA was examined using 15 different sulfur-containing purine and pyrimidine compounds. Saturated solutions of each compound were prepared in standard incubation buffer, and four compounds showed weak activity with either MA or RA as indicated by APM gel shift (6-thiopurine, 2-methyl-6-thiopurine, 2,6-dithiopurine, and 6,8-dithiopurine). As many of these compounds were suspected to be contaminated with 6-thiopurine, they were HPLC purified using a linear acetonitrile gradient on a C18 column. The reaction efficiency of these compounds was then compared to ^{6S}Gua by incubating the purified material at a uniform concentration of 0.26 mM with the ribozymes. Only 6-thiopurine reacted at all close to that of ^{6S}Gua with MA displaying an apparent efficiency of ∼0.1−0.5 M^{-1} min^{-1}, while RA was 5−10-fold slower still (assuming a linear reaction rate, Figure 6). MA also reacted with 2-methyl-6-thiopurine and 6,8-dithiopurine, but about 1−2 and 4−10 times slower than with 6-thiopurine, respectively, while RA did not react detectably with either compound. All four product bands had unique APM gel shifts. Measurements with radiolabeled guanine, though highly desirable, were not attempted due to the very low solubility of guanine and the relatively small amount of ribozyme that could be used to perform such an assay.

Discussion

How is the structural information encoded in an RNA's primary sequence able to influence the outcome of an in vitro selection? We observed a bias resulting from imposing the secondary structure of the family A pyrimidine nucleotide synthase onto one of our high-diversity pools. While the overall structure never appeared to be preserved, nearly all of the round 10 biased pool isolates contain the hairpin-stem V of the family A pyrimidine nucleotide synthase.[11] The conservation of this stem was expected due to chance 46% of the time (excluding wobbles) implying its high conservation in nearly every isolate (18/20) conferred an overall benefit to ribozyme function. This hairpin was also found in the fastest most populated families MA and MB. In these 18 families, 12 contained a UCUUU motif (of which six can be extended to AGGCGUUCUUU) immediately upstream of the helix. The distinctly different sequence of these families in other regions makes it likely that these families are independent representatives of one overarching motif class defined at least partially by a UCUUU-hairpin motif. This motif was not found in the random pool even though the random pool contained RNA with terminal nucleotide sequences identical to that of the biased pool. This suggests that the motif is a direct consequence of the imposed secondary structure. It is curious in this regard that while the structural bias increased the frequency of occurrence for the UCUUU−hairpin loop motif, it did not result in ribozymes notably more efficient than

