Biophysical Journal

Supporting Material

Transient-State Kinetic Analysis of the RNA Polymerase I Nucleotide Incorporation Mechanism

Francis D. Appling,[1] Aaron L. Lucius,[2,*] and David A. Schneider[1,*]

[1]Department of Biochemistry and Molecular Genetics and [2]Department of Chemistry, University of Alabama at Birmingham, Birmingham, Alabama

Supplemental Methods

Buffers

With the exception of electrophoresis components, all buffers were filtered through 0.22 μm Millipore express plus vacuum-driven filters (Millipore) unless specified otherwise.

Buffer A: (40 mM KCl, 20 mM Tris-OAc, pH 7.9 at 25 °C); Buffer B: (40 mM KCl, 20 mM Tris-Acetate (OAc) pH 7.9 at 25 °C, 2 mM dithiothreitol, 0.2 $mg \cdot ml^{-1}$ bovine serum albumin (BSA)); Protein measurement buffer: (6 M guanidine-HCl, 0.05 M K_2HPO_4, pH adjusted to 7.0 with phosphoric acid); Storage buffer: (0.55 M K-OAc, 10 mM K-HEPES, 0.5 mM $MgCl_2$, 45% (v/v) glycerol pH 7.8); Gel loading buffer: (90% formamide, 25 mM EDTA-Na pH 8.54, 0.025 $mgml^{-1}$ bromophenol blue); Extraction buffer: (10 mM Tris-Cl pH 7.9, 1 mM EDTA-NA pH 8.54); Digestion buffer: (Tris-Cl pH 9.2, 3 mM $MgCl_2$); NA measurement buffer: (100 mM NaCl, 50 mM K_2HPO_4, pH adjusted to 7.0 with phosphoric acid); Neutralization buffer: (2.67 M Tris-Cl pH 8.4)

Proteins

Pol I was purified according to (1). Pol I is stored in storage buffer at -20 °C.

Heparin and bovine serum albumin (BSA) preparation

BSA stocks were prepared by dissolving lyophilized BSA (Fisher Scientific) in buffer A to a concentration of ≈40 $mg \cdot ml^{-1}$ and passing through a 0.2 μm nylon syringe filter (Fisher Scientific). This solution was dialyzed against three 1:500 exchanges of buffer A at 4 °C using 3,500-5,000 MWCO cellulose ester dialysis tubing (Spectra/Por). The concentration of BSA was determined by a spectroscopic assay in protein measurement buffer by taking absorbance scans between 360 nm and 220 nm. Absorbance values at 280 nm were corrected by subtracting absorbance due to scattering effects at 320 nm and an extinction coefficient of 48,150 $M^{-1} \cdot cm^{-1}$ was used to calculate concentration as described (2).

Heparin stocks were prepared by dissolving Na-heparin salt (Acros Organics) in buffer A to achieve ≈40 $mg \cdot ml^{-1}$ and filtering through a 0.2 μm nylon syringe filter (Fisher Scientific). This solution was dialyzed against three 1:500 exchanges of buffer A at 4 °C using 3,500-5,000 MWCO cellulose ester dialysis tubing (Spectra/Por). The concentration of heparin stocks were determined by a colorimetric assay (3). Absorbance at 505 nm of a 5:1:1 mixture of heparin, 0.06 M Barbital buffer (Sigma Aldrich), and 0.09 $mg \cdot ml^{-1}$ azure A was measured and concentration was quantified by comparison to a standard curve.

Nucleic Acids

RNA oligonucleotides were commercially synthesized and purified by HPLC by the manufacture (Integrated DNA Technologies; IDT). RNA oligonucleotides were dialyzed against three 1:5000 exchanges of buffer A at 4 °C using 100-500 MWCO cellulose ester dialysis tubing (Spectra/Por) and were stored at -20 °C.

