Processivity of Ribozyme-Catalyzed RNA Polymerization†

Michael S. Lawrence and David P. Bartel*

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

Received February 10, 2003; Revised Manuscript Received May 19, 2003

ABSTRACT: The "RNA world" hypothesis proposes that early in the evolution of life, before the appearance of DNA or protein, RNA was responsible both for encoding genetic information and for catalyzing biochemical reactions. Ribo-organisms living in the RNA world would have replicated their RNA genomes by using an RNA polymerase ribozyme. Efforts to provide experimental support for the RNA world hypothesis have focused on producing such a polymerase, and *in vitro* evolution methods have led to the isolation of a polymerase ribozyme that catalyzes primer extension which is accurate and general, but slow. To understand the reaction of this ribozyme, we developed a method of measuring polymerase processivity that is particularly useful in the case of an inefficient polymerase. This method allowed us to demonstrate that the polymerase ribozyme, despite its inefficiency, is partially processive. It is currently limited by a low affinity for the primer−template duplex, but once it successfully binds the primer−template duplex in the productive alignment, it catalyzes an extension reaction that is so rapid that it can occur multiple times during the short span of a single binding event. This finding contributes to the understanding of one of the more sophisticated activities yet to be generated *de novo* in the laboratory and sheds light on the parameters to be targeted for further optimization.

One part of the mystery of life's origins is the emergence of catalyzed information propagation. Little is known about the evolutionary precursors of modern DNA and RNA polymerases, those data-copying enzymes which, together with the ribosome, make up the core of contemporary molecular biology. Information replication, in the view of many theorists, was achieved by even the earliest, most primitive forms of life (*1*, *2*). According to the RNA world hypothesis, information replication was occurring in RNA-based "ribo-organisms" before the emergence of proteins or DNA (*3*−*6*). This idea flows from the unique dual nature of RNA as both an inherent information carrier and a versatile biological catalyst. One ribozyme indispensable to the RNA world would have been a polymerase, responsible for copying all the ribozymes of a ribo-organism by catalyzing the templated polymerization of mono- or oligonucleotides. The RNA world hypothesis thus presupposes that some RNA sequences can fold into RNA-directed RNA polymerases.

Experimental support for this presumption is starting to accumulate. *In vitro* evolution methods have yielded an RNA polymerase ribozyme that catalyzes an accurate and general primer extension reaction (*7*). This polymerase (Figure 1A) uses ribonucleoside triphosphates and the coding information of an RNA template to successively extend an RNA primer. It accepts primers and templates of any sequence or length, provided the 3′ terminus of the primer pairs to the template, and it catalyzes polymerization with an average fidelity of 0.97. In its present form, however, the polymerase is slow. It requires at least 6 h to extend a primer−template duplex

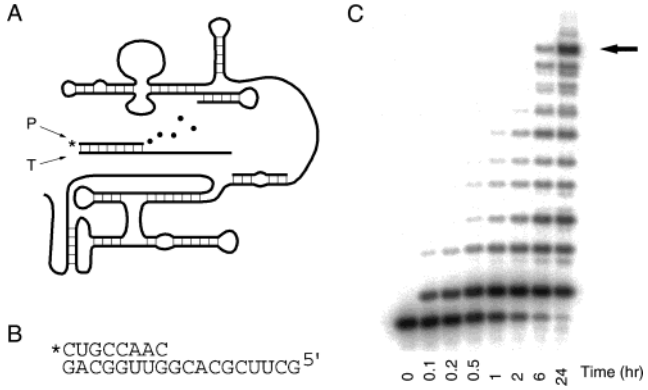


FIGURE 1: Ribozyme-catalyzed RNA polymerization. (A) Stylized representation of the polymerase ribozyme secondary structure (*7*), with P indicating primer, T indicating template, and dots representing NTPs. (B) Example of an RNA PT (PT1). The primer is radiolabeled (*) at its 5′ end. (C) Sequencing gel showing the results of a polymerization reaction using the PT from panel B. The arrow denotes the fully extended product.

(PT)[1] by one helical turn (Figure 1B,C). Furthermore, because the buffer conditions that maximize its polymerization rate also promote its own hydrolysis, the polymerase is substantially inactivated by the long incubation. With its current limitations, this polymerase ribozyme is incapable of synthesizing the long stretches of RNA that would be needed in the RNA world. To generate a better polymerase, it would be useful to understand which steps of the polymerization reaction are most in need of improvement.

† This work was supported by NIH Grant GM61835.

* To whom correspondence should be addressed. Phone: (617) 258-5287. Fax: (617) 258-6768. E-mail: dbartel@wi.mit.edu.

[1] Abbreviations: NTP, nucleoside triphosphate; EDTA, ethylenediaminetetraacetic acid; Tris, tris(hydroxymethyl)aminomethane; PT, primer−template duplex.

10.1021/bi034228l CCC: $25.00
Published on Web 07/02/2003

The polymerase ribozyme can be compared with its proteinaceous counterparts, which are generally many orders of magnitude faster, adding up to 750 bases per second (*8*, *9*). Some protein polymerases, notably those responsible for replicating chromosomal DNA, catalyze highly processive catalysis, adding many thousands of nucleotides without releasing the nascent chain (*10*, *11*). This can be contrasted with completely nonprocessive, or distributive, polymerization, in which a PT is iteratively bound, extended by a single nucleotide, and then released. Modern replicative DNA polymerases owe much of their efficiency to their high processivity, which allows them to minimize time lost in repeatedly releasing and rebinding the template.

Given the importance of processivity in protein-catalyzed polymerization, we set out to examine the processivity of the RNA-catalyzed polymerization reaction. The pattern of products that the polymerase ribozyme has been observed to generate already suggests some degree of processivity. In the case of completely distributive polymerization, by the time the fully extended primer had begun to accumulate, short primers would be nearly used up, and the distribution of products would peak at some intermediate length. However, cursory inspection of Figure 1C shows the inverse product distribution; as the reaction proceeds, it becomes dominated by fully extended product and very short products. It is not straightforward to reconcile the underrepresentation of intermediate-length products with a completely distributive mechanism. On the other hand, the inefficiency of the ribozyme-catalyzed polymerization seems to be at odds with the notion of processivity. To shed some light on this apparent paradox, we investigated the polymerase reaction more closely.

