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 Kₘ 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. 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. As purine and pyrimidine bases are known to be pyridine nucleotide cofactors, and the amino acids histidine and tryptophan, these ribozymes displayed high specificity for their 6S Ura substrate and had apparent efficiencies (k_{cat} 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. 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 affinities and catalytic parameters.

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. 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. 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). These ribozymes displayed high specificity for their 4S Ura substrate and had apparent efficiencies (k_{cat} K_{m}) in the 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. 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 affinities and catalytic parameters.

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

**Methods**

**AppRpp and App^6^G Synthesis.** The synthesis of adenylated PRPP (AppRpp) was as previously described. Adenylated 6-thioguanosine-5′-monophosphate (App^6^G) was synthesized by reacting 100 mM GMP with 50 mM adenosine 5′-phosphorimidazolide in the presence of 100 mM MgCl₂ and 200 mM HEPEs (pH 7.4) at 50 °C for 3 h followed by HPLC purification.

**Purine Nucleotide Synthesis**

(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.

**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.


biphosphate (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-[32P] 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/mL T2, Sigma) for 2 h at 37 °C. A 6-thioguanosine 3′-mononucleotide (6Gp) control was synthesized by ligation of the marker RNA (derivatized with 3′ pG) 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 (NH₄)₂SO₄:H₂O. The first dimension was developed with 80% ethanol, second with 40:1 saturated (NH₄)₂SO₄:2-propanol; both solvents were supplemented with 100 mM β-mercaptoethanol. Samples were spotted 1 cm in from the corner of the TLC plates.

**Kinetic Analysis.** Kinetic studies were performed in incubation buffer supplemented with MgCl₂ 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 6Gua 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

\[ F = \beta (1 - e^{-kt_{obs}}) \]

using the program KaleidaGraph (Synergy Software), \( F \) being the fraction reacted at time \( t \), \( k_{obs} \) the first-order apparent rate constant, and \( \beta \) the total fraction able to react. The resulting rates were fit to the Michaelis–Menten equation

\[ k_{obs} = k_{cat}\text{app}[6S\text{Gua}]/(K_m + [6S\text{Gua}]) \]

A weighted error analysis was performed to obtain \( k_{cat}\text{app} \) and \( K_m \).

**Analogues.** Purine analogues 6-thioguanine \( \epsilon (347 \text{ nm, pH 1}) = 20 \, 900 \text{ M}^{-1} \cdot \text{cm}^{-1} \) \text{,} \ 6\text{-thiopurine} \( \epsilon (325 \text{ nm, pH 1}) = 20 \, 500 \text{ M}^{-1} \cdot \text{cm}^{-1} \) \text{,} \ 2\text{-methyl-6-thiopurine assumed} \( \epsilon (330 \text{ nm, pH 1}) = 20 \, 000 \text{ M}^{-1} \cdot \text{cm}^{-1} \), \text{ 6,8-dithiopurine} \( \epsilon (358 \text{ nm, pH 1}) = 27 \, 800 \text{ M}^{-1} \cdot \text{cm}^{-1} \) \text{,} \text{ 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 purified using a 6% APM gel in order to isolate the thiol-containing material. The recovered RNA was digested into 3′-mononucleotides (6S Gua) control was synthesized by ligation of the primer and thus would be expected to be conserved by chance 46% of the time. Only one isolate (MF) 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 indicated that this sequence was likely to adopt a fold considerably different from that of the family A nucleotide synthase.

**Results.** 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 × 10¹³ 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. 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 6Gua substrate using an APM gel shift strategy (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. After 6 rounds of in vitro selection, 6Gua-dependent ribozymes having the same mobility as a RNA-p6G 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 6Gua substrate (see methods). By round 10, nearly 1% of each pool had reacted after 15 s of incubation. Both pools of 6G 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 6G synthases of both pools seemed to have reached maximum reaction rates by round 9, slightly faster than the previously selected 6U 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 six residues to AGGCUGUCUUU, a UCUUU sequence motif with an 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 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

