Supporting Information

Larson et al. 10.1073/pnas.1200939109

SI Results and Discussion

Fitting the *F-v* Data to Different Kinetic Models. To determine the location of the translocation step within the nucleotide-addition cycle, as well as the biochemical steps that are affected by ammonium (and related) ions, we fit the global data to each of four different Brownian ratchet-type models (Fig. 2). Three of these pathways place the NTP binding event before, after, or either before or after translocation, but prior to pyrophosphate release; the fourth pathway allows pyrophosphate release to occur either before or after translocation. In Model 1 (Fig. 2*A*), RNA polymerase II (RNAPII) oscillates stochastically between the pre- and posttranslocated states with thermal motion. Assuming a Boltzmann-type behavior for the load-dependent step, this model predicts that the pause-free elongation rate, v, will depend upon the applied force and NTP concentrations as follows

$$v(F,[\mathrm{NTP}]) = \frac{k_{\mathrm{cat}}}{1+\frac{K_D}{[\mathrm{NTP}]}(1+K_\delta\cdot\exp[-F\cdot\delta/k_\mathrm{B}T])}, \qquad \textbf{[S1]}$$

where K_D is the NTP dissociation constant, K_δ describes the equilibrium between the pre- and posttranslocated states ($K_\delta = [\mathrm{pre}]/[\mathrm{post}]$), and δ is a characteristic distance parameter, which here represents the distance between the pre- and posttranslocated states (fixed to 1 bp = 0.34 nm in these fits). The condensation reaction and PP_i release steps are not separately visualized in the single-molecule assay, and are therefore combined here into a single kinetic parameter representing the catalysis event, k_{cat}.

In Model 2 (Fig. 2*B*), substrate binding occurs prior to translocation, suggesting that the incoming NTP drives translocation (1), and the thermal motion between the pre- and posttranslocated states is rectified by the NTP condensation reaction. Model 2 predicts the following *F-v* behavior:

$$v(F,[\mathrm{NTP}]) = \frac{k_{\mathrm{cat}}}{1+K_\delta\cdot\exp[-F\cdot\delta/k_\mathrm{B}T]\left(1+\frac{K_D}{[\mathrm{NTP}]}\right)}. \qquad \textbf{[S2]}$$

In Model 3, the incoming NTP can bind in either the pre- or posttranslocated state (Fig. 2*C*) (2), We assume that both binding sites are energetically equivalent, which leads to identical equilibrium constants for NTP binding the pre- and posttranslocated states. As such, Model 3 has the same number of free parameters ($m = 3$) as both Models 1 and 2, and predicts the following *F-v* relationship:

$$v(F,[\mathrm{NTP}]) = \frac{k_{\mathrm{cat}}}{1+\frac{K_D}{[\mathrm{NTP}]}\{1+K_\delta\cdot\exp(-F\cdot\delta/k_\mathrm{B}T)\}+K_\delta\cdot\exp(-F\cdot\delta/k_\mathrm{B}T)}. \qquad \textbf{[S3]}$$

Model 4, in which PP_i release can take place from either the pre- or posttranslocated state (Fig. 2*D*), is inspired by kinetic studies demonstrating that the rate of PP_i release is stimulated by the incoming cognate NTP (3). Similar to the assumption made for Model 3, PP_i is assumed to be equally likely from either the pre- or posttranslocated state, resulting in a *F-v* relationship that has the same number of free parameters as Models 1, 2, and 3:

$$v(F,[\mathrm{NTP}]) = \frac{k_p}{1+\frac{K_D}{[\mathrm{NTP}]}(K_\delta\cdot\exp[-F\cdot\delta/k_\mathrm{B}T])}. \qquad \textbf{[S4]}$$

Global fits to the measured pause-free elongation rates for WT, E1103G, and H1085A/E1103G RNAPII at all our experimental conditions are shown for all four models in Fig. S1.