DNA oligonucleotides were synthesized and de-salted by IDT. DNA oligonucleotides were further purified by denaturing polyacrylamide gel electrophoresis (PAGE) using a method based on that described by Ellington and Pollard (4). Specifically:

Lyophylized oligonucleotide was dissolved in ultrapure water at room temperature for at least 30 min to an estimated concentration of >200 μM. Oligonucleotide solution is diluted into gel loading buffer to achieve a concentration of 10-20 μM oligonucleotide. This solution was loaded onto 20 well (16 cm wide) 0.75 mm thick 18% polyacrylamide gels (19:1 acrylamide:bisacrylamide, 7 M urea, TBE). Resolution falls with increasing amounts of oligonucleotide loaded per lane and we have found that 150-300 pmole·lane^{-1} provides an acceptable balance between purity and yield. Material was electrophoresed at constant temperature (53 °C) in TBE which usually results in ≈1,500 V. Gels are laid on top of thin layer chromatography plates with UV-shadowing capability (Millipore) and exposed to a 254 nm UV lamp. Oligonucleotides were located by UV shadowing and excised with a stainless steel razor blade. Gel slices are crushed, mixed with extraction buffer to achieve 11.11 ml·g^{-1} (buffer/gel), and frozen at -80 °C. Samples were rapidly thawed by incubation in a 65 °C water bath and were incubated at 95 °C after thawing for 5 min. Samples were shaken for 16 hrs in an air shaker at 37 °C. Gel fragments were pelleted by centrifugation for 2 min at 1,000 times g, room temperature and solution was removed and filtered through 0.2 μm nylon syringe filters (Fisher Scientific). Nucleic acid was concentrated by extracting water with butanol. Oligonucleotide was precipitated by adding 0.3 volumes of 3 M Na-OAc pH 5.2 and 1.5 volumes (total volume after addition of Na-OAc) of 100% ethanol and incubated >30min at -20 °C. To calculate volumes of 3 M Na-OAc pH 5.2 and ethanol to add to oligonucleotide solutions, volume of oligonucleotide solution was estimated by weight assuming a density of 1 g·ml^{-1}. Precipitated nucleic acid was pelleted by centrifugation for 10min at 17,000 times g, 4 °C. Supernatant was removed and the precipitate was dried 15 min at room temperature. Precipitate was dissolved with 105 μL buffer A for >15 min at room temperature.

DNA oligonucleotides were dialyzed against three 1:5000 exchanges of buffer A at 4 °C using 3,500-5,000 MWCO cellulose ester dialysis tubing (Spectra/Por) and were stored at -20 °C. We did not observe degradation after long-term storage of RNA or DNA oligonucleotides under these conditions. Purity and integrity of nucleic acids were assayed by 5'-end labeling a small sample with ^{32}P and standard T4 polynucleotide kinase approaches. Labeled samples were separated on denaturing PAGE gels as described for nucleic acid purification, exposed to phosphoimager plates, and scanned with a typhoon imager for analysis.

The concentrations of oligonucleotides were determined by a spectroscopic assay. Nucleic acids were enzymatically degraded to nucleoside monophosphates by incubation with phosphodiesterase I (PDEI) (Worthington Biochemical Corp.)(5). PDEI was dissolved and stored in digestion buffer supplemented with 40% glycerol (v/v) at -20 °C. Typically 5 μL of dialyzed nucleic acid was digested in a 50 μL reaction. Digestion reactions were carried out in digestion buffer with ≈12 u·ml^{-1} PDEI at room temperature >20 min. Absorbance scans were collected between 360 nm and 220 nm in NA measurement buffer. Absorbance values at 260 nm were corrected by subtracting absorbance due to scattering effects at 320 nm. Oligonucleotide extinction coefficients were calculated by summing the extinction coefficients of each NMP multiplied by their stochiometric constant.

Extinction coefficients were obtained from Gray, Hung, and Johnson (6).

All data presented here were collected using ECs assembled with the following nucleic acid sequences:

RNA(5'-AUCGAGAGG)

DNA template strand (DNA-t) (5'
ACCAGCAGGCCGATTGGGATGGGTATTCCCTCCTGCCTCTCGATGGCTGTAAGTATC
CTATAGG)

DNA non-template strand (DNA-nt) (5'
CCTATAGGATACTTACAGCCATCGAGAGGCAGGAGGGAATACCCATCCCAATCGGC
CTGCTGGT)

RNA – DNA-t annealing reactions were carried out as follows: a solution of 10 μM RNA, 3.33 μM DNA-t in buffer A was heated to 95 °C for 5 min then cooled at a rate of 0.1 $°C \cdot s^{-1}$ to a final temperature of 4 °C.