---

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 MgCl2. 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′-[32P]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 2. Overall in vitro selection scheme for purine nucleotide synthase ribozymes.](image-url)
than the initial reference marker. As this TLC system has also 6-thio-9-methylpurine did not react detectably, see below). We are reasonably confident that an N-9 linkage is in fact being produced (note for example, pseudouridine from uridine), we are reasonably

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. Reaction rates were calculated by dividing the observed first-order reaction rates by the substrate concentration. Percentage APM polyacrylamide gel, we concluded that the ribozyme-dependent spot comigrated precisely with the ribozyme sample with the five nucleotide references 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 6\(^6\)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 6\(^6\)Gua with MA displaying an apparent efficiency of \(\sim 0.1–0.5 \text{ M}^{-1} \text{min}^{-1}\), 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 (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 AGGCGUU/UCUU) 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 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 \(6\(^6\)\)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 \(6\(^6\)\)Gua as judged by APM gel shift (less than five parts in \(10^4\) detected, uncatalyzed rate \(< 2 \times 10^{-4} \text{ M}^{-1} \text{min}^{-1}\)).

**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 \(6\(^6\)\)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 \(6\(^6\)\)Gua with MA displaying an apparent efficiency of \(\sim 0.1–0.5 \text{ M}^{-1} \text{min}^{-1}\), 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.

**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. Kinetic analysis since both isolates had the highest frequency of occurrence and initial velocity compared to other sequences found in both pools. \(\text{6}\(^6\)\)Gua, which is sparingly soluble in water, was titrated over its solubility range. First-order rate constants were extracted as a function of \(6\(^6\)\)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 \(\mu\)M and a \(k_{\text{cat app}}\) of 0.018 ± 0.007 min\(^{-1}\) for the RA isolate (Figure 5) with an apparent efficiency (\(k_{\text{cat app}}/K_m\)) of 230 M\(^{-1}\) min\(^{-1}\). In contrast, the reaction rate of MA was directly proportional to \(6\(^6\)\)Gua concentration, giving
those found in the random pool (see Figure 3 to compare final pool activities).

A selection involving a small substrate such as $^{65}$Gua represents a significantly different catalytic challenge than presented by selections involving substrates capable of high-affinity binding through Watson-Crick pairing. 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 $k_{\text{cat}}$ 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 $^{65}$Gua tightly ($K_m \approx 80 \mu M$), discriminates well against quite similar compounds, and performs a slow chemical step ($k_{\text{cat, app}} \approx 0.02 \text{ min}^{-1}$). In contrast, MA has slightly worse substrate discrimination than RA, binds its substrate with low affinity, and by implication has a $k_{\text{cat, app}}$ significantly higher than RA.

The ability of RA and MA to distinguish between many closely related $^{65}$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. A similar pattern is displayed by a naturally occurring guanine aptamer found in the xpt-pbuX mRNA and an artificially selected guanine aptamer, both

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' pCpG 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 AFM 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: $^{32}$Gp, Gp, Ap, Up, and Cp. (e) Ribozyme isolates reacted with $^{65}$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.
discriminate weakly against xanthine and hypoxanthine. While these correlations may be coincidental, the isolation of two ribozymes suggests that RNA catalysis might be possible with a variety of naturally occurring purine substrate.

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 105–106 times faster than the equivalent pyrimidine glycosidic linkage (calculated using deoxyribose nucleosides which are much more labile than ribose nucleotides in acid31,32). 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

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.\textsuperscript{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
Supplementary Figure. Alignments of round 10 isolates from both the biased and random sequence pools. a. Families resulting from the biased pool. The sequences have been aligned with respect to the lightly mutagenized secondary structure elements of the family A pyrimidine nucleotide synthase which was used to design the pool. The majority of sequences isolated contain a UCUUU-hairpin-loop motif upstream of their 3’ primer binding sequence. b. Families resulting from the random pool. Even through this pool had an identical 3’ primer binding sequence, a common motif of the sort found in the biased pool is not observed. The bracketed number following a sequence in either alignment indicates the number of repeats observed for that isolate. Unnamed sequences were isolated only once.