The experimental data best support the secondary binding-site model (Model 3, Fig. 2*C*), where the NTP can bind RNAPII in either the pre- or the posttranslocated state. As before, we have assumed that the equilibrium constants for both NTP-addition pathways, K_δ and K_D, are identical (i.e., that NTP binding to either the pre- or the posttranslocated state is energetically equivalent) and that the energetics of the translocation event are unchanged by NTP binding. If we relax that particular assumption, we arrive at the more complex Model 5 (Fig. S2*A*). The *F-v* relationship predicted by this model is:

$$v(F,[\mathrm{NTP}]) = \frac{k_{\mathrm{cat}}}{1+\frac{K_D}{[\mathrm{NTP}]}\{1+K_\delta\cdot\exp(-F\cdot\delta/k_\mathrm{B}T)\}+A\cdot K_\delta\cdot\exp(-F\cdot\delta/k_\mathrm{B}T)}. \qquad \textbf{[S5]}$$

Here, an additional free parameter, A, has been introduced, which represents the energetic penalty associated with having an NTP in the secondary binding site whereas the enzyme remains in its pretranslocated position. A value of $A = 0$ would preclude any binding at all in this position, whereas a value of $A = 1$ suggests that binding to either the primary or secondary binding site is energetically equivalent (2). Global fits of the *F-v* curves for WT RNAPII and its trigger loop mutants to various A values spanning an order of magnitude are shown in Fig. S2 *B–D*. Fitting to a range of A values still captured the main features of the *F-v* relationship for the WT RNAPII and its two trigger loop mutants. Moreover, any qualitative differences among the remaining parameters (i.e., changes to k_{cat}, K_D, or K_δ) for WT RNAPII compared to its trigger loop (TL) mutants were unaltered. The fits are therefore not particularly sensitive to the value of A (Model 5) under the conditions of these experiments, making the simpler version represented by Model 3 more attractive.

ITP Misincorporation Assay. The addition of 200 μM ITP to the transcription buffer containing saturating concentrations of NTPs led to a large increase in long pause density (LPD) for WT RNAPII (Fig. 4*B*), consistent with previous single-molecule studies of ITP misincorporation by *Escherichia coli* RNAP (4). By contrast, the long pause density (LPD) for the E1103G mutant enzyme dropped somewhat in the presence of ITP under identical conditions (Fig. 4*B*). The inability of ITP to increase the LPD in the case of E1103G could come about in two ways, in principle, that are not mutually exclusive (1). The E1103G mutant enzyme fails to incorporate ITP into the transcript, thereby precluding any need for mismatch recognition and the subsequent triggering of a long pause, or (2) The E1103G mutant enzyme misincorporates ITP, but thereafter fails to recognize the mismatch and trigger a long pause. To distinguish between these alternatives, we monitored transcript elongation in bulk by both the WT and E1103G enzymes in the presence of ITP.

WT and E1103G ECs were assembled using the preformed synthetic scaffold described in the supporting methods below, resulting in ECs with a 14 nt transcript, and both EC preparations

were incubated with four different NTP mixtures. To correct for differences in the concentration of active ECs, WT, and E1103G ECs were incubated in the presence of *AG mix (0.33 μM α-^{32}P-ATP [Specific Activity: 3, 000 Ci/mmol], and 1.3 μM each of ATP and GTP) for 2 min at 30 °C, allowing extension along an AG cassette (AGGAAA) until the enzyme stalled at A20 due to CTP deprivation. We found that E1103G complexes were present in a 1.8-fold excess compared to the WT (compare lanes 1 and 5 in Fig. S4), and both ECs were able to extend to the end of the template in the presence of *ACG mix (0.33 μM α-^{32}P-ATP, and 1.3 μM each of ATP, CTP, and GTP) as evidenced by lanes 2 and 7. In the presence of *AI mix (0.33 μM α-^{32}P-ATP, and 6.6 μM ITP), both ECs showed extension to the A20 position (lanes 3 and 7), indicating ITP misincorporation at positions coding for GTP. We did not observe extension to A20 in the presence of ATP alone, ruling out the possibility of GTP contamination in the ATP stocks. Additionally, lanes 4 and 8 indicate that the production of run off transcript was inhibited for both WT and E1103G in the *ACI chase (0.33 μM α-^{32}P-ATP, and 1.3 μM each of ATP, CTP, and ITP) compared to the *ACG chase, further confirming the misincorporation of ITP, and not (for example) residual GTP contamination of the ATP stocks. Comparison of lanes 3 and 7 shows that E1103G incorporates approximately equimolar amounts of ITP (1.1-fold excess after normalization) compared to WT, indicating that E1103G is deficient in both substrate selection and mismatch recognition in the case of ITP.