EC Assembly

ECs were formed in buffer B at ambient temperature (≈20-23 °C). BSA and DTT aliquots were stored at 10X concentration at -20 °C and were thawed and used once. Pol I stock was diluted 1:40 into elongation complex assembly reactions.

Due to the sequential nature of the EC formation procedure the concentrations of reactants change during assembly due to dilution. The following list of concentrations corresponds to final reactant concentrations. This list will be followed by an assembly protocol in which each added component is presented as a fraction of the total reaction volume. The bold numbers on the far left of the reagent list denote the order of addition.

We found that under our conditions single nucleotide incorporation time courses are very sensitive to the presence of DNA-nt. In our experiments we used a DNA-nt / DNA-t ratio of 3 / 1. We found that beyond a DNA-nt / DNA-t ratio of 2 / 1 single nucleotide incorporation time courses are insensitive to further increase in [DNA-nt].

(≈16 nM Pol I, 162.75 nM RNA, 54.26 nM DNA-t, 162.75 nM DNA-nt, ≈5 nM α-^{32}P-CTP, 100 μM $Mg(OAc)_2$, 1.1 mM ethylenediamine teraacetic acid (EDTA) – K_3, 5 μM unlabeled CTP)

1) 0.07675 10X concentrated buffer A
2) 0.1 2 $mg \cdot ml^{-1}$ BSA
3) 0.1 20 mM DTT
4) 0.5657 H_2O
5) 0.025 Pol I
6) 0.016275 RNA – DNA-t (20min incubation)
7) 0.016275 DNA-nt (40min incubation)
8) 0.05 labeling mixture (2 mM $Mg(OAc)_2$, 100 nM α-^{32}P-CTP; in buffer A) (10min incubation)

9) 0.05 EDTA / unlabeled CTP chase mixture (22 mM EDTA-K_3, 100 μM unlabeled CTP; in buffer A)

ATP Substrates

Adenosine triphosphate (ATP) substrates were purchased from Sigma Aldrich as lyophilized sodium salts. ATP substrate stocks were prepared by dissolving in buffer A, filtering through 0.2 μm nylon syringe filters (Fisher Scientific), and dialyzing extensively into buffer A at 4 °C using 100-500 MWCO cellulose ester dialysis tubing (Spectra/Por).

ATP substrate mixtures for polymerization/nuclease reactions are prepared in buffer B and supplemented with 1.1 mM EDTA-K_3. In ATP substrate mixtures ATP, heparin, and $Mg(OAc)_2$ are adjusted to 2X concentration (with respect to reaction concentrations) to result in 1X concentration following a one to one mixing event.

Quenched Flow Time Courses

Substrate ATP mixtures were prepared in buffer B supplemented with $Mg(OAc)^2$ and ATP substrate to the desired concentration and heparin to a concentration of 0.05 $mg \cdot ml^{-1}$. Drive syringes were filled with ultrapure ddH_2O and quench syringe was filled with 1 M HCl. Each reaction was initiated by rapidly mixing 20 μL EC solution with 20 μL ATP substrate solution. Reactions were quenched by delivery of 1 M HCl. The volume of quench solution delivered is dependent upon the reaction loop and excess 1M HCl was added to bring final quench solution volume to 200 μL. Quenched reactions were neutralized by the addition of 115 μL of neutralization buffer. An aliquot of neutralized reaction was mixed in a 1:2 ratio with gel loading buffer and placed on ice for the remainder of time course collection. Samples were stored at -20 °C. We have not observed degradation of reactant or product nucleic acids during storage.

Quenched flow time courses were thermo stated at 25 °C by a circulating water bath. The order of time point collection was randomized by Microsoft Excel's random number generation function.

Nucleotide Addition Reaction Electrophoresis

Samples were heated to 95 °C >5 min and 15 μL of sample per well were loaded onto gels of the dimensions described for nucleic acid purification. Gels for analysis of nucleotide incorporation reactions were 28 % acrylamide (19:1 acrylamide:bisacrylamide), 7 M urea, TBE. A gradient buffer system was employed in which the top tank contains TBE and the bottom tank contains TBE supplemented with 0.3 M Na-OAc. We have found that it is important to pre-run gradient buffer gels ≈1 hr at 53 °C to establish the optimal state of the gradient. During nucleic acid separation gels are run at 53 °C (≈900 V). Gels were exposed and imaged as described for analysis of nucleic acid purity.