Classical methods of determining polymerase processivity rely on exclusion of multiple binding events. Experiments are designed such that all observed primer extension comes from a single round of processive polymerization. For instance, in a study of yeast polymerase η, a substrate trap was used for this purpose (*12*). The polymerase was preincubated with radiolabeled PT, and then the polymerization reaction was initiated by simultaneous addition of dNTPs and a saturating amount of herring sperm DNA as a trap. As the reaction proceeded, any PT that fell off the polymerase was replaced by herring sperm DNA. Therefore, all observed extension of the labeled primer was the result of a single binding event. Earlier studies of *Escherichia coli* DNA polymerase I ensured a single binding event by adding PT in large excess over polymerase so that the polymerases were statistically extremely unlikely ever to rebind the same PT molecule (*10*, *13*). Subsequent studies of *E. coli* DNA polymerase I and mouse DNA polymerase α used a large excess of PT over labeled dNTPs (*14*, *15*). Under those conditions, the predominant way for two labeled dNTPs to end up in the same product molecule was through addition during a single round of processive polymerization.

The experimental approaches used in the studies described above could not be adapted to our situation. This was because of the low affinity of our polymerase for its PT. Using a substrate trap (such as herring sperm DNA) was impossible, because saturating levels of RNA could not be achieved, and moreover, high concentrations of RNA appeared to inhibit the polymerase. Reducing polymerase and NTP concentrations to restrict observed polymerization to single binding events would have had the unacceptable consequence of abolishing the signal available for measurement. Faced with the limitations imposed by our polymerase ribozyme, chiefly its low affinity for the PT, we devised a strategy for measuring its processivity that did not rely on being able to exclude multiple binding events. Our approach involved constructing computational kinetic models of distributive and processive polymerization and determining which of the two models better mirrored the experimentally observed polymerization.

MATERIALS AND METHODS

Polymerization Reactions. The polymerase was the round-18 ribozyme from Johnston et al., prepared as described therein (*7*). Polymerization reaction mixtures contained 2 μM ribozyme (with 2.5 μM GGCACCA oligoribonucleotide), 1 μM RNA template, and 0.5 μM 5′-radiolabeled RNA primer. These RNAs were mixed together in water, heated (80 °C for 2 min), and then incubated at 22 °C for at least 5 min before the reaction was started by the simultaneous addition of nucleoside triphosphates (NTPs, 4 mM each), 200 mM $MgCl_2$, and 100 mM Tris (pH 8.5) (final concentrations). The typical reaction volume was 30 μL. Reaction mixtures were incubated at 22 °C. At selected time points, reaction aliquots were withdrawn and the reactions quenched by addition of 4 volumes of 50 mM EDTA and 8 M urea. Quenched aliquots were heated (80 °C for 2 min) in the presence of competitor RNA designed to hybridize to the template RNA and then analyzed on sequencing gels. Variations of this protocol were as described in the text.

Prebinding Experiment. Two parallel polymerization reactions were performed, each with 5 μM ribozyme (with 6.25 μM GCCACCA oligo), 0.5 μM radiolabeled primer and 2.5 μM template (PT6, 18-nucleotide template), 200 mM $MgCl_2$, 100 mM Tris (pH 8.5), and 4 mM GTP. In the "prebound" reaction, all RNAs were heated together in water (80 °C for 2 min) and cooled for 5 min to 22 °C, then salt and buffer were added, and the mixture was incubated for an additional 5 min at 22 °C. Reaction was started by 40-fold dilution into GTP, salt, and buffer. In the "non-prebound" reaction, ribozyme and PT were kept separate during the heating, cooling, and salt/buffer incubation. Preincubated PT was diluted into GTP, salt, and buffer, and the reaction was started by addition of the preincubated ribozyme. Upon this step, final concentrations of all reaction components became exactly as in the prebound reaction, and time courses proceeded as described above.

Kinetic Modeling. Polymerizations were modeled computationally by numerical integration of a set of equations that formalize the chemical reactions occurring in the polymerization. All numerical integration was performed using a Microsoft Excel spreadsheet in which each column tracked one variable, and each row recomputed all the variables a small time interval (Δt) after the preceding row. For example, in a time course of single-nucleotide addition, such as the irreversible extension of PT3 to yield PT4, only a single reaction needs be considered. This reaction is first-order with respect to the concentration of PT3 and obeys the observed rate constant $k_{obs(3)}$. Two complementary equations are required to formalize this reaction: one to represent the disappearance of PT3 and another to represent

Table 1: PTs, Observed Extension Rate Constants, and Processivity Coefficients

		18-nt template 0.1 mM ea. NTP		18-nt template 4 mM ea. NTP		21-nt template 4 mM ea. NTP	
Primer-template[a]		k_{obs} (hr^{-1})	P	k_{obs} (hr^{-1})	P	k_{obs} (hr^{-1})	P
PT1	5'CUGCCAAC 3'GACGGUUGGCACGCUUCG(CAG)	0.16		1.4		0.08	
PT2	5'CUGCCAACC 3'GACGGUUGGCACGCUUCG(CAG)	0.0036	0.01	0.09	0.03	0.09	0.00
PT3	5'CUGCCAACCG 3'GACGGUUGGCACGCUUCG(CAG)	0.08	0.07	0.56[b]	0.15	1.8	0.30
PT4	5'CUGCCAACCGU 3'GACGGUUGGCACGCUUCG(CAG)	0.014	0.16	0.16[b]	0.45[b]	1.5	0.47
PT5	5'CUGCCAACCGUG 3'GACGGUUGGCACGCUUCG(CAG)	0.15	0.00	1.3	0.07	11	0.9
PT6	5'CUGCCAACCGUGC 3'GACGGUUGGCACGCUUCG(CAG)	0.22	0.04	1.4	0.25	3.8	0.9
PT7	5'CUGCCAACCGUGCG 3'GACGGUUGGCACGCUUCG(CAG)	0.048	0.01	0.41	0.02	6.0	0.33

