# Recognition of Nucleoside Triphosphates during RNA-Catalyzed Primer Extension[†]

Margaret E. Glasner, Catherine C. Yen, Eric H. Ekland,[‡] and David P. Bartel*

*Whitehead Institute for Biomedical Research and Department of Biology, Massachusetts Institute of Technology,
9 Cambridge Center, Cambridge, Massachusetts 02142*

ABSTRACT: In support of the idea that certain RNA molecules might be able to catalyze RNA replication, a ribozyme was previously generated that synthesizes short segments of RNA in a reaction modeled after that of proteinaceous RNA polymerases. Here, we describe substrate recognition by this polymerase ribozyme. Altering base or sugar moieties of the nucleoside triphosphate only moderately affects its utilization, provided that the alterations do not disrupt Watson−Crick pairing to the template. Correctly paired nucleotides have both a lower $K_m$ and a higher $k_{cat}$, suggesting that differential binding and orientation each play roles in discriminating matched from mismatched nucleotides. Binding of the pyrophosphate leaving group appears weak, as evidenced by a very inefficient pyrophosphate-exchange reaction, the reverse of the primer-extension reaction. Indeed, substitutions at the $\gamma$-phosphate can be tolerated, although poorly. Thio substitutions of oxygen atoms at the reactive phosphate exert effects similar to those seen with cellular polymerases, leaving open the possibility of an active site analogous to those of protein enzymes. The polymerase ribozyme, derived from an efficient RNA ligase ribozyme, can achieve the very fast $k_{cat}$ of the parent ribozyme when the substrate of the polymerase (GTP) is replaced by an extended substrate (pppGGA), in which the GA dinucleotide extension corresponds to the second and third nucleotides of the ligase. This suggests that the GA dinucleotide, which had been deleted when converting the ligase into a polymerase, plays an important role in orienting the 5′-terminal nucleoside. Polymerase constructs that restore this missing orientation function should achieve much more efficient and perhaps more accurate RNA polymerization.

The RNA world hypothesis proposes that early life relied on catalytic RNAs rather than protein enzymes for catalysis (*1*). Much of the appeal of this theory comes from the notion that ribozymes, which could have served as their own genes, would have been much simpler to duplicate than proteins (*2−5*). Hence, a vital component of the RNA world scenario is an RNA polymerase ribozyme responsible for replicating the essential ribozymes of early life (including itself). Whether there might be RNA sequences that can promote RNA polymerization with sufficient accuracy and efficiency for RNA self-replication is one of the fundamental unanswered questions related to the RNA world hypothesis (*1*, *6*). Finding such a sequence would support the idea of the RNA world and provide a key component for the synthesis of simple living cells (*7*).

The class I ligase (Figure 1A) is a promising starting point for efforts to develop a self-replicating ribozyme. This ribozyme ligates two RNAs, generating a 3′,5′-phosphodiester linkage and releasing pyrophosphate, a reaction similar to that of RNA and DNA polymerases (*8*). Engineered forms of the ligase can polymerize short segments of RNA (Figure 1C). In the presence of appropriate RNA templates and nucleoside triphosphates (NTPs),[1] such ribozymes can extend an RNA primer by three or, in special cases, more nucleotides (*9*). The overall Watson−Crick error rate is about 10%; i.e.,

on average, 88−92% of the extension is by the Watson−Crick match to the template, whereas an average of 3−4% of the extension is by each of the three mismatches. Although an error rate of 10% is too large to support RNA self-replication, it is significantly smaller than the 60% error rate calculated solely on the basis of the observed stabilities of Watson−Crick matches relative to mismatches at the ends of helices (*6*). The ribozyme is particularly accurate at C and G template residues; at a C template residue, GTP is utilized 1000 times more efficiently than CTP, and at a G, CTP is utilized 600 times more efficiently than GTP (*9*).

Polymerase-like ribozymes derived from naturally occurring ribozymes, such as self-splicing introns, have not demonstrated comparable Watson−Crick fidelity (*6*, *10*). Another advantage of the ligase derivatives is that polymerization yields new RNA linkages, whereas the self-splicing derivatives catalyze disproportionation reactions, in which one RNA becomes longer at the expense of another. Furthermore, pyrophosphate activation provides a chemical driving force (2.1 kcal mol$^{-1}$ nucleotide$^{-1}$) for the reaction (*11*). This disfavors the reverse reaction and might eventually be harnessed to promote translocation and displacement of a noncoding strand or intramolecular template structure (*6*).

Here we describe the recognition and utilization of NTPs during RNA-catalyzed primer extension. We investigated the

---

* Corresponding author. Fax: (617)258-6768. E-mail: dbartel@wi.mit.edu.

[‡] Present address: Department of Microbiology and Immunology, University of California, San Francisco, CA 94143-0414.

[1] Abbreviations: NTP, nucleoside triphosphate; PP$_i$, inorganic pyrophosphate; EDTA, ethylenediaminetetraacetic acid; 3′-dATP, cordycepin triphosphate; MES, 2-(*N*-morpholino)ethanesulfonic acid; BES, *N,N*-bis(2-hydroxyethyl)-2-aminoethanesulfonic acid; EPPS, *N*-(2-hydroxyethyl)piperazine-*N*′-(3-propanesulfonic acid).

basis for the high Watson−Crick discrimination at G and C template residues and the degree to which the primer-extension reaction is reversible. We also discovered the importance of a GA dinucleotide segment for substrate utilization, explaining why deleting this dinucleotide from the parent ligase to create the primer-extension construct resulted in a large decrease in activity. Our results reinforce the promise of the polymerase ribozyme as a starting point for developing a self-replicating ribozyme and identify specific elements of substrate binding and utilization to be targeted for improvement during future design and evolution efforts.