Data Analysis

Exposed gels were quantified according to EQ. S1,

$$Frac_i(t) = \frac{\frac{[RNA_i](t)}{\sum_i [RNA_i](t)} - \frac{[RNA_i](0)}{\sum_i [RNA_i](0)}}{1 - \frac{[RNA_i](0)}{\sum_i [RNA_i](0)}} \quad \text{S1}$$

where $Frac_i(t)$ is the fraction of RNA in the state i at time t and $[RNA_i](t)$ is the pixel counts of RNA corresponding to RNA in state i at time t. The summations in EQ. 2 included pixel counts corresponding to 10-mer, 11-mer, and CA. Quenched flow nucleotide addition time courses were collected three times and the time course data points plotted in figure 3 of the text represent the average data point. The uncertainty bars plotted in figure 3 of the text represent the standard deviation about this average.

The majority of data analysis was performed using custom built MATLAB functions. All code can be made available upon request.

Fitting to Analytical Functions

11-mer and CA time courses were fit individually by EQ. S2

$$Frac_i(t) = \sum_{j=1}^{n} \alpha_j e^{-k_{obs} t} \quad \text{S2}$$

to obtain observed rate constants k_{obs}. For 11-mer time courses $j = 2$ and for CA time courses $j = 1$. The observed rate constants obtained from each experimental replicate were averaged. This analysis yielded three observed rate constants corresponding to $11-mer\ k_{obs1}$, $11-mer\ k_{obs2}$, and $CA\ k_{obs}$. The uncertainty associated with the observed rate constant values obtained from fitting time course data to EQ. S2 were assumed to be equal to the standard deviation about the average of the three observed rate constants obtained from fitting experimental replicates. Observed rate constants were plotted as a function of [ATP]. *11-mer* k_{obs1} was fit using EQ. S4

$$11-mer\ k_{obs1}([ATP]) = \frac{k_{\max}[ATP]}{K_{1/2} + [ATP]} \quad \text{S3}$$

where $11-mer\ k_{obs1}([ATP])$ is the observed rate constant as a function of [ATP], $k_{\max}$ is the extrapolated observed rate constant at infinite [ATP], and $K_{1/2}$ is the [ATP] at which $11-mer\ k_{obs1}([ATP]) = 0.5k_{\max}$. $11-mer\ k_{obs2}([ATP])$ and $CA\ k_{obs}([ATP])$ were fit using EQ. S5

$$k_{obs}([ATP]) = k_{obs} + 0[ATP] \quad \text{S4}$$

which is a constant function. Analysis using EQ.s S2-S4 was performed in MATLAB using the MATLAB function Lsqnonlin. The data were fit to EQ.s S4 and S5 by minimizing the weighted sum of squared residuals (WSSR) in which each data point was weighted by the standard deviation calculated for each observed rate constant value. Uncertainty on parameter values obtained from fitting the data to EQs S3 and S4was calculated by grid searching (see below).

Fitting by Numerical Integration

11-mer and CA time courses were globally fit using the ordinary differential equations (ODEs) derived from each reaction scheme. Specifically, WSSR was calculated by summing weighted residuals of each time course (both 11-mer and CA time courses collected at each [ATP]). Numerical integration was performed in MATLAB using the Ode23s numerical integration algorithm. During numerical integration the jacobian matrix (EQ. S5) defined by the ODEs was passed to the numerical integration algorithm.

$$\frac{\partial\left\{\frac{ds_i}{dt}\cdots\frac{ds_n}{dt}\right\}}{\partial\{s_i\cdots s_n\}}=\begin{pmatrix}\frac{\partial\frac{ds_i}{dt}}{\partial s_i} & \cdots & \frac{\partial\frac{ds_i}{dt}}{\partial s_n}\\ \vdots & \ddots & \vdots\\ \frac{\partial\frac{ds_n}{dt}}{\partial s_i} & \cdots & \frac{\partial\frac{ds_n}{dt}}{\partial s_n}\end{pmatrix} \qquad \text{S5}$$

In EQ. 6 s_n denotes the concentration of an individual species in a given reaction pathway. When fitting by numerical integration of ODEs WSSR was minimized using a custom built algorithm in MATLAB. This algorithm is based on a genetic algorithm but uses a gradient-based non-linear least squares algorithm (GBNLLS) to further minimize WSSR at each iteration. The structure of the algorithm is briefly described below.