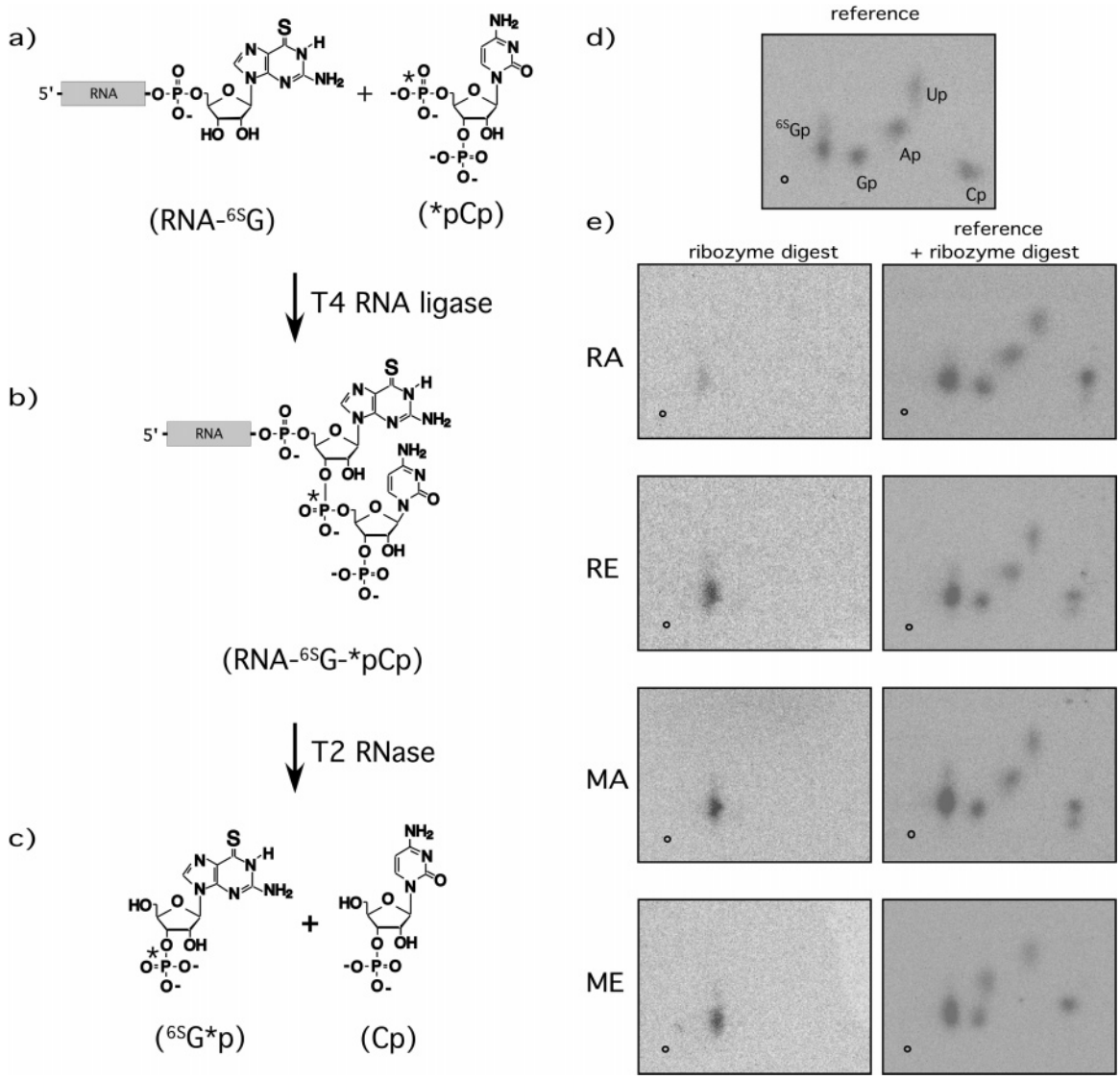


Figure 4. Two-dimensional TLC analysis of the reaction products resulting from four different purine nucleotide synthase reactions. (a) Reacted RNA or an RNA construct synthesized so as to have a terminal 3′ p^{6S}G is derivatized with pCp (^{32}P-labeled phosphate denoted by asterisk) using T4 RNA ligase. (b) The radiolabeled, thiol-containing material is purified on an APM gel. (c) The recovered material is digested into mononucleotides using ribonuclease T2. (d) TLC showing a mixture of known 3′ radiolabeled mononucleotides. From left to right spots are as follows: ^{6S}Gp, Gp, Ap, Up, and Cp. (e) Ribozyme isolates reacted with ^{6S}Gua and treated as described in panels a−c are shown in the left column. A mixture of ribozyme digests and marker mononucleotides are shown in the right column. The first axis is vertical. TLC origins are indicated by black open circles.

those found in the random pool (see Figure 3 to compare final pool activities).

A selection involving a small substrate such as ^{6S}Gua represents a significantly different catalytic challenge than presented by selections involving substrates capable of high-affinity binding through Watson−Crick pairing.[7] With small substrates, even though both the substrate concentration and the incubation time can be experimentally varied, at short enough times and low enough substrate concentrations only the ratio between k_{cat} and K_m directly effect ribozyme survival. It is therefore not unexpected that ribozymes resulting from a given pool should have roughly similar purine nucleotide synthesis efficiencies (but does leave unexplained why separate pools would have the same efficiency). What is interesting is that two ribozymes with similar efficiencies appear to have exploited quite different catalytic strategies. RA binds ^{6S}Gua tightly ($K_m \approx 80\ \mu$M), discriminates well against quite similar compounds, and performs a slow chemical step ($k_{cat\ app} \approx 0.02\ min^{-1}$). In contrast, MA has slightly worse substrate discrimination than RA, binds its substrate with low affinity, and by implication has a $k_{cat\ app}$ significantly higher than RA.

The ability of RA and MA to distinguish between many closely related ^{6S}Gua derivatives indicates that appreciable contacts must be formed with the substrate at some point during the course of glycosidic bond formation. While the MA and RA purine synthases generally did not react detectably with a range of 6-thioguanine derivatives, they were somewhat tolerant of substitutions at the 2 position reacting thousands of times slower with, for example, 6-thiopurine (Figure 6). It is curious in this regard that the protein enzyme HGPTase is also unable to discriminate strongly between hypoxanthine and guanine and normally accepts both substrates.[22] A similar pattern is displayed by a naturally occurring guanine aptamer found in the *xpt-pbuX* mRNA[23] and an artificially selected guanine aptamer;[24] both

(22) Craig, S. P.; Eakin, A. E. *J. Biol. Chem.* **2000**, *275*, 20231−20234.