## MATERIALS AND METHODS

*Ribozyme and Substrates.* Ribozymes were prepared by runoff transcription of a PCR template as described (*12*). The oligonucleotide substrate analogous to a primer was a synthetic DNA−RNA chimera (5′-aaaCCAGUC; lower case indicates DNA). This primer was 5′-radiolabeled using T4 polynucleotide kinase and [$\gamma$-$^{32}$P]ATP. G(5′)ppp(5′)G, dGTP, ddGTP, and ultrapure grade ATP, CTP, and GTP were purchased from Pharmacia. GDP, guanosine diphospho-mannose, and nicotinamide guanine dinucleotide were from Sigma; $\alpha$-GTP-phosphorothioates were from Dupont/NEN; and 7-deaza-dGTP was from Amersham. Polynucleotide substrates were generated using T7 RNA polymerase in abortive in vitro transcription reactions in which only GTP and ATP (for pppGGAA and pppGAAA) or GTP, CTP, and UTP (for pppGGCU) were supplied. DNA oligonucleotide templates were GGTGTTCC*TATAGTGAGTCGTATTA*CGC, TCTCATTTC*TATAGTGAGTCGTATTA*CGC, and TGGT-TCCAGTCGTATAGCC*TATAGTGAGTCGTATTA*CGC (italics denotes T7 polymerase promotor) for the pppGGAA, pppGAAA, and pppGGCU transcriptions, respectively. To body-label transcription products, [$\alpha$-$^{32}$P]GTP was included. The pppGGA from abortive in vitro transcription was eluted from a gel and desalted on a Sep-Pak Classic C18 cartridge (Millipore). Analysis by MALDI-TOF mass spectrometry (PerSeptive Biosystems) was consistent with the major purified product being pppGGA, with a detectable amount of contaminating ppGGA.

*Primer-Extension Assays.* All reactions were performed under single-turnover conditions, with ribozyme in sufficient excess to saturate the labeled substrate. Typical reactions included 1 $\mu$M ribozyme and 0.5 $\mu$M primer. Prior to initiating the reaction, the ribozyme was heated in water (2−3 min at 80 °C) and cooled (2 min at 22 °C). Reactions were initiated with the simultaneous addition of salt, buffer, primer, and the triphosphate-containing substrate. Unless stated otherwise, reactions were in 30 or 50 mM Tris, pH 8.0, 60 mM MgCl$_2$, 200 mM KCl, and 0.6 mM EDTA at 22 °C (or room temperature for incubations less than 1 min). Because NTPs chelate Mg$^{2+}$ and the reaction rate is known to be sensitive to Mg$^{2+}$ concentration, the reaction mixture was supplemented with MgCl$_2$ (at a concentration stoichiometric with that of the NTP) to maintain a free Mg$^{2+}$ concentration of 60 mM. Reactions were stopped by addition of at least 1 vol of 8 M urea/120 mM EDTA and run on denaturing sequencing gels. Radiolabeled reactants and products were quantified with a Fujix BAS 2000 phosphorimager. Nucleotide-analogue experiments were performed with a ribozyme that differed from that used in the rest of the experiments: A18−A20 was changed to CGU, A67 was changed to G,

and U78 was changed to C. This minor alteration of the ribozyme sequence is not expected to change the relative efficiencies of primer extension.

Rates were determined using

$$F = 1 - e^{-kt} \tag{1}$$

where $t$ equals time, $k$ is the observed rate of catalysis, and $F$ is the fraction of radiolabeled substrate converted to product. Analysis was complicated by a brief lag (<5 s) in the course of the reaction. This lag was not observed with the parent ligase (*13*). A similar phenomenon was encountered with the Group II intron and was attributed to a Mg$^{2+}$-induced conformation change (*14*). Rates were determined using the part of the curve that best fit eq 1.

Apparent second-order rate constants for polynucleotide substrates were determined by first calculating the fold preference for the polynucleotide substrate compared to GTP using

$$\text{preference} = \frac{P-X}{P-G} \times \frac{GTP_o}{pppX_o} \tag{2}$$

where P−X is the fraction of primer extended using the polynucleotide substrate, P−G is the fraction of primer extended using GTP, GTP$_o$ is the initial GTP concentration, and pppX$_o$ is the initial polynucleotide substrate concentration. The preference value was then multiplied by the apparent second-order rate constant for GTP addition (590 M$^{-1}$ min$^{-1}$) to give the apparent second-order rate constant for the polynucleotide substrate.

The substrate for the reverse reaction, aaaCCAGUCGGdA (lowercase, DNA; dA, 3′-deoxyadenosine), represented the product of a polynucleotide extension reaction in which the standard primer had been extended using pppGGdA. This oligonucleotide was generated by extending aaaCCAGUCGG (Dharmacon Research) with a 3′-deoxyadenosine nucleotide, using cordycepin triphosphate (3′-dATP) and yeast poly(A) polymerase (Amersham). To radiolabel the oligonucleotide, [$\alpha$-$^{32}$P]-3′-dATP was used.