A population of parameter sets is defined by randomly selecting individual parameter values from a Gaussian distribution of defined mean and standard deviation. WSSR is calculated for each individual parameter set in the population by the above described numerical integration procedure. The population is ranked according to WSSR. Like standard genetic algorithms; a subset of the population advances to the next iteration (generation) unaltered, a subset is subjected to mutation at a specified probability, and a subset is subjected to crossover at a specified probability (7). In our algorithm a subset of the population is further optimized by the GBNLLS algorithm Lsqnonlin. The number of individuals and the identity of individuals subjected to GBNLLS is adjustable. All fitting by numerical integration in this study was accomplished by subjecting four individuals to GBNLLS at each iteration of the algorithm. Three individuals are the top three based on WSSR and one is chosen randomly from the population. After a specified number of iterations the parameter values and WSSR is reported.

In globally fitting Schemes 1 and 2 we observed that the rate constants governing the steps corresponding to pyrophosphate release and dinucleotide production (k_5 and k_6) were interchangeable with little effect on the fit. This lack of constraint is due to the fact that only one of these steps must be rate limiting (with a rate constant of $\approx$0.4 s^{-1}) to describe the time course

data. Both published (8) and unpublished data (data not shown) indicate that during processive elongation Pol I catalyzes nucleotide incorporation at ≈20-50 nucleotides s^{-1}. These rates are not consistent with a rate constant governing pyrophosphate release (a step that must occur after ever nucleotide addition event) of ≈0.4 s^{-1}. We therefore constrained the rate constant governing pyrophosphate release (k_5) in Schemes 1 and 2 to a lower limit of 10 s^{-1} in all global fitting routines.

$k_{\mathbf{obs}}\left(\left[\mathbf{ATP}\right]\right)$ *Derivation and Simulation*

$k_{obs}\left(\left[ATP\right]\right)$ are the negative eigenvalues of the coefficient matrix defined by the DEs. Specifically, DEs derived from reaction pathways studied in this work are of the form EQ. S6

$$\frac{d}{dt}\mathbf{c} = \mathbf{M}\mathbf{c} \qquad \text{S6}$$

where **c** is the solution vector of concentrations of species in a given reaction pathway and **M** is the coefficient matrix. Importantly, for the sake of derivation, these equations were derived under pseudo first order conditions, i.e. [ATP]=constant. This does not affect the comparison of simulated $k_{obs}\left(\left[ATP\right]\right)$ functions to experimental data because all reactions were performed under pseudo first order conditions with respect to ATP, i.e. [EC]<<[ATP]. The negative analytical eigenvalues of **M** were found using MATLAB. The parameter values obtained from fitting DEs by numerical integration were substituted into the analytical eigenvalues. k_{obs} values were calculated over a range of [ATP] and the real component of the k_{obs} values were compared to experimental data.

Grid Searching

Grid searching was accomplished by systematically varying a single parameter's value and subsequently minimizing WSSR by allowing all other parameter values to be optimized. Each WSSR minimization was accomplished by Lsqnonlin. F-statistics at each parameter value were calculated according to EQ. S7

$$F-stat\left(param\ value\right) = \frac{WSSR_{grid}\left(param\ value\right)}{WSSR_{fit}} \qquad \text{S7}$$