SI Materials and Methods

Materials. All DNA and RNA oligonucleotides were from IDT and were purified by denaturing PAGE before use. FPLC-purified NTPs were from Promega Corp. The α^{32}P-ATP was from Perkin-Elmer. T4 Polynucleotide Kinase and T4 DNA Ligase were from NEB.

Purification of Yeast RNA Polymerase II. pBS1479 (5), a plasmid designed for C-terminal tagging proteins in yeast with the tandem-affinity purification (TAP)-epitope tag, was derivatized to additionally encode a high-affinity peptide [the Avitag (6)] BirA biotinylase sequence upstream of the calmodulin-binding peptide sequence and the TEV cleavage site sequence (Avitag-TAP). The 3′-end of *RPB3* was targeted for fusion of Avitag-TAP encoding sequence in CKY438 (*MAT***a** *his3Δ200 ura3-52 leu2Δ1 or Δ0 met15Δ0 trp1Δ63 lys2-128∂ gal10Δ56 rpb1Δ::natMX*) creating CKY680 (same as above, except *rpb3*::*Avitag-TAP*::*KlTRP1*). CKY680 was used for purification of WT RNAPII or *LEU2 rpb1* plasmids were shuffled into CKY680 by standard procedures to remove pRP112 *RPB1 CEN URA3*. Wild-type or *rpb1* mutant derivatives of CKY680 were then transformed with pRS313-BirA-NLS (7) for the purpose of in vivo biotinylation of RNAPII. His$^+$ transformants were grown up overnight in 5 mL of Synthetic Complete minus histidine (SC-His) medium (8), to maintain selection for pRS313-BirA-NLS, diluted into 200 mL of SC-His medium for 24 hr additional growth, and finally diluted into 12 L of 2× SC-His medium supplemented with biotin to 250 nM and grown to saturation. Biotinylated WT and mutant RNAPII enzymes were purified as described previously (9, 10), except 10-subunit (RNAPII lacking Rpb4/Rpb7) and 12-subunit RNAPII complexes were not pooled following Uno-Q column (BioRad) fractionation, and 12-subunit enzymes were used for single-molecule experiments.

Template DNA. We used the 4.75 kb fragment from the human *POLR2A* gene 5′-region encoding the UTR and several exons and introns as a template to study transcript elongation. The 4.75 kb fragment was PCR-amplified from HeLa cell genomic DNA and blunt end cloned into pJET1/Blunt according to the manufacturer's instructions (Fermentas). The resulting plasmid, pPM172, was used to generate the 4.75 kb fragment to ligate downstream of the scaffold. This was accomplished by PCR amplifying the fragment using the forward primer 5′-CACA GGCGCGCCCACGGGGTGAGCCAGTCACG that coded for a *Dra*III site (underscored) immediately upstream of the sequence of the *POLR2A* gene, and the reverse primer 5′-TGCG GCGGGAACACAACTGG with a 5′-digoxigenin modification. The 2.7 kb DNA fragment to be ligated upstream of the scaffold was PCR amplified from the plasmid pRL702 using the forward primer 5′-CAGCGGTAATTCCGAGCTGCA and the reverse primer 5′-CGATTTCCAGCGCACGTTTGTC with a 5′-biotin modification. The upstream and downstream DNA fragments were restricted with *Sty*I and *Dra*III respectively, and the 2.7 kb upstream and the 4.75 kb downstream DNA fragments were gel purified before use.