## RESULTS AND DISCUSSION

This study focuses on a ribozyme referred to as the "GTP ribozyme," which extends a primer by a single nucleotide using GTP (Figure 1D). The GTP ribozyme differs from the polymerase construct (Figure 1C) by the covalent attachment of the template. Constructs with this intramolecular configuration are more convenient to produce and characterize as well as about 10 times more efficient than those with the intermolecular configuration. While the ribozyme core can accommodate templates coding for multiple nucleotides, utilization of templates coding for a single nucleotide is most efficient. Thus, our prototype ribozyme codes for primer extension by a single nucleotide. While the GTP ribozyme was chosen for maximal convenience and efficiency, there is every indication that with other templates and configurations the ribozyme core will display the same properties and preferences, albeit with slower overall rates (*9*). Using the GTP ribozyme, we investigated the efficiency of RNA-catalyzed primer extension with respect to the three components of the NTP substrate: the reactive phosphate, the nucleoside, and the pyrophosphate leaving group.
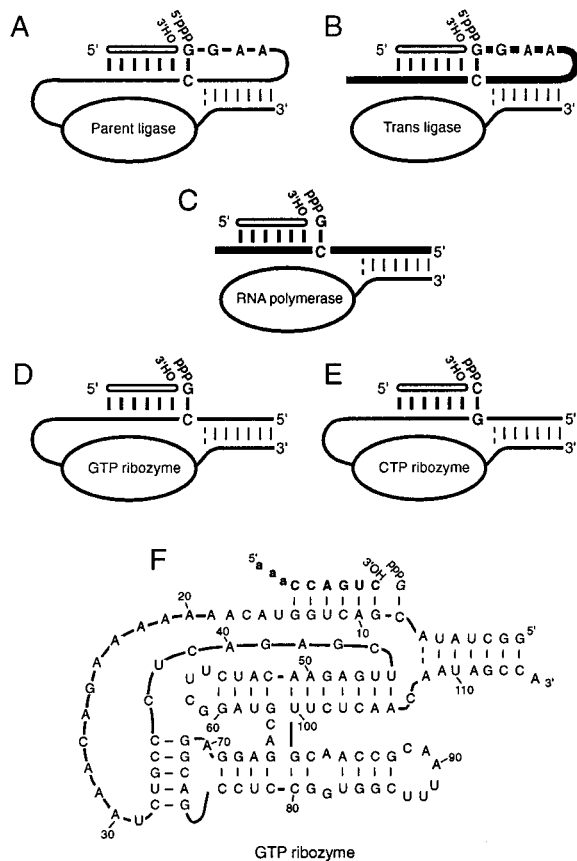
FIGURE 1: Design of RNA polymerase-like ribozymes. In the schematic representations (A–E) the core of the ribozyme is indicated by an oval, and Watson–Crick pairing involving the substrates is indicated by vertical dashes (with a proposed A:A interaction indicated by a double dash). (A) The parent class I ligase (solid line, GenBank U26413) hybridized to the primer (open line) (*8*, *12*). The 5′-terminal residues (pppGGAA) are specified, as is the cytidine that pairs with the first nucleotide of the ribozyme. (B) The trans ligase, a multiple-turnover ribozyme (*8*, *13*). In this ribozyme, the former 5′-leader of the ligase (heavy line) is detached from the ribozyme. (C) The RNA polymerase (*9*). This ribozyme requires three components: a template (heavy line), primer (open line) and NTPs. (D) The GTP ribozyme (GenBank AF303466). This ribozyme is similar to the polymerase ribozyme except that it is covalently attached to a template specifying primer extension using a single GTP. (E) The CTP ribozyme. To specify extension using CTP, the template base (residue 8) was changed from C to G. (F) Sequence and secondary structure of the GTP ribozyme.

*The Reactive Phosphate.* The primer-extension reaction centers on the α-phosphate of the NTP. Substitution of either of the nonbridging oxygens of this phosphate decreased the efficiency of primer extension (Figure 2). Substitution of the pro-$R_P$ oxygen decreased the apparent second-order rate constant ($k_{cat}/K_m^{GTP}$) of primer extension by more than 1000-fold, while substitution of the pro-$S_P$ oxygen had a more modest, 37-fold effect. Some decrease in reactivity would be expected from these substitutions due to their destabilization of the associative transition state (*15*). For instance, a 60-fold decrease in rate for DNA polymerase upon substitution of the pro-$S_P$ oxygen is attributed to this elemental effect on the transition state (*16*). This effect on intrinsic reactivity could also fully account for the 37-fold effect seen with the pro-$S_P$ substitution on the activity of the GTP ribozyme, provided that the chemical step is rate limiting during primer extension.
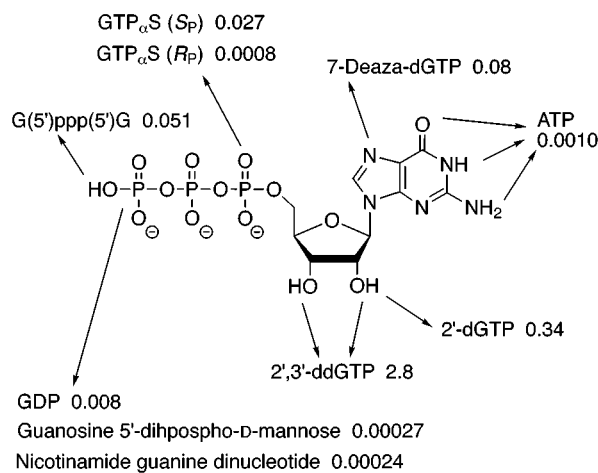


FIGURE 2: Primer extension using GTP analogues. Primer extension was assayed using 1 mM of the indicated analogue (except GpppG, which was used at 0.5 mM to compensate for its 2-fold symmetry). Observed rates are reported relative to that using 1 mM GTP.
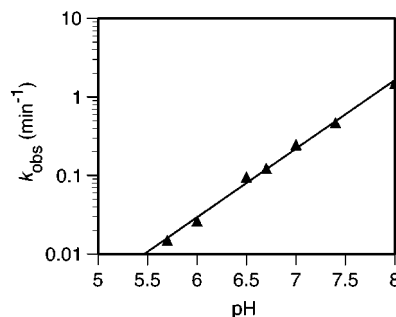


FIGURE 3: pH dependence of the GTP ribozyme. The reaction was carried out with 9 mM GTP in standard conditions, except Tris buffer was replaced by 50 mM MES (pH 5.7–6.5), BES (pH 6.7–7.4), or EPPS (pH 8.0). The slope of the line is 0.9.

To investigate whether the chemical step was rate limiting, the pH dependence of primer extension was examined. The logarithm of the rate as a function of pH is a line with a slope of 0.9 (Figure 3), indicating that a single deprotonation is rate limiting. Such behavior for ribozyme reactions involving phosphoryl transfer is taken as evidence of a rate-limiting chemical step, because the attacking hydroxyl must be deprotonated during this step (*15*, *17–19*). Thus, the thio effect seen upon substitution of the pro-$S_P$ oxygen can be attributed solely to the elemental effect on the reactivity of the phosphate reaction center.