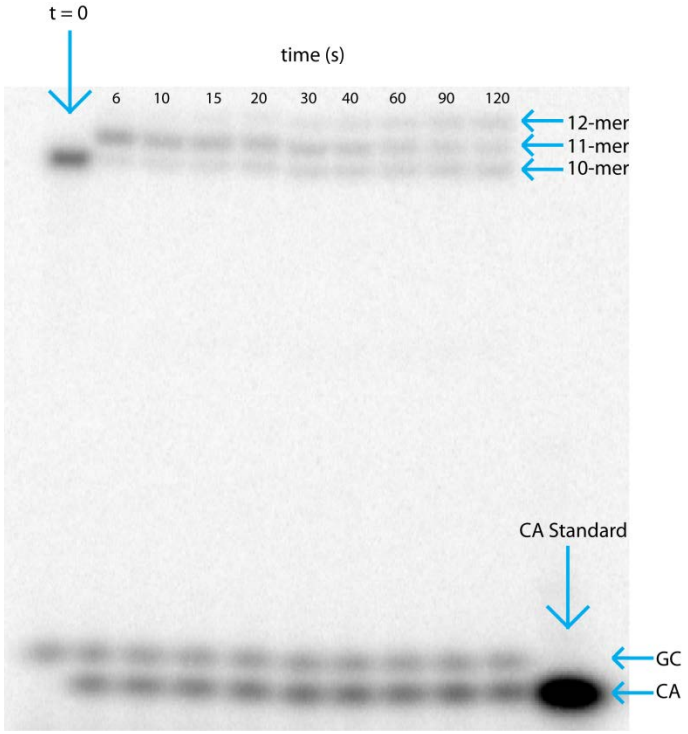
where the WSSR subscript refers to the best-fit WSSR value obtained from the original fit (*fit*) or that obtained at a fixed parameter value during a given grid search iteration (*grid*); note: $WSSR_{grid}$, and subsequently $F-stat$, are functions of the parameter values of the parameter being scanned. An F-critical ($F-crit$) value at the 68% confidence interval was calculated using Microsoft Excel's FINV function. The relationship between each parameter and F-statistics ($F-stat\left(param\ value\right)$) was defined by polynomial interpolation of F-statistic vs parameter value data in the vicinity of $F-stat = F-crit$. Polynomial fitting was performed using Kaleidagraph. *param value* at $F-crit$ was calculated using MATLAB by finding the roots of the polynomial used to fit the F-stat vs parameter value data. The roots defining the

parameter bounds of a given parameter were chosen based on inspection of $F-stat$ vs $param\ value$ plots. Simply stated, the parameter value bounds are defined by where the interpolating polynomial defining $F-stat\left(param\ value\right)$ crosses the F-crit threshold.