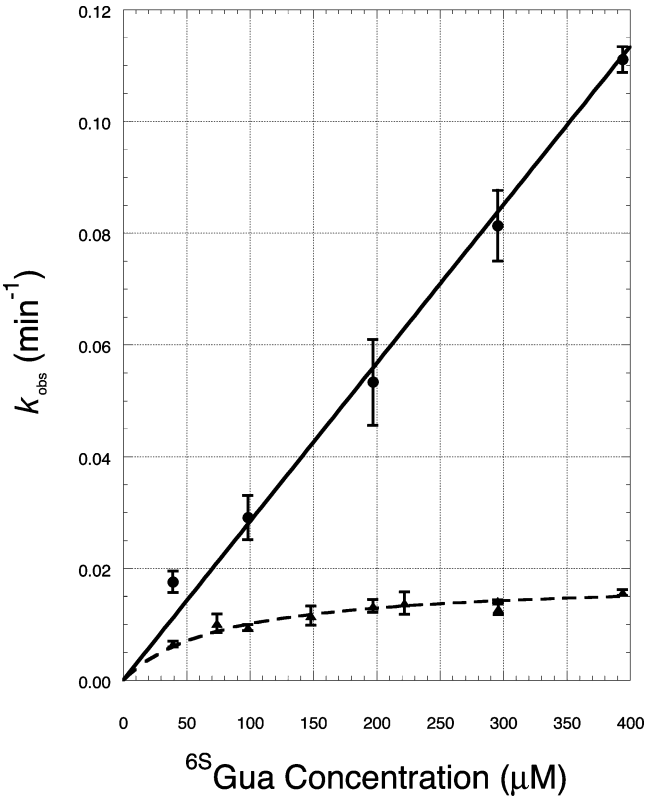


Figure 5. Nucleotide synthesis rate (k_{obs}) for two purine nucleotide synthase isolates MA (circles) and RA (triangles) as a function of ^{6S}Gua concentration. MA had kinetics which fit a straight line, indicating an apparent efficiency of 284 M^{-1} min^{-1}. RA fit well the to Michaelis−Menten equation, which indicated a K_m of 78 μM and $k_{cat\ app}$ of 0.02 min^{-1}. The linear behavior of MA suggests that aggregation or some other form of general ribozyme inhibition is not responsible for the RA kinetics.

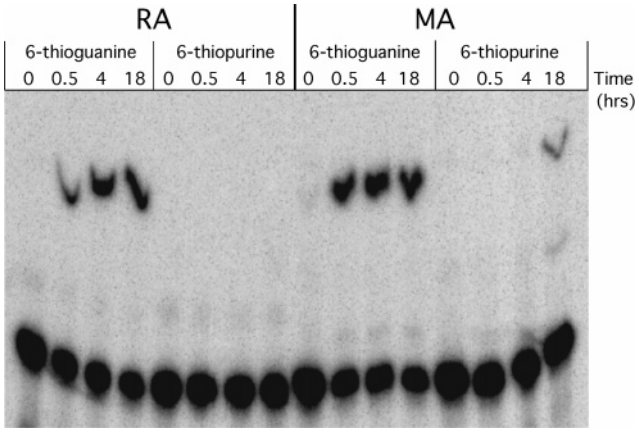


Figure 6. Ribozyme incubation with 6-thioguanine or 6-thiopurine. Reactions were performed for both random (RA) and biased pool isolates (MA). Substrate concentrations were held at 0.26 mM. Incubation with ^{6S}Gua for 30 min, 4 h, and 18 h resulted in 13%, 30%, and 30%, respectively, of ribozyme RA reacting and 26%, 30%, and 29%, respectively, for MA. Incubation of both ribozymes with 6-thiopurine for 18 h resulted in signals of ≈0.5% for RA and 3.6% for MA.

discriminate weakly against xanthine and hypoxanthine. While these correlations may be coincidental, the isolation of two artificial ribozymes that share the same general recognition characteristics found with a variety of naturally occurring guanine aptamers, as well as protein enzymes that recognize guanine, suggest that this pattern of recognition may be universal.

The purine nucleotide synthases were considerably more efficient than the equivalent pyrimidine nucleotide synthase previously isolated. Even after performing further selections which improved the family A pyrimidine nucleotide synthase by 35-fold,[11] the purine synthases we found directly from both random and biased pools were still 2-fold faster (50−100 times faster than the initial pyrimidine nucleotide synthase isolates). The most obvious explanation for this rate difference is that ^{4S}Ura is simply harder to recognize than ^{6S}Gua. As uracil aptamers are currently unknown, this assumption is difficult to access objectively but appears reasonable given the superior stacking potential expected from purine substrates.