The effect of substituting the pro-$R_P$ oxygen was much more dramatic (1000-fold), suggesting an important ribozyme contact or metal coordination to the pro-$R_P$ oxygen. Intriguingly, the steriospecificity of the thio effects matches that of protein-enzyme polymerases (*20*). Both the ribozyme and protein enzymes suffer mere elemental effects upon substitution of the pro-$S_P$ oxygen and much larger effects with the pro-$R_P$ substitution. This leaves open the possibility that both the ribozyme and the protein active sites use the same strategy for promoting the chemistry of primer extension, each employing similar two-metal-ion mechanisms for phosphoryl transfer (*21–24*). Furthermore, the similar elemental effects upon pro-$S_P$ substitution (37- and 60-fold) suggest that the two transition states lie at similar positions along the continuum between associative and dissociative character (*15*).

*The Nucleoside: (i) Michaelis–Menten Parameters for Matched and Mismatched Substrates.* Protein-enzyme poly-
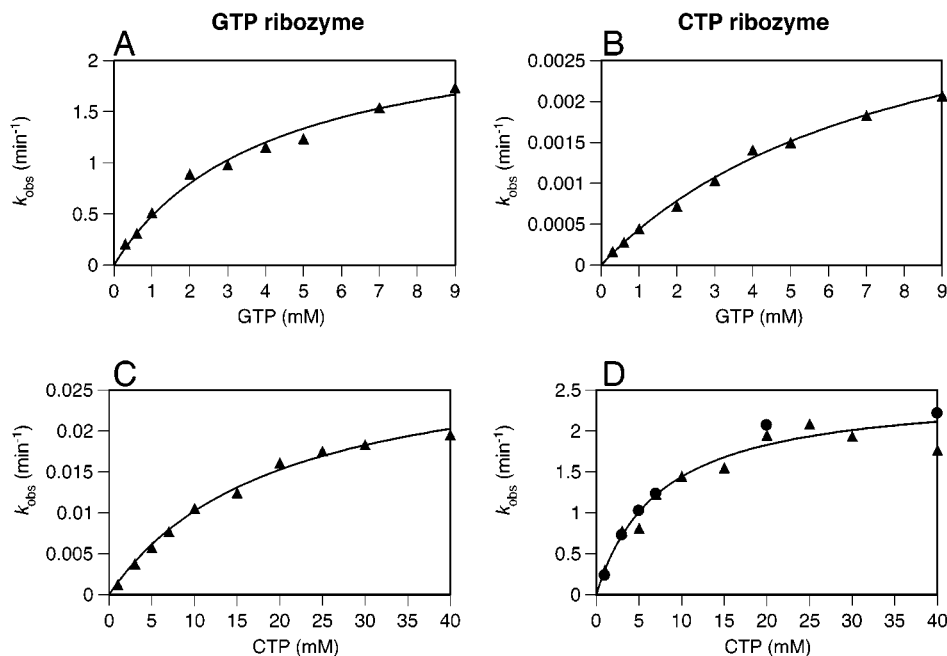
FIGURE 4: Single-nucleotide extension as a function of nucleoside-triphosphate concentration. (A) GTP with the GTP ribozyme. The line corresponds to $k_{cat}^{GTP} = 2.4$ min$^{-1}$, $K_m^{GTP} = 4.1$ mM, and $k_{cat}/K_m^{GTP} = 590$ M$^{-1}$ min$^{-1}$. (B) GTP with the CTP ribozyme. The line corresponds to $k_{cat}^{GTP} = 4 \times 10^{-3}$ min$^{-1}$, $K_m^{GTP} = 8$ mM, and $k_{cat}/K_m^{GTP} = 0.5$ M$^{-1}$ min$^{-1}$. (C) CTP with the GTP ribozyme. The line corresponds to $k_{cat}^{CTP} = 3 \times 10^{-2}$ min$^{-1}$, $K_m^{CTP} = 20$ mM, and $k_{cat}/K_m^{CTP} = 1.5$ M$^{-1}$ min$^{-1}$. (D) CTP with the CTP ribozyme. The line corresponds to $k_{cat}^{CTP} = 2.5$ min$^{-1}$, $K_m^{CTP} = 7.4$ mM, and $k_{cat}/K_m^{CTP} = 340$ M$^{-1}$ min$^{-1}$. Reactions were performed with either 50 (▲) or 100 mM Tris (●).

merases discriminate between matched and mismatched NTPs at the level of NTP binding as well as preferential utilization of NTPs bound with Watson−Crick geometry, although the relative importance of these two modes of discrimination varies (*16, 25, 26*). To investigate the means of discrimination during RNA-catalyzed polymerization, single-turnover kinetic parameters were determined for matched and mismatched substrates. With increasing concentration of GTP, the primer-extension rate of the GTP ribozyme displayed saturable behavior that fit well to a Michaelis−Menten curve (Figure 4A, apparent $k_{cat}^{GTP} = 2.4$ min$^{-1}$ and apparent $K_m^{GTP} = 4.1$ mM). However, the concentration course had to be abbreviated because GTP formed a visible precipitate at concentrations greater than 10 mM. This led to concerns that these parameters might merely reflect aggregation of GTP or noncompetitive inhibition. To control for these nonspecific effects, the concentration course was repeated using the "CTP ribozyme," a ribozyme variant that differed at the template residue such that it coded for extension using CTP (Figure 1E). Although the rates for the CTP ribozyme were nearly a 1000 times slower, they also began to display saturable behavior at high GTP concentrations, fitting well to a Michaelis−Menten curve (Figure 4B, apparent $k_{cat}^{GTP} = 4 \times 10^{-3}$ min$^{-1}$ and apparent $K_m^{GTP} = 8.0$ mM). Thus, some degree of GTP aggregation or noncompetitive inhibition cannot be ruled out. Nevertheless, the 2-fold lower apparent $K_m^{GTP}$ of the GTP ribozyme implies that GTP is at least beginning to saturate the GTP-binding site of this ribozyme.