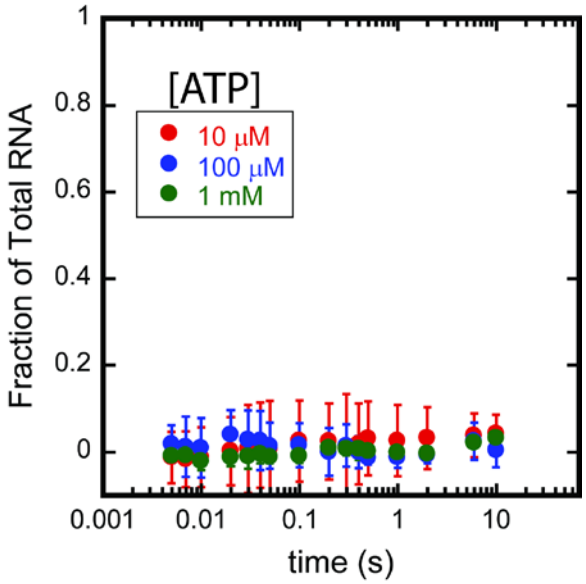
Supplemental References

1. Appling, F. D., and D. A. Schneider. 2015. Purification of active RNA polymerase I from yeast. Methods in molecular biology 1276:281-289.
2. Gill, S. C., and P. H. v. Hippel. 1989. Calculation of Protein Extinction Coefficients from Amino Acid Sequence Data. Analytical Biochemistry:319.
3. Jaques, L. B. 1977. Determination of Heparin and Related Sulfated Mucopolysaccharides. Methods of Biochemical Analysis 24:203-312.
4. Ellington, A., and J. Jack D. Pollard. 1998. Purification of Oligonucleotides Using Denaturing Polyacrylamide Gel Electrophoresis. Current Protocols in Molecular Biology:2.12.11-12.12.17.
5. Lucius, A. L., A. Vindigni, R. Gregorian, J. A. Ali, A. F. Taylor, G. R. Smith, and T. M. Lohman. 2002. DNA Unwinding Step-size of E.coli RecBCD Helicase Determined from Single Turnover Chemical Quenched-flow Kinetic Studies. Journal of molecular biology 324:409-428.
6. Gray, D. M., S. H. Hung, and K. H. Johnson. 1995. Absorption and Circular Dichroism Spectroscopy of Nucleic Acid Duplexes and Triplexes. Methods in Enzymology 246:19-34.
7. Back, T., U. Hammel, and H.-P. Schwefel. 1997. Evolutionary Computation: Comments on the History and Current State. IEEE Transactions on Evolutionary Computation 1:3-17.
8. Viktorovskaya, O. V., K. L. Engel, S. L. French, P. Cui, P. J. Vandeventer, E. M. Pavlovic, A. L. Beyer, C. D. Kaplan, and D. A. Schneider. 2013. Divergent contributions of conserved active site residues to transcription by eukaryotic RNA polymerases I and II. Cell reports 4:974-984.

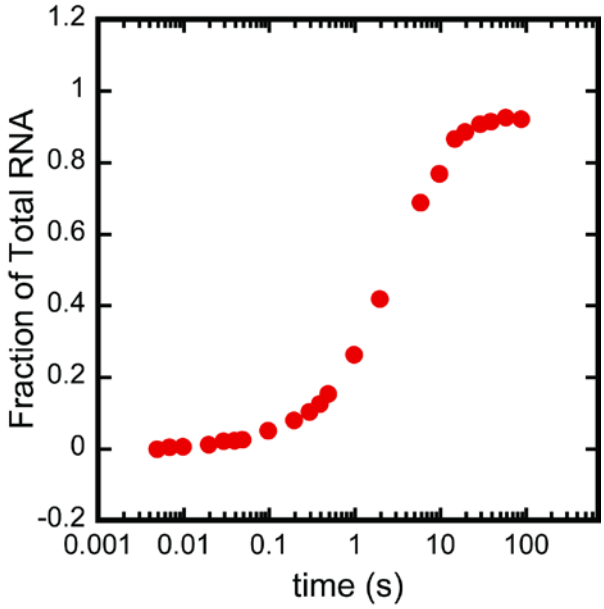
Supplemental Figures



Supplemental Figure S1. Identification of RNA species. This is a representative gel of an AMP incorporation time course collected by hand. The reactions were carried out as described in the text and materials and methods sections for quenched flow time courses except all volumes were reduced by half and time points were collected by hand. As opposed to the short time scale monitored using the quenched flow instrument, on this time scale 12-mer RNA is observed. We do not know the sequence of this RNA but it is likely due to misincorporation or trace contamination of NTP other than ATP. Its presence on these long time scales does not affect the interpretation of our data. The first lane on the left represents t = 0 material. For this lane labeled elongation complexes were denatured and electrophoresed. In this lane two bands are visible corresponding to 10-mer and GC RNA. The 10-mer RNA is the expected species following incorporation of a single ^{32}P-labeled CMP into the 9-mer substrate RNA converting unlabeled 9-mer RNA into labeled 10-mer RNA. The GC dinucleotide is the expected product from a nuclease reaction removing the incorporated ^{32}P-labeled CMP and the terminal GMP from the synthetic RNA. Beginning with the second lane from the left and moving from left to right each lane represents an increased delay time (denoted by numbers above each lane) between addition of ATP substrate and addition of HCl quench. The appearance of a slower migrating species is consistent with a single AMP incorporation extending the 10-mer RNA to the 11-mer state. The appearance of a species migrating slightly faster than GC is consistent with a dinucleotide production event immediately following incorporation of an AMP. This species co migrates with a CA standard.



Supplemental Figure S2. [GC] vs. time at 10 μM, 20 μM, and 1mM ATP. The fraction of RNA in the GC state was quantified according to EQ S1 (supplemental materials and methods). The circles represent the average of three measurements and the error bars represent the standard deviation about this average.



Supplemental Figure S3. CA time course collected at 300 μM ATP. This time course demonstrates that the ≈ 0.75 peak height values of 11-mer discussed in the text are not due to enzyme inactivation or RNA degradation.