[a] Template bases shown in parentheses are absent in the 18-nucleotide template. [b] These measurements, carried out once already during the procedure illustrated in Figure 3, were repeated as part of the expanded set of experiments reported in this table; this explains their slight deviations from the values reported in Figure 3. These differences of 0−20% provide an indication of the experimental variability associated with these measurements.

the concomitant appearance of PT4. The two equations are (1) $\Delta[PT3] = -k_{obs(3)}[PT3]\Delta t$ and (2) $\Delta[PT4] = k_{obs(3)}[PT3]\Delta t$. The spreadsheet for this time course had three columns, one each for t (time), [PT3], and [PT4]. All columns were initialized to zero in the first row of the spreadsheet, except for [PT3], which was initialized to 0.5 μM, the starting concentration of PT3 in the reaction. Each subsequent row recomputed the three columns as follows. The new value of t was the old value of t plus the constant Δt. The new value of [PT3] was the old value of [PT3] (found in the previous row) minus the product of the constant $k_{obs(3)}$ times the old value of [PT3] times Δt. The new value of [PT4] was the old value of [PT4] plus that same product. The best fit for $k_{obs(3)}$ was determined by adjusting $k_{obs(3)}$ until the root-mean-square discrepancy between the modeled time course and the experimental data was minimized. As long as Δt was sufficiently small, the results of the simulation did not depend on its value. All simulations reported herein used a Δt of 2 s. Simulations using a Δt of 10 s produced indistinguishable results.

Distributive Model. The completely distributive model shown in Figure 3C was constructed by generalizing the system described above. For each extension step in the polymerization, one chemical reaction was added to the model, reflecting the single-nucleotide extension of unbound PTn to yield bound PTn+1, obeying the rate constant $k_{obs(n)}$. Subsequent release of PTn+1 was modeled as occurring instantaneously. The parameters of the distributive model are the set of $k_{obs(n)}$ values for each PTn. These parameters were fit independently of one another, one from each time course of single-nucleotide extension.

Processive Model. For each extension step in the processive model shown in Figure 3D, three chemical reactions were included in the model. The first was for single-nucleotide extension of unbound PTn to yield bound PTn+1, as described above. The second was for processive extension of bound PTn+1 to yield bound PTn+2, obeying the rate constant $k_{cat(n+1)}$. The third was for release of bound PTn+1, obeying the rate constant $k_{off(n+1)}$. The processivity coefficient P was defined as the ratio $k_{cat}/(k_{cat} + k_{off})$. Two versions of the processive model were implemented, one with bound PTn partitioning instantaneously and one with the partitioning rate set equal to 6 min^{-1}, the lower bound on $k_{cat} + k_{off}$ calculated from the burst duration in the prebinding experiment. These two implementations gave indistinguishable results. The parameters of the processive model are the set of $k_{obs(n)}$ and $P_{(n)}$ values for each PTn. Values of $k_{obs(n)}$ were inherited from the distributive model. Values of $P_{(n)}$ were then fit independently of one another, one from each time course of two-nucleotide extension. Extension beyond two nucleotides was not examined during the parameter fitting. It was reserved for testing the model's predictive power.

RESULTS

Affinity for PT. Processivity depends on two competing processes: extension and release. A polymerase is processive only to the extent that it can hold onto its PT substrate longer than it takes to extend it. Therefore, to determine whether the polymerase is processive or distributive, it was informative to measure its affinity for the PT. When a PT (PT2, Table 1) was titrated into a polymerization reaction mixture, measured reaction rates were linearly proportional to the PT concentration over the entire testable range, revealing a Michaelis−Menten parameter (K_m) of at least 400 μM, and possibly much higher. However, titration of the ribozyme into a similar reaction mixture showed that polymerization rates decline at ribozyme concentrations above 15 μM, suggesting that high concentrations of RNA could inhibit the polymerase.

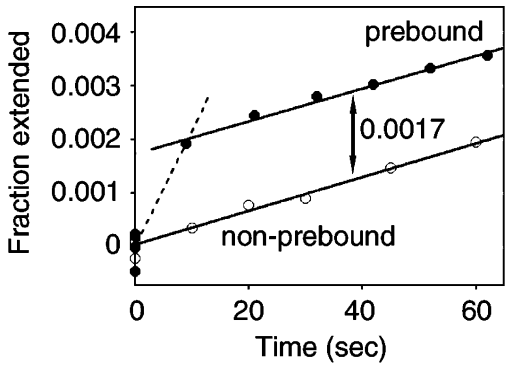
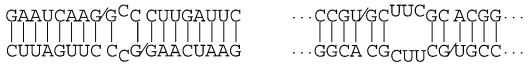


FIGURE 2: Prebinding and dilution experiment, showing single-nucleotide extension of PT6 (Table 1). In the prebound reaction (●), 5 μM ribozyme and 0.5 μM PT were incubated together before the reaction was initiated. In the non-prebound reaction (○), the ribozyme and PT were kept separate until the moment the reaction was initiated. The amplitude of the burst phase (0.0017) is equal to the distance between lines fit to the linear portion of each time course. The dotted line represents the lower bound on the trajectory of the burst phase.