There is, however, a second possibility. The chemistry of glycosidic bond formation is surprisingly and dramatically influenced by nucleobase composition. Purine nucleotides are much more thermodynamically stable than pyrimidine nucleotides. The synthesis of OMP by EC 2.4.2.10 has a ΔG of nearly zero,[25,26] while AMP or GMP synthesis by EC 2.4.2.7 or 2.4.2.8 has a ΔG of ∼−7 kcal/mol.[27−29] While the reason for this large free energy difference does not appear to have been well explored, it is striking that the hydrolysis of PRPP has a ΔG of −8.4 kcal/mol,[30] only 1.4 kcal/mol more negative than that observed for purine nucleotide synthesis. Kinetically the uncatalyzed cleavage of a purine glycosidic bond at low pH can be estimated to occur at a rate 10^5−10^6 times faster than the equivilent pyrimidine glycosidic linkage (calculated using deoxyribose nucleosides which are much more labile than ribose nucleotides in acid[31,32]). These thermodynamic and kinetic statements are consistent with the finding that purine nucleosides and not pyrimidine nucleosides can be synthesized by dehydration.[5] Taken together, these observations suggest that the uncatalyzed rate of purine nucleotide synthesis starting from PRPP is likely to be much higher than for pyrimidine nucleotide synthesis even though neither rate could be detected in our hands.[9]

The pertinent question may therefore be not why is purine nucleotide synthesis superior to pyrimidine nucleotide synthesis, but why is it only 50−100 fold better? It appears most likely that RNA has difficulty precisely positioning its limited range of functional groups in a catalytic pocket small enough to optimize both small substrate binding and transition-state stabilization simultaneously. This is generally consistent with our observation that both random and structurally biased pools resulted in ribozymes with very similar efficiencies (Figure 3), even though the initial amount of structural information available in the two pools differed. More specifically, our kinetic data suggests that RA and MA have considerably different substrate

(23) Mandal, M.; Boese, B.; Barrick, J. E.; Winkler, W. C.; Breaker, R. R. *Cell* **2003**, *113*, 577−586.
(24) Kiga, D.; Futamura, Y.; Sakamoto, K.; Yokoyama, S. *Nucleic Acids Res.* **1998**, *26*, 1755−1760.
(25) Tavares, A.; Lee, C. S.; Osullivan, W. J. *Biochim. Biophys. Acta* **1987**, *913*, 279−284.
(26) Bhatia, M. B.; Vinitsky, A.; Grubmeyer, C. *Biochemistry* **1990**, *29*, 10480−10487.
(27) Kornberg, A.; Lieberman, I.; Simms, E. S. *J. Biol. Chem* **1955**, *215*, 417−427.
(28) Dewolf, W. E.; Emig, F. A.; Schramm, V. L. *Biochemistry* **1986**, *25*, 4132−4140.
(29) Xu, Y.; Eads, J.; Sacchettini, J. C.; Grubmeyer, C. *Biochemistry* **1997**, *36*, 3700−3712.
(30) Frey, P. A.; Arabshahi, A. *Biochemistry* **1995**, *34*, 11307−11310.
(31) Zoltewicz, J. A.; Clark, D. F.; Sharpless, T. W.; Grahe, G. *J. Am. Chem. Soc.* **1970**, *92*, 1741−1750.
(32) Garrett, E. R.; Seydel, J. K.; Sharpen, A. J. *J. Am. Chem. Soc.* **1966**, *31*, 2219−2227.

recognition strategies and yet have very similar overall efficiencies. It has been hypothesized that a general relationship between informational complexity (the amount of information required to specify an RNA structure) and function may exist.[33,34] We suggest that the precise positioning of functional groups required for small molecule catalysis conflicts with the scale of RNA's relatively large monomers and provides a natural basis to relate structural information to catalytic function. This complexity−function relationship might be expected to saturate or change its character if this conflict is in fact the limiting process that governs the emergence of small molecule RNA catalysts. It will therefore be of considerable interest to study purine and pyrimidine nucleotide synthase ribozymes in greater structural and kinetic detail in order to explore this important aspect of RNA catalysis.

Acknowledgment. We thank E. Chiu for assistance with sequencing, M. Tracey for NMR spectra, G. Taylor for assistance with RNA folding software, and the Unrau laboratory for careful reading of the manuscript. This work was supported by funds from the National Science and Engineering Research Council of Canada and the Michael Smith Foundation for Health Research.

Supporting Information Available: Alignments of round 10 isolates from both the biased and random sequence pools. This material is available free of charge via the Internet at http://pubs.acs.org.

JA045387A

(33) Carothers, J. M.; Oestreich, S. C.; Davis, J. H.; Szostak, J. W. *J. Am. Chem. Soc.* **2004**, *126*, 5130−5137.
(34) Szostak, J. W. *Nature* **2003**, *423*, 689−689.