We performed the analogous set of experiments using CTP, which is more soluble than GTP (Figure 4C,D; for the CTP ribozyme, apparent $k_{cat}^{CTP} = 2.5$ min$^{-1}$ and apparent $K_m^{CTP} = 7.4$ mM; for the GTP ribozyme, apparent $k_{cat}^{CTP} = 3 \times 10^{-2}$ min$^{-1}$ and apparent $K_m^{CTP} = 20$ mM). Again, saturable behavior was seen for both ribozymes, with the

$K_m$ of the cognate ribozyme lower (3-fold) than that of the other ribozyme. This supports the idea that the saturable behavior seen for the cognate substrate of each ribozyme (Figure 4A,D) reflects occupation of the NTP-binding site. The simplest explanation for the Michaelis−Menten behavior of Figure 4B,C is that the NTPs also begin to occupy the NTP-binding sites of the noncognate ribozymes, though with slightly less affinity than for the cognate ribozymes. The dramatic $k_{cat}$ differences between cognate and noncognate ribozymes can be explained by much more favorable NTP orientation with the cognate ribozyme.

Although we do not rule out the possibility that NTP aggregation or noncompetitive inhibition might generate some of the nonlinear behavior seen in Figure 4, we can at least qualitatively surmise that the polymerase uses both binding and orientation to discriminate between matched and mismatched NTPs. The contribution of binding would be least (only 2−3-fold) and that of orientation would be greatest (>100-fold) in the case where $K_m = K_d$ at the NTP-binding site and the mismatched nucleotides are truly saturating the active site. If, for one or both of the ribozymes with mismatched NTPs, the deviation from linearity were instead due to aggregation or noncompetitive inhibition, then the contribution of binding would be correspondingly higher. Even in such a scenario, it is doubtful that binding differences could account for all the discrimination, as this would imply $K_d$ (dissociation constant) values for the mismatches of 12 and 30 M for the GTP and CTP ribozymes, respectively,[2] which approach the concentration of water.

*(ii) GTP Analogues.* Features of the NTP beyond its Watson−Crick face could be important for productive binding by the ribozyme. The contributions of different moieties of the NTP were probed by measuring extension rates with different GTP analogues (Figure 2). Altering base and sugar moieties had a minor effect, provided that Watson−Crick

pairing was maintained. Deleting the 2′-oxygen (2′-dGTP) decreased activity 3-fold, yet the dideoxy analogue (2′,3′-ddGTP) was favored nearly 3-fold over GTP. Another possible contact to the nucleoside was investigated with 7-deaza-dGTP, which replaces a potential hydrogen bond donor, N7, with carbon. Taking the absence of the 2′-hydroxyl into account, the 7-deaza substitution reduced the rate 4-fold.

The slight discrimination between ribo- and 2′-deoxyribonucleotides contrasts with the strict discrimination made by protein enzymes (*27*). Biological polymerases are under intense selective pressure to differentiate between ribo- and 2′-deoxyribonucleotides. Both sets of NTPs are present in the cell, and incorporation of ribonucleotides into DNA or deoxyribonucleotides into RNA is detrimental. Lack of discrimination might have posed a less serious problem in the RNA world before deoxyribonucleotides became common. In contrast to their discrimination at the 2′ position, some bacteriophage RNA and DNA polymerases have only a 3-fold preference for a 3′-hydroxyl (*28, 29*); in their native context, these enzymes are rarely challenged with 3′-deoxynucleoside triphosphates. For the GTP ribozyme, the lack of recognition of the 3′-hydroxyl was expected because the parental ligase ribozyme has a phosphodiester linkage at this position.

*(iii) 3′-Terminal Extensions.* Substrate recognition was further examined using different 3′-terminal extensions of the nucleoside. The key modification in the conversion of the ligase (Figure 1B) to a polymerase (Figure 1C) was the replacement of the first four nucleotides of the ligase (pppGGAA) with exogenous GTP (compare Figure 1A with 1D). This engineering exacted a large cost in reaction rate. For example, in conditions where the apparent $k_{cat}^{GTP}$ was 2.4 min$^{-1}$ for the GTP ribozyme (Figure 4A), the $k_{cat}$ for multiple-turnover ligation is 140 min$^{-1}$ and the rate of self-ligation of the parent ligase is 300 min$^{-1}$ (*13*) (M.E.G., N. Bergman, and D.P.B., manuscript in preparation). To determine if the activity of the GTP ribozyme could be restored to that of the parent ligase, we examined its activity with polynucleotide-triphosphate substrates.

When primer extension using GTP was challenged with a combination of polynucleotide triphosphates, including pppGG, pppGGA, pppGGG, and pppGGAA, the GTP ribozyme preferentially utilized the polynucleotides (Figure 5A).

[2] These $K_d$ values were derived by first fitting for the $K_i$ of the NTP with the mismatched substrate using the standard formula for noncompetitive inhibition (eq 3)

$$k_{obs} = \frac{[S]k_{cat}/(1 + [I]/K_i)}{[S] + K_m} \quad (3)$$

Assuming that all of the deflection from linearity in the plots for the mismatches is due to inhibition, $K_i^{GTP}$ would be 8 mM (Figure 4B), and $K_i^{CTP}$ would be 21 mM (Figure 4C). These $K_i$ values were used to fit for $K_d$ and $k_{chem}$ (rate constant of the chemical step) with the matched substrates, again using eq 3 but substituting $k_{chem}$ for $k_{cat}$ and $K_d$ for $K_m$ (reasonable substitutions given a simple single-turnover reaction with rate-limiting $k_{chem}$). This generated rough estimates of $K_d$ and $k_{chem}$ for matched substrates in the case of maximal noncompetitive substrate inhibition. For the GTP ribozyme $K_d^{GTP}$ would be 30 mM, and $k_{chem}^{GTP}$ would be 16 min$^{-1}$; for the CTP ribozyme $K_d^{CTP}$ would be 50 mM, and $k_{chem}^{CTP}$ would be 13 min$^{-1}$. Equation 4 was used to calculate the $K_d$ of the mismatched substrate in the event that all discrimination was at the level of NTP binding.