Other PTs had saturable titration curves, which on the surface suggested tighter binding to the polymerase. For instance, in a titration of PT4, reaction rates leveled off quickly, with a half-maximal rate observed at 6 μM. A similar result was seen with the unrelated primer−template PTX (5′-GAAUCAAG/3′-CUUAGUUCCCG), which exhibited half-saturation at 7 μM. However, doubt was cast on this interpretation of these low apparent K_m values by experiments attempting to detect competitive or noncompetitive inhibition. If PT4 truly bound to the polymerase active site with micromolar affinity, then at concentrations above 7 μM it would interfere with polymerization of PTX. However, only a very slight inhibitory effect was observed when excess PT4 was titrated as a potential inhibitor into the polymerization reaction mixture of PTX; at an inhibitor concentration of 75 μM, the polymerization rate dropped only 5%. In the reciprocal experiment, no decline in the reaction rate was observed when PTX was titrated into the reaction mixture of PT4. On the basis of these experiments, the dissociation constants (K_d) of the polymerase and these PTs must be ~1 mM or higher. The low apparent K_m values measured in the titration experiments could be explained by substrate self-sequestration. The homodimers of PTX and PT4 could be stabilized by base pairing, as well as by stacking interactions at the monomer junctions, whereas there is no obvious way for PT4 and PTX to form stable heterodimers.



A prebinding and dilution experiment was performed to ascertain directly how much of the PT binds to the polymerase. The polymerase and PT6 were preincubated together in reaction buffer, allowing them time to bind to one another. Then, at the moment of NTP addition, the reactants were diluted 40-fold to minimize later binding of the polymerase and the PT. A small "burst" of rapid extension was observed, with 0.17% of the PT being extended within the first 10 s of the reaction. After that burst, extension proceeded at a much lower rate (Figure 2). In a control reaction performed without the preincubation of the polymerase and the PT, no burst phase was observed; extension proceeded only at the slow association-limited rate, identical to that measured after the burst in the first experiment.

The burst phase observed in the prebinding experiment can be quantified in two ways: its size and its duration. The size of the burst phase correlates with how much PT was prebound to the polymerase, and the duration correlates with how long the prebound complex survived in the polymerization reaction. The prebound complex has two possible fates; either it successfully reacts to yield the extension product (with a rate constant k_{cat}), or else it simply dissociates (with a rate constant k_{off}). Because the burst was complete within 10 s, a lower bound of 6 min^{-1} can be placed on the sum of $k_{cat} + k_{off}$. Interpreting the size of the burst requires knowing what fraction of the prebound complex reacted to yield product instead of simply dissociating. At one extreme, if one assumes that much more than 0.17% of the PT was productively prebound to the polymerase (implying a low K_d), but only a very small fraction of it reacted, then this would imply very fast dissociation from the polymerase, with a k_{off} of at least 6 min^{-1}. At the other extreme, if one assumes that only 0.17% of the PT was prebound to the polymerase in the productive alignment and all of it was extended during the burst, then this would imply that the K_d of the polymerase−PT interaction is approximately 3 mM. Given the earlier measurement suggesting a K_d in the millimolar range or higher, this second extreme seemed to be closer to reality.

Taken together, the above experiments suggested that the affinity of the polymerase for its RNA substrate is weak, probably in the millimolar range, shedding light on the slow polymerization observed in Figure 1C. In that experiment, the polymerase and its PT were both present at low micromolar concentrations, 1000-fold below saturation. Moreover, significant increases in rate could not be achieved simply by increasing the ribozyme concentration, because polymerization was inhibited by high concentrations of ribozyme. These limitations confined our experiments to subsaturating conditions for all analyses of the ribozyme and its reactions.

Measuring Processivity. We devised a method of measuring processivity that did not rely on being able to saturate the polymerase with substrate trap or otherwise exclude multiple binding events. First, a set of RNA primers was chemically synthesized, all of which could anneal to a single template. Each primer in the series was one nucleotide longer than the primer preceding it. This set of PTs (Table 1) can be viewed as a series of intermediates in an RNA polymerization reaction. Each PT in the series has two possible histories: either it can be generated from its one-nucleotide-shorter precursor by the ribozyme, or else it can be added to the reaction mixture before the start of the reaction. Each PT was used as the starting material for a ribozyme-catalyzed polymerization reaction. For each reaction, the observed rate constant (k_{obs}) was measured for the binding and extension of that PT. These observed rate constants were then combined into a model of distributive polymerization, in which every step of polymerization started from unbound PT. This model was tested by comparison to the multiple-nucleotide polymerization observed experimentally. The observed polymerization was too fast to be explained by the distributive model, and so a processive model was developed to replace it.

Distributive Model. We began with an extension reaction using PT3 of the Table 1 series as the starting material (Figure 3A, gel). For each time point, extension was quantitated by summing the ladder of products accumulating above the starting band and dividing by the sum of all bands, including the starting band, thereby yielding the fraction of primer extended by at least a single nucleotide. The extension was modeled as a single irreversible chemical reaction obeying the observed rate constant of extension $k_{obs(3)}$, determined by a best fit to be 0.46 h^{-1} (Figure 3A, graph). The reaction was first-order with respect to the PT concentration, because the experiment was performed under subsaturating conditions, well within the concentration regime where the polymerization rate exhibited this simple linear dependence. The polymerization rate also depended on the concentration of NTPs, magnesium, hydroxide, and the polymerase itself; however, these did not vary appreciably over the time course of the reaction, and so they were treated as part of the constant standard reaction conditions. Therefore, throughout this work, observed first-order rate constants (k_{obs}) of polymerization are reported, with respect to these standard conditions.

Next, the analogous polymerization assay was performed, starting with PT4 instead of PT3 (Figured 3B, gel). This PT was identical to the single-extension product of the first experiment. The observed rate constant of extension for PT4, $k_{obs(4)}$, was fitted as 0.14 h^{-1} (Figure 3B, graph).

The two observed rate constants thus determined were then combined into a model of distributive polymerization. The two-step process, starting with PT3 and yielding PT5, was treated as a system of two consecutive irreversible reactions (Figure 3C, scheme). The time course predicted by this kinetic scheme was computed and plotted (Figure 3C, graph). This is the time course of extension of PT3 by at least two nucleotides, as it would be observed if the polymerase catalyzed completely distributive polymerization. In such a scenario, the starting material PT3 diffuses through the reaction medium until it encounters the polymerase, at which point it is bound, extended by one nucleotide to become PT4, and promptly released back into solution. It then diffuses until binding again to a polymerase molecule, when it undergoes another nucleotide addition and is released back into solution as PT5.