Scaffold Assembly and Elongation Complex (EC) Reconstitution. RNA and template DNA scaffold was assembled by annealing the template strand (100 nM) oligonucleotide 5′-GTGGTGTCGCTTGG GTTCTCTTTTCGCCTTGTCGGCTGCGCGTCGGTGGGTG TTTCCT-GATGGCTGTTTGTTTCCTATAGC and the 14mer RNA (100 nM) 5′-UUUUUACAG-CCAUC in reconstitution buffer (10 mM Tris HCl, pH 7.9, 40 mM KCl, 5 mM $MgCl_2$) using a thermal cycler, as described previously (11). ECs were reconstituted in elongation buffer (25 mM Hepes-KOH, pH 8.0, 130 mM KCl, 5 mM $MgCl_2$, 1 mM DTT, 0.15 mM EDTA, 5% glycerol and 25 μg of acetylated bovine serum albumin/mL) by incubating RNAPII (10 nM) with the scaffold (7.5 nM) containing template strand DNA and RNA at 30 °C for 10 min followed by the addition of nontemplate strand (10 nM) oligonucleotide 5′-CAAGGCTATAGGA-AACAAACAGCCATCAGGAAACA CCCACCGACGCGCAGCCGACAAGGCGAAAAGA-GAAC CCAAGCGACACCACGGG, and incubated for another 10 min at 30 °C. The template and nontemplate DNA oligonucleotides were designed to generate unique restriction sites for *Sty*I and *Dra*III at the upstream and downstream end of the scaffold respectively, upon annealing.

Transcript elongation by A20 ECs. After ligation of the DNA fragments to the labeled A20 ECs, the complexes were incubated with ATP, CTP, and GTP (100 μM each) for 5 min at 30 °C to form G76 ECs, which were then elongated in the presence of all 4 NTPs (1 mM each) for 30 min at 30 °C.

1. Nedialkov YA, et al. (2003) NTP-driven translocation by human RNA polymerase II. *J Biol Chem* 278:18303–18312.
2. Abbondanzieri EA, Greenleaf WJ, Shaevitz JW, Landick R, Block SM (2005) Direct observation of base-pair stepping by RNA polymerase. *Nature* 438:460–465.
3. Johnson RS, Strausbauch M, Cooper R, Register JK (2008) Rapid kinetic analysis of transcription elongation by *Escherichia coli* RNA polymerase. *J Mol Biol* 381:1106–1113.
4. Shaevitz JW, Abbondanzieri EA, Landick R, Block SM (2003) Backtracking by single RNA polymerase molecules observed at near-base-pair resolution. *Nature* 426:684–687.
5. Puig O, et al. (2001) The tandem affinity purification (TAP) method: A general procedure of protein complex purification. *Methods* 24:218–229.
6. Beckett D, Kovaleva E, Schatz PJ (1999) A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. *Protein Sci* 8:921–929.
7. van Werven FJ, Timmers HT (2006) The use of biotin tagging in *Saccharomyces cerevisiae* improves the sensitivity of chromatin immunoprecipitation. *Nucleic Acids Res* 34:e33.
8. Rose MD, Winston F, Hieter P (1990) *Methods in Yeast Genetics: A Laboratory Course Manual* (Cold Spring Harbor Laboratory Press, Plainview, NY).
9. Wang D, Bushnell DA, Westover KD, Kaplan CD, Kornberg RD (2006) Structural basis of transcription: Role of the trigger loop in substrate specificity and catalysis. *Cell* 127:941–954.
10. Kaplan CD, Larsson KM, Kornberg RD (2008) The RNA polymerase II trigger loop functions in substrate selection and is directly targeted by alpha-amanitin. *Mol Cell* 30:547–556.
11. Sidorenkov I, Komissarova N, Kashlev M (1998) Crucial role of the RNA:DNA hybrid in the processivity of transcription. *Mol Cell* 2:55–64.