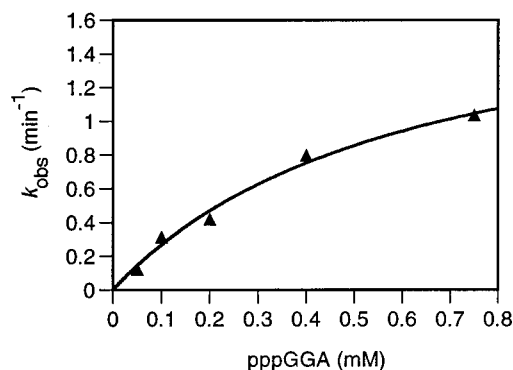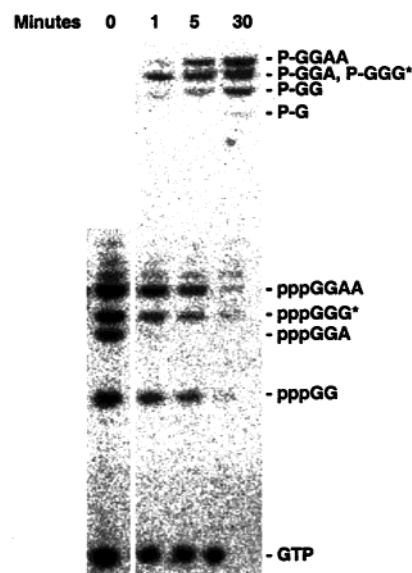$$K_d^{mismatch} = \frac{K_d^{match}(k_{cat}/K_m^{match})}{k_{cat}/K_m^{mismatch}} \quad (4)$$



FIGURE 5: Primer extension using polynucleotide triphosphates. (A) A mixture of GTP and polynucleotide triphosphates with combined concentration <5 μM was incubated with 5 μM unlabeled primer and 5 μM GTP ribozyme. Shown is a phosphorimage of a 22.5% sequencing gel used to separate the radiolabeled substrates (GTP, pppGG, pppGGA, pppGGG, and pppGGAA) and extension products (P-G, P-GG, P-GGA, P-GGG, and P-GGAA, where P denotes the primer). The abortive transcription product indicated by an asterisk (*) was surmised to be pppGGG based on its gel mobility and the known propensity of T7 polymerase to add an untemplated G to the 5′-terminus of transcripts (*40*). (B) Primer extension as a function of pppGGA concentration at pH 6.0. The line corresponds to $k_{cat}^{pppGGA}$ = 1.9 min$^{-1}$, $K_m^{pppGGA}$ = 0.6 mM, and $k_{cat}/K_m^{pppGGA}$ = 3 × 10$^3$ M$^{-1}$ min$^{-1}$. Substrate precipitation prevented extension of the concentration course beyond 0.75 mM pppGGA.

Table 1: Efficiency of Extension Using Polynucleotide Triphosphate Substrates

| substrate | $k_{cat}/K_m$ (M$^{-1}$ min$^{-1}$) | substrate | $k_{cat}/K_m$ (M$^{-1}$ min$^{-1}$) |
|---|---|---|---|
| pppG | 590 | pppGGAA | 25 000 |
| pppGG | 21 000 | pppGGCU | 1 800 |
| pppGGA | 800 000 | pppGAAA | 310 |

Examining appropriate time points and normalizing to the efficiency of GTP utilization allowed the determination of the $k_{cat}/K_m$ for these and other polynucleotide substrates (Table 1). With polynucleotide substrates corresponding to the first three nucleotides of the parental self-ligating ribozyme, the efficiency dramatically increased. The importance of G at the second nucleotide of the substrate can be seen in the increased extension efficiency of pppGG as compared

to that of GTP (40-fold) as well as the increased efficiency of pppGGAA relative to that of pppGAAA (80-fold). The importance of A at the third nucleotide can be seen in the increased extension efficiency of pppGGA as compared to that of pppGG (40-fold) as well as the increased efficiency of pppGGAA versus that of pppGGCU (14-fold).

The most favored polynucleotide substrate tested, pppGGA, was used 1300-fold more efficiently than GTP. We determined the Michaelis−Menten parameters of pppGGA to investigate whether this high catalytic efficiency was due primarily to tighter binding or faster catalysis. For these reactions the pH was lowered from 8.0 to 6.0 in order to slow the reaction sufficiently for time points to be acquired by manual pipetting. The concentration course of pppGGA revealed an apparent $k_{cat}$ of 1.9 min$^{-1}$ and an apparent $K_m$ of 0.6 mM. Extrapolating to pH 8.0 gives a $k_{cat}$ of 190 min$^{-1}$ and a $k_{cat}/K_m$ of $3 \times 10^5$ M$^{-1}$ min$^{-1}$, which is within range of the $k_{cat}/K_m$ value determined earlier by less direct methods ($8 \times 10^5$ M$^{-1}$ min$^{-1}$, Table 1).

A $k_{cat}$ of 190 min$^{-1}$ for pppGGA at pH 8.0 approaches that seen for the rate of self-ligation by the parental ligase, which is 300 min$^{-1}$ (*13*) (M.E.G., N. Bergman, D.P.B., manuscript in preparation). Thus, extending the substrate with a GA dinucleotide, which corresponds to the second and third nucleotides deleted when converting the ligase to a polymerase, nearly restores the ligation rate to that of the parent— even though there is no covalent attachment linking pppGGA to the GTP ribozyme. The first G of pppGGA is thought to form a Watson−Crick pair with C8 of the GTP ribozyme. However, there is no obvious Watson−Crick partner for either nucleotide of the dinucleotide extension. To the degree that the apparent $K_m$ values of GTP and pppGGA (4.1 and 0.6 mM, respectively) reflect substrate binding, the GA extension increases the affinity of the substrate. However, a more important role for the extension appears to be the increased utilization of bound substrate, as reflected in the 80-fold difference in apparent $k_{cat}$ (2.4 and 190 min$^{-1}$, respectively). We suspect that the GA extension interacts with nucleotides in the core of the ribozyme by means of non-Watson−Crick contacts and positions the pppG portion of the substrate for attack by the 3′-hydroxyl group of the primer. In the absence of the GA extension, bound GTP achieves productive orientation only a small fraction of the time. One target for improving the efficiency (and perhaps fidelity) of primer extension would be to generate ribozyme derivatives that restore this putative orientation function.