To evaluate how well this modeled time course predicted the actual behavior of the polymerase in extending PT3 by two nucleotides, the results of the first experiment starting with PT3 (Figure 3A, gel) were requantitated in a different way, this time summing all product bands at least two nucleotides larger than the starting material, thereby yielding the fraction of primer extended at least twice. These data were compared to the distributively modeled time course (Figure 3C, graph). Clearly, the data and distributive model were in disagreement. The accumulation of PT5 occurred too fast to be explained by a fully distributive mechanism. Furthermore, the distributive model assumed that product release occurred immediately after nucleotide addition. Any delay due to product release would shift the modeled distributive time course even lower. The curve shown in Figure 3C therefore reflects the best-case scenario for the distributive model. Moreover, the distributive model could not be redeemed by simply increasing the individual rate constants. Doing so yielded incorrectly shaped curves that

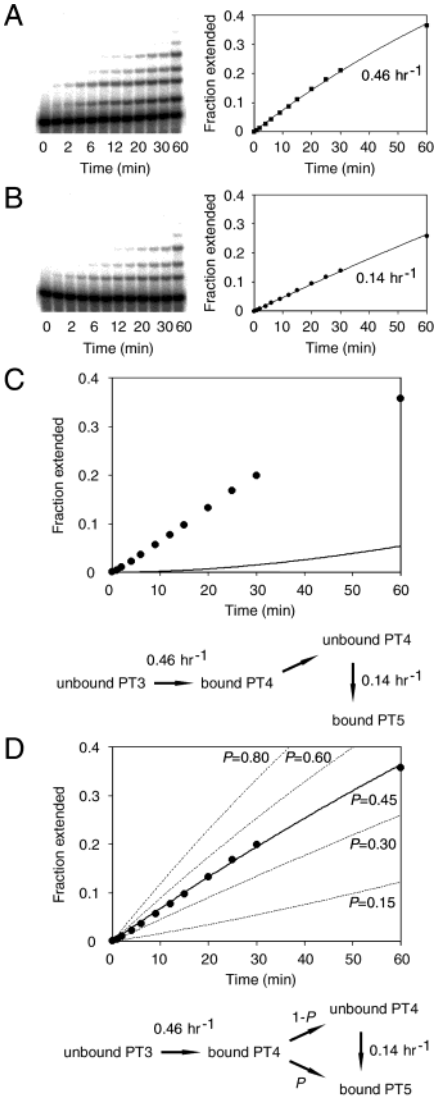


FIGURE 3: Development of the processive model. (A) Single-nucleotide extension of PT3. Single-nucleotide extension was quantitated by summing all bands above the lowest band (the starting material), and dividing by the sum of all bands in the lane, to yield the fraction of starting material extended by at least one nucleotide. This fraction is plotted along with a fitted time course with an observed rate constant $k_{obs(3)}$ of 0.46 h^{-1}. (B) Single-nucleotide extension of PT4, quantitated as in panel A and plotted along with a fitted time course with an observed rate constant $k_{obs(4)}$ of 0.14 h^{-1}. (C) Two-nucleotide extension of PT3 and a minimal scheme for the distributive model. Plotted is the fraction of starting material that has been extended by at least two nucleotides, quantitated from the gel in panel A by summing all bands except the lowest two bands, and dividing by the sum of all bands in the lane. The distributive scheme reflects conversion of unbound PT3 to bound PT4 with $k_{obs(3)}$, followed by obligate release of PT4, and then conversion of unbound PT4 to bound PT5 with $k_{obs(4)}$. Superimposed on the data is the modeled time course of distributive polymerization. The distributive model is a poor fit to the data. (D) Two-nucleotide extension of PT3 and a minimal scheme for the processive model. Data points are as in panel C. The processive scheme reflects conversion of unbound PT3 to bound PT4 with $k_{obs(3)}$, followed by one of two possible paths: either further extension of bound PT4 to bound PT5, occurring with probability P, or release of bound PT4, occurring with probability $1 - P$, followed by conversion of unbound PT4 to bound PT5 with $k_{obs(4)}$. Modeled processive time courses (solid and dotted lines) superimposed upon the data show the effect of varying P in the processive model. The best fit to the data is achieved with a P of 0.45 (solid line).

undershot early data points and overshot late ones (not shown).

Processive Model. Having demonstrated that the polymerase catalyzed addition of two nucleotides faster than a fully distributive mechanism allowed, we still needed to explain just how it accomplished this feat. One reasonable hypothesis was that the polymerase could hold onto its PT long enough to add a second nucleotide some fraction of the time. We defined the processivity coefficient P as the probability of adding a second nucleotide before PT release. This parameter ranges from zero (completely distributive polymerization, already ruled out) to one (completely processive polymerization, in which the polymerase holds onto its substrate and extends it all the way to the end of the template before letting go).

In the processive model of polymerization (Figure 3D, scheme), the starting material PT3 diffuses to and binds to the polymerase, where it is extended by a single nucleotide, yielding PT4. This step was modeled with a $k_{obs(3)}$ of 0.46 h^{-1}, as before. The distributive model required that PT4 then be released from the polymerase and rebound before it could be extended again. At this juncture, however, the processive model diverged, allowing PT4 to be extended a second time (yielding PT5) while still bound to the polymerase. This occurs with a probability of $P_{(4)}$. The other possible outcome, with a probability of $1 - P_{(4)}$, is for PT4 to fall off the polymerase and continue along the distributive pathway by binding again to a polymerase molecule and being extended to PT5. As before, the binding and extension of PT4 were modeled with a $k_{obs(4)}$ of 0.14 h^{-1}. The Figure 3D graph shows the effect of varying $P_{(4)}$ in the model so described. Each curve represents the partially processive time course of extending PT3 to PT5, as calculated for a given value of $P_{(4)}$. A close match to the observed time course was obtained with a $P_{(4)}$ of 0.45. In other words, the polymerase binds PT3 and extends it a first time, and then with nearly equal probability either releases it or extends it a second time.