*The Leaving Group.* For natural RNA and DNA polymerases, much of the NTP affinity is derived from contacts to the triphosphate (*22, 30, 31*). When T7 DNA polymerase binds a matched NTP in the closed complex, it contacts all seven nonbridging oxygen atoms of the triphosphate, either directly through interactions with amino acid residues or indirectly via metal ions (*22*). In the case of the GTP riboyzme, substitutions of the pyrophosphate leaving group significantly reduced the efficiency of primer extension (Figure 2). Phosphate and substituted phosphates (mannose phosphate and nicotinamide phosphate) were poor leaving groups, reducing the rate 2 or 3 orders of magnitude, respectively. GDP, the leaving group for the cap analogue G(5′)ppp(5′)G, had a smaller effect, reducing the rate 20-fold. The large rate reduction by leaving group substitutions may implicate important phosphate contacts, but it may also
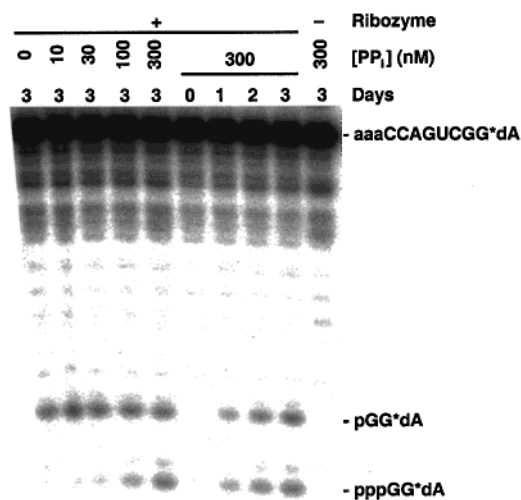


FIGURE 6: Pyrophosphate-exchange and hydrolysis reactions promoted by the GTP ribozyme as a function of time and pyrophosphate concentration. Reactions were performed with 1 $\mu$M ribozyme, <0.7 $\mu$M radiolabeled oligonucleotide (aaaCCAGUCGG*dA; * denotes $^{32}$P), 100 $\mu$M pppGGA, and the indicated concentration of sodium pyrophosphate. The radiolabeled pyrophosphate-exchange product, pppGG*dA, was identified by comigration with an authentic standard generated by T7 transcription using GTP and [$\alpha$-$^{32}$P]-3′-dATP (data not shown). The radiolabeled hydrolysis product, pGG*dA, was identified by comigration with a standard generated by partial digestion of aaaCCAGUCGG*dA with S1 nuclease (data not shown). The slower migration of pGG*dA relative to pppGG*dA can be explained by their respective charges of −4 and −6.

reflect the altered chemical nature of the leaving group. Phosphate is expected to be a poor leaving group, since its last p$K_a$ is 12.0 as compared to 9.1 for pyrophosphate (*32, 33*). Substituted phosphates may have additional p$K_a$ perturbations or steric effects.

The affinity of proteinaceous polymerases for the pyrophosphate leaving group is so great that excess pyrophosphate can drive the reaction backward (*34, 35*). This pyrophosphate-exchange reaction provides an avenue for increased RNA polymerization fidelity; incorporated mismatched nucleotides that block translocation can be excised and replaced (*35*). For the GTP ribozyme, attempts to detect the reverse of single-nucleotide extension have not been successful (N. Bergman and D.P.B., unpublished result). Nevertheless, the reverse of extension using pppGGA can be detected (Figure 6), presumably because of the more productive orientation imparted by the 3′-terminal GA dinucleotide. The GTP ribozyme was incubated with pyrophosphate, an excess of unlabeled pppGGA, and a synthetic oligonucleotide, aaaCCAGUCGG*dA, representing the product of extension using pppGGdA (lowercase, DNA; *, $^{32}$P-radiolabeled phosphate; dA, 3′-deoxyadenosine). The unlabeled pppGGA served as a chase, preventing the newly formed pyrophosphate-exchange product, pppGG*dA, from reassociating with the ribozyme and re-ligating to the primer. The formation of pppGG*dA depended on the pyrophosphate concentration, with an apparent second-order rate constant, $k_{cat}/K_m^{PPi}$, of $1.3 \times 10^{-2}$ M$^{-1}$ min$^{-1}$ (Figure 6). The reaction did not begin to display saturable behavior with soluble concentrations of pyrophosphate ($\leq$0.3 mM). Thus, the GTP ribozyme has a detectable, though nonquantifiable, affinity for pyrophosphate.

Considering the relative efficiencies of the forward and reverse reactions, the forward reaction would be preferred

by a factor of $10^7$ in the presence of moderate concentrations of pppGGA and pyrophosphate (e.g., 0.3 mM each). The preference cannot be determined for single-nucleotide extension because the reverse reaction is too inefficient to be detected, but the preference is expected to be of a similar magnitude. Furthermore, as concentrations are raised, pyrophosphate will begin precipitating before NTPs, further favoring the forward reaction. Thus, unless improved forms of the polymerase have greatly increased affinity for pyrophosphate, their reactions will be essentially irreversible, obviating any need for a pyrophosphatase activity to drive polymerization to completion. On the other hand, a meager reverse reaction prevents utilization of pyrophosphate exchange for increased fidelity (*35*).

A second activity was detected during our investigation of the reverse reaction. A product migrating slower on the sequencing gel than pppGG*dA was not pyrophosphate-dependent (Figure 6). This product appears to be pGG*dA, generated by ribozyme-catalyzed hydrolysis at the reactive phosphate, occurring at a very slow rate ($4.5 \times 10^{-6}$ min$^{-1}$). This reaction, in which an RNA linkage is hydrolyzed to generate a 3′-hydroxyl and a 5′-phosphate, is analogous to the reactions of RNAse P and group I-like ribozymes as well as certain side reactions of group I and group II self-splicing introns (*36−39*). Observation of this reaction with the GTP ribozyme demonstrates that, as expected, not all of its transition-state stabilization is mediated through contacts to the pyrophosphate leaving group; at least part occurs at the reactive phosphate or 3′-oxygen.