Analysis was extended to other positions along the template by iteration of the process just described. For example, a polymerization assay was performed using PT5 (not shown). From this additional data, $k_{obs(5)}$ and $P_{(5)}$ were determined. Each subsequent polymerization assay yielded an additional pair of parameters. For the final data set (Table 1), the measurements depicted in Figure 3 were repeated as part of an expanded set of experiments encompassing seven consecutive positions along the template. A strong dependence on sequence context was observed, with both k_{obs} and P varying approximately 20-fold, but in nearly all cases, some degree of processivity was detectable.

Quantitation of the processivity of the polymerase shed additional light on the prebinding experiment (Figure 2), allowing a clearer interpretation of its results. The burst phase observed in that experiment had established a lower bound of 6 min^{-1} for the sum of $k_{cat} + k_{off}$, but because it was unknown what fraction of bound PT reacted during the burst instead of falling off, it was impossible to establish lower bounds on the individual rate constants k_{cat} and k_{off}. Measuring a processivity at that position of 0.25 (PT6 in Table 1) revealed that one out of four bound PTs had been extended, and the other three had dissociated from the polymerase before they could be extended. This in turn implied that k_{off} had a magnitude 3 times greater than that of k_{cat} and allowed us to establish lower bounds of 4.5 min^{-1} for k_{off} and 1.5 min^{-1} for k_{cat} (Figure 4). Furthermore, measuring a processivity of 0.25 implied that the extension observed during the burst phase (0.17%) represented only one-fourth of the prebound PT; the total amount of prebound PT was 0.7%. This value (along with the ribozyme concentration, 5 μM) allowed us to refine our estimate of the K_d of this interaction to 700 μM. As polymerization is more efficient in the context of PT6 than in some other contexts, some PTs are likely to have higher K_d values than PT6.

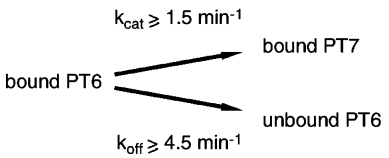


FIGURE 4: Lower bounds on k_{cat} and k_{off}. Polymerase-bound PT6 is either extended or released, at rates no lower than the indicated limits.

Validating the Processive Model. For the processive model to capture the behavior of two-nucleotide extension, it required two parameters, k_{obs} and P. At this point, the crucial question arose: would these parameters, determined locally for pairs of successive PTs, be sufficient for modeling long-range polymerization? This question was addressed by challenging the processive model to predict polymerization of three, four, five, and more nucleotides. As hoped, the results of the simulation agreed with all observed data. No further refinement or extra parameters were required to successfully predict the time course of many-nucleotide extension. Figure 5 shows the results of this test. In each of panels A−E, a different PT was used as starting material, and the observed polymerization data are plotted in both the top and bottom graphs. In the top graph, the predictions of the distributive model are superimposed on the data. In the bottom graph, the predictions of the processive model are superimposed. In all cases, both models correctly predicted the first extension (uppermost curve), but only the processive model succeeded in predicting subsequent extensions.

Effect of Reduced NTP Concentrations. At the core of the processive model is the partitioning of the bound PT along two pathways: extension and release. If some change in reaction conditions were to affect one of these two pathways disproportionately, then it should also cause a change in the observed processivity. We tested this prediction by reducing the concentration of NTPs in the reaction. This change was expected to reduce the observed extension rate without dramatically affecting the rate of PT release. According to the processive model, it should also cause a drop in processivity. All parameters were remeasured using $^1/_{40}$ of the original NTP concentration, and as expected, all observed extension rate constants and processivity coefficients fell (Table 1).

Effect of a Lengthened Template. As an additional test of the model's robustness, polymerization was re-examined using a template with three additional nucleotides at its 5′ end. This slight change was found to influence the behavior of the polymerase dramatically (Table 1). In almost all cases, the observed extension rate constants and the processivity coefficients both increased, in some cases by more than an order of magnitude. Nevertheless, when these markedly different parameters were combined into a new processive

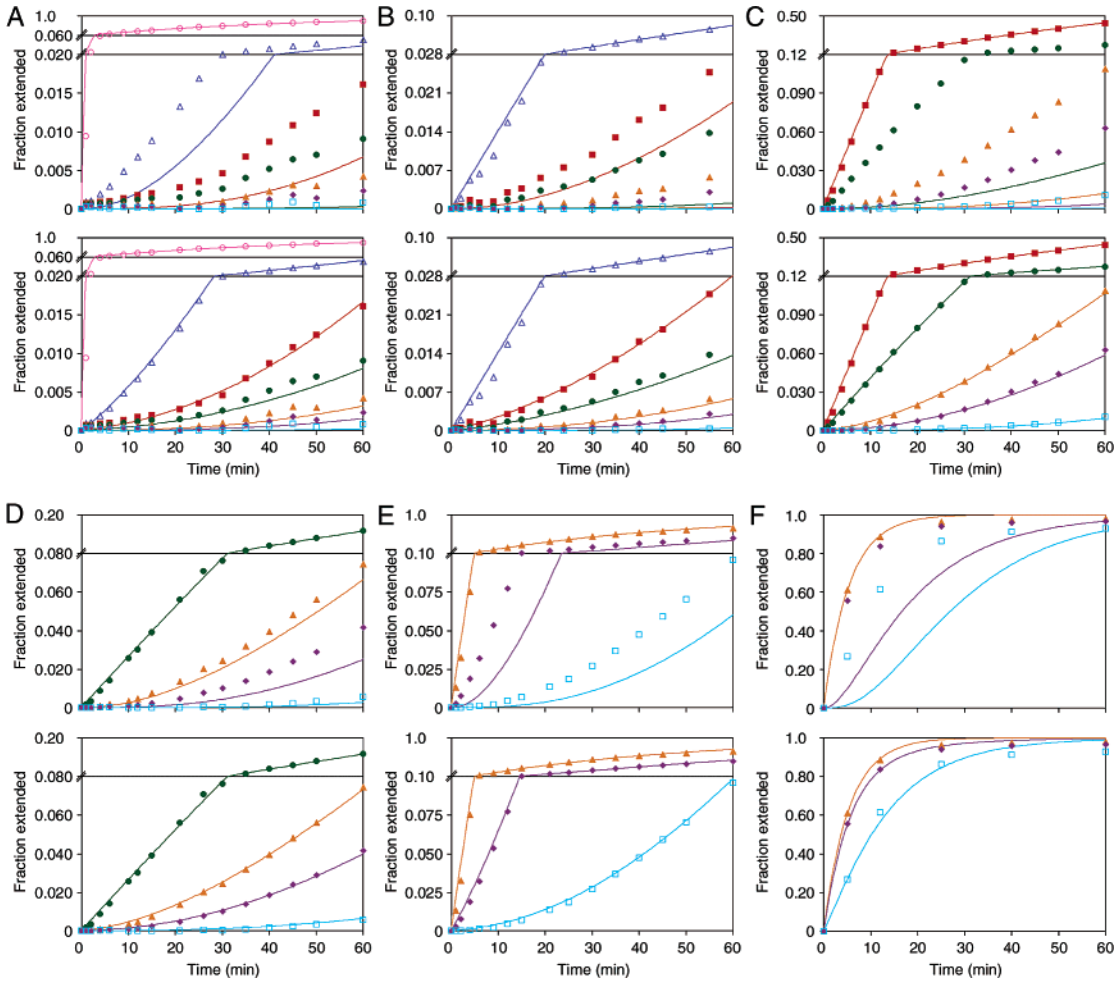


FIGURE 5: Polymerization of many nucleotides. (A) Starting from PT1. (B) Starting from PT2. (C) Starting from PT3. (D) Starting from PT4. (E) Starting from PT5. In panels A−E, the 18-nucleotide template was used. (F) Starting from PT5 and using the 21-nucleotide template. In panels A−F, top and bottom plots show the same data, measured from polymerization assays. The top plot superimposes the predictions of the distributive model, and the bottom plot superimposes the predictions of the processive model. Plotted is the fraction of starting material that has been extended to at least PT2 (pink lines and empty circles), at least PT3 (blue lines and empty triangles), at least PT4 (red lines and filled squares), at least PT5 (green lines and filled circles), at least PT6 (yellow lines and filled triangles), at least PT7 (purple lines and filled diamonds), and at least PT8 (light blue lines and empty squares). Upper regions of vertical axes are compressed (except in panel F).

simulation, it yielded correct predictions of the multiple-nucleotide extension observed using the lengthened template. The analysis illustrating the greatest processivity detected thus far [$P_{(5)} = 0.9$ when using the 21-nucleotide template] is shown in Figure 5F. In this context, the processivity is sufficiently high to confer a "running start" advantage: the polymerase generates PT7 from PT5 by two extension reactions faster than it generates PT7 from PT6 by a single extension reaction (Figure 6). This result implies that the polymerase binds PT6 considerably slower than it binds PT5, and that the easier route to productively bound PT6 is through the binding of PT5 and its extension to PT6.

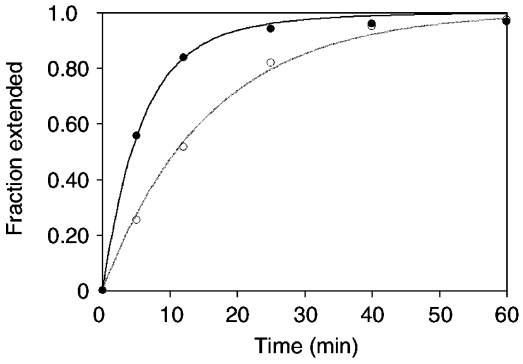


FIGURE 6: Example of a "running start" advantage conferred by processivity. Two time courses of polymerization are shown, one starting from PT5 (●) and the other starting from PT6 (○). Plotted for each time course is the fraction of starting material extended to at least PT7. Processive model time courses are superimposed (lines). Accumulation of PT7 occurs faster in the case of the further removed starting point, PT5.

DISCUSSION

The kinetic parameters of the processive model were all determined on a strictly local basis, yet in aggregate, they succeeded in capturing the experimentally observed global behavior of the system, thereby significantly validating the model. This finding provides strong evidence that the polymerase ribozyme can catalyze the processive addition of multiple nucleotides in a single binding event.

The processivity of this polymerase depends in large part on the particular sequence context of polymerization. In two contexts, its processivity reached 0.9, placing it among low-processivity proteinaceous polymerases such as yeast polymerase η (*12*). In other contexts, its processivity was weak or undetectable. Sequence-dependent variation in k_{obs} was

also seen, a phenomenon observed with proteinaceous polymerases (*16*, *17*). There was no systematic relationship between the k_{obs} and *P* values in different contexts, implying dramatic differences in the rate constants of PT release (k_{off}) in different sequence contexts. Such variation must also explain the differences in *P* when measured using the 18- and 21-nucleotide templates.

Given the weak affinity of the polymerase for the PT, the detectable processivity of the polymerase came as a surprise. Although the polymerase has great difficulty in achieving productive PT binding and alignment, once it succeeds, the extension reaction occurs with great speed, so fast, in fact, that the reaction can occur multiple times before the complex dissociates. This finding makes sense considering that the catalytic core of the polymerase is a ligase ribozyme capable of promoting a self-ligation reaction with a k_{cat} of 500 min^{-1} (*18*). It appears that much of the innate catalytic ability of the ligase was retained when its core was moved into the context of the polymerase pool. Indeed, when the polymerase was given the advantage of a covalent tether to its primer, its polymerization rate increased 300-fold (data not shown).

For a ribo-organism in the RNA world, processivity could have been instrumental in overcoming the problem of strand displacement. Because of the tremendous stability of long RNA duplexes, dislodging a newly synthesized RNA strand from its template is a formidable task. One way of achieving it is to wait until the subsequent round of synthesis, letting the next nascent RNA strand gradually displace the previous one as it is synthesized (*19*). This strategy requires that the polymerase be processive, for if the polymerase fell off the template early in the synthesis, the previous full-length product would simply reanneal, displacing the much shorter nascent RNA strand.

The finding that the polymerase is currently limited by its weak affinity for the PT suggests that future *in vitro* selection experiments with this polymerase ribozyme should target PT binding rather than only chemical proficiency and NTP binding. Doing so could increase its efficiency to the point where it would be able to synthesize a complementary strand of its own length. The polymerase is nearly 200 nucleotides in length, but with its current limitations, it can polymerize only ~14 nucleotides in 6 h. Longer incubations yield little further extension because the incubation conditions degrade the polymerase. However, computational modeling can provide an estimate of how long it would take the polymerase to synthesize its complementary strand if it were freed from the constraints imposed by the incubation conditions. We modeled polymerization starting from an eight-nucleotide primer, using a 200-nucleotide template. We assigned values of k_{obs} and *P* along the template by reiterating the succession of parameters from Table 1. This was done so that our model would take into account the large fluctuations in both parameters observed along a given template. Using the values for the 18-nucleotide template, we found that the polymerase would require more than 3 weeks to fully extend 50% of the starting material. For the 21-nucleotide template values, the polymerase would require just more than 2 weeks. Next we modeled the polymerization that would be expected if the PT affinity of the polymerase could be improved 100-fold (decreasing the K_d of PT6, for example, to ~7 μM). We modeled this improvement as a 100-fold decrease in k_{off} at each position along the template. This decrease in k_{off} was transferred to the parameters of both the processive and nonprocessive modes of extension as follows. First, because *P* is determined by the k_{cat}/k_{off} ratio, *P* was increased at each position according to the 100-fold increase in that ratio. Second, because the observed rates of nonprocessive extension starting from unbound material (k_{obs}) were limited by PT affinity, k_{obs} at each position was increased 100-fold. The model using the 18-nucleotide template parameters predicted that the improved polymerase would require just more than 1.5 h to fully extend 50% of the primer. Using the 21-nucleotide template parameters, it would require just more than 4 h. These time scales are within the polymerase lifetime imposed by the incubation conditions. Thus, sufficient gains in PT affinity may eventually yield the efficient polymerization that would be required of an RNA replicase.

ACKNOWLEDGMENT

We thank Uli Müller, Ed Curtis, and other members of the Bartel lab for helpful discussions.

REFERENCES

1. Orgel, L. E. (1968) Evolution of the genetic apparatus, *J. Mol. Biol. 38*, 381−393.
2. Eigen, M. (1971) Self-organization of matter and the evolution of biological macromolecules, *Natürwissenschaften 58*, 465−523.
3. Pace, N. R., and Marsh, T. L. (1985) RNA catalysis and the origin of life, *Origins Life 16*, 97−116.
4. Gilbert, W. (1986) The RNA world, *Nature 319*, 618.
5. Orgel, L. E. (1986) RNA catalysis and the origins of life, *J. Theor. Biol. 123*, 127−149.
6. Joyce, G. F., and Orgel, L. E. (1999) in *The RNA World* (Gesteland, R. F., Cech, T. R., and Atkins, J. F., Eds.) pp 49−77, Cold Spring Harbor Laboratory Press, Plainview, NY.
7. Johnston, W. K., Unrau, P. J., Lawrence, M. S., Glasner, M. E., and Bartel, D. P. (2001) RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension, *Science 292*, 19−25.
8. Fay, P. J., Johanson, K. O., McHenry, C. S., and Bambara, R. A. (1981) Size classes of products synthesized processively by DNA polymerase III and DNA polymerase III holoenzyme of *Escherichia coli*, *J. Biol. Chem. 256*, 976−983.
9. Kornberg, A., and Baker, T. A. (1991) *DNA Replication*, W. H. Freeman, New York.
10. McClure, W. R., and Chow, Y. (1980) The kinetics and processivity of nucleic acid polymerases, *Methods Enzymol. 64*, 277−297.
11. Xiang-Peng, K., Onrust, R., O'Donnell, M., and Kuriyan, J. (1992) Three-dimensional structure of the beta subunit of *E. coli* DNA polymerase III holoenzyme: A sliding DNA clamp, *Cell 69*, 425−437.
12. Washington, M. T., Johnson, R. E., Prakash, S., and Prakash, L. (1999) Fidelity and processivity of *Saccharomyces cerevisiae* DNA polymerase *η*, *J. Biol. Chem. 274*, 36835−36838.
13. Bambara, R. A., Uyemura, D., and Choi, T. (1978) On the processive mechanism of *Escherichia coli* DNA polymerase I, *J. Biol. Chem. 253*, 413−423.
14. Detera, S. D., Becerra, S. P., Swack, J. A., and Wilson, S. H. (1981) Studies on the mechanism of DNA polymerase *a*, *J. Biol. Chem. 256*, 6933−6943.
15. Detera, S. D., and Wilson, S. H. (1982) Studies on the mechanism of *Escherichia coli* DNA polymerase I large fragment, *J. Biol. Chem. 257*, 9770−9780.
16. Echols, H., and Goodman, M. F. (1991) Fidelity mechanisms in DNA replication, *Annu. Rev. Biochem. 60*, 477−511.
17. Kunkel, T. A. (1992) Biological asymmetries and the fidelity of eukaryotic DNA replication, *BioEssays 14*, 303−308.
18. Glasner, M. E., Bergman, N. H., and Bartel, D. P. (2002) Metal ion requirements for structure and catalysis of an RNA ligase ribozyme, *Biochemistry 41*, 8103−8112.
19. Bartel, D. P. (1999) Re-creating an RNA replicase, *The RNA World (2nd Ed.)*, 143−162.

BI034228L