*Conclusion.* The GTP ribozyme and related constructs have some intriguing properties, particularly when considering that they were derived from a ribozyme that was selected from random sequences based on its ability to perform a simple, sequence-specific ligation reaction—not an ability to perform template-dependent primer extension using NTPs. The stereospecificity of the thio effects at the reactive phosphate matches those of protein enzymes, leaving open the possibility of a common mechanism. Discrimination among NTPs depends predominantly on Watson−Crick interaction with the template. Furthermore, the reaction is essentially irreversible. It will be interesting to examine the extent to which combinatorial methods can generate ribozymes capable of more accurate, general, and extensive RNA polymerization. A promising approach would be to generate a large pool of RNA sequences based on the core of the GTP ribozyme and to select for variants based on their ability to perform primer-extension reactions rather than RNA ligation.

## ACKNOWLEDGMENT

## REFERENCES

1. Joyce, G. F., and Orgel, L. E. (1999) Prospects for understanding the origin of the RNA world, in *The RNA World* (Gesteland, R. F., Cech, T. R., and Atkins, J. F., Eds.) 2nd ed., pp 49−77, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
2. Pace, N. R., and Marsh, T. L. (1985) *Origins Life 16*, 97−116.
3. Sharp, P. A. (1985) *Cell 42*, 397−400.
4. Cech, T. R. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 4360−4363.
5. Orgel, L. E. (1986) *J. Theor. Biol. 123*, 127−149.
6. Bartel, D. P. (1999) Re-creating an RNA replicase, in *The RNA World* (Gesteland, R. F., Cech, T. R., and Atkins, J. F., Eds.) 2nd ed., pp 143−162, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
7. Bartel, D. P., and Unrau, P. J. (1999) *Trends Biochem. Sci. 24*, M9−M13.
8. Ekland, E. H., Szostak, J. W., and Bartel, D. P. (1995) *Science 269*, 364−370.
9. Ekland, E. H., and Bartel, D. P. (1996) *Nature 382*, 373−376.
10. Bartel, D. P., Doudna, J. A., Usman, N., and Szostak, J. W. (1991) *Mol. Cell. Biol. 11*, 3390−3394 .
11. Erie, D. A., Yager, T. D., and von Hippel, P. H. (1992) *Annu. Rev. Biophys. Biomol. Struct. 21*, 379−415.
12. Ekland, E. H., and Bartel, D. P. (1995) *Nucleic Acids Res. 23*, 3231−3238.
13. Bergman, N. H., Johnston, W. K., and Bartel, D. P. (2000) *Biochemistry 39*, 3115−3123.
14. Costa, M., and Michel, F. (1995) *EMBO J. 14*, 1276−1285.
15. Herschlag, D., Piccirilli, J. A., and Cech, T. R. (1991) *Biochemistry 30*, 4844−4854.
16. Wong, I., Patel, S. S., and Johnson, K. A. (1991) *Biochemistry 30*, 526−537.
17. Beebe, J. A., and Fierke, C. A. (1994) *Biochemistry 33*, 10294−10304.
18. Mei, R., and Herschlag, D. (1996) *Biochemistry 35*, 5796−5809.
19. Dahm, S. C., Derrick, W. B., and Uhlenbeck, O. C. (1993) *Biochemistry 32*, 13040−13045.
20. Eckstein, F. (1985) *Annu. Rev. Biochem. 54*, 367−402.
21. Steitz, T. A. (1998) *Nature 391*, 231−232.
22. Doublie, S., Tabor, S., Long, A. M., Richardson, C. C., and Ellenberger, T. *Nature 391*, 251−258.
23. Kiefer, J. R., Mao, C., Braman, J. C., and Beese, L. S. (1998) *Nature 391*, 304−307.
24. Brautigam, C. A., and Steitz, T. A. (1998) *J. Mol. Biol. 277*, 363−77.
25. Kuchta, R. D., Benkovic, P. A., and Benkovic, S. J. (1988) *Biochemistry 27*, 6716−6725.
26. Kati, W. M., Johnson, K. A., Jerva, L. F., and Anderson, K. S. (1992) *J. Biol. Chem. 267*, 25988−25997.
27. Joyce, C. M. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 1619−1622.
28. Axelrod, V. D., and Kramer, F. R. (1992) *Biochemistry 24*, 5716−5723.
29. Tabor, S., and Richardson, C. C. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 6339−6343.
30. Beese, L. S., Friedman, J. M., and Steitz, T. A. (1993) *Biochemistry 32*, 14095−14101.
31. Pelletier, H., Sawaya, M. R., Kumar, A., Wilson, S. H., and Kraut, J. (1994) *Science 264*, 1891−1903.
32. Rohatgi, R., Bartel, D. P., and Szostak, J. W. (1996) *J. Am. Chem. Soc. 118*, 3332−3339.
33. Dawson, R. M. C., Elliott, D. C., Elliott, W. H., and Jones, K. M. (1969) in *Data for Biochemical Research*, pp 400−412, Clarendon Press, Oxford.
34. Deutscher, M. P., and Kornberg, A. (1969) *J. Biol. Chem. 244*, 3019−3028.
35. Kahn, J. D., and Hearst, J. E. (1989) *J. Mol. Biol. 205*, 291−314.
36. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., and Altman, S. (1983) *Cell 35*, 849−857.
37. Decatur, W. A., Einvik, C., Johansen, S., and Vogt, V. M. (1995) *EMBO J. 14*, 4558−4568.
38. Inoue, T., Sullivan, F. X., and Cech, T. R. (1986) *J. Mol. Biol. 189*, 143−165.
39. Jarrell, K. A., Peebles, C. L., Dietrich, R. C., Romiti, S. L., and Perlman, P. S. (1988) *J. Biol. Chem. 263*, 3432−3439.
40. Moroney, S. E., and Piccirilli, J. A. (1991) *Biochemistry 30*, 10343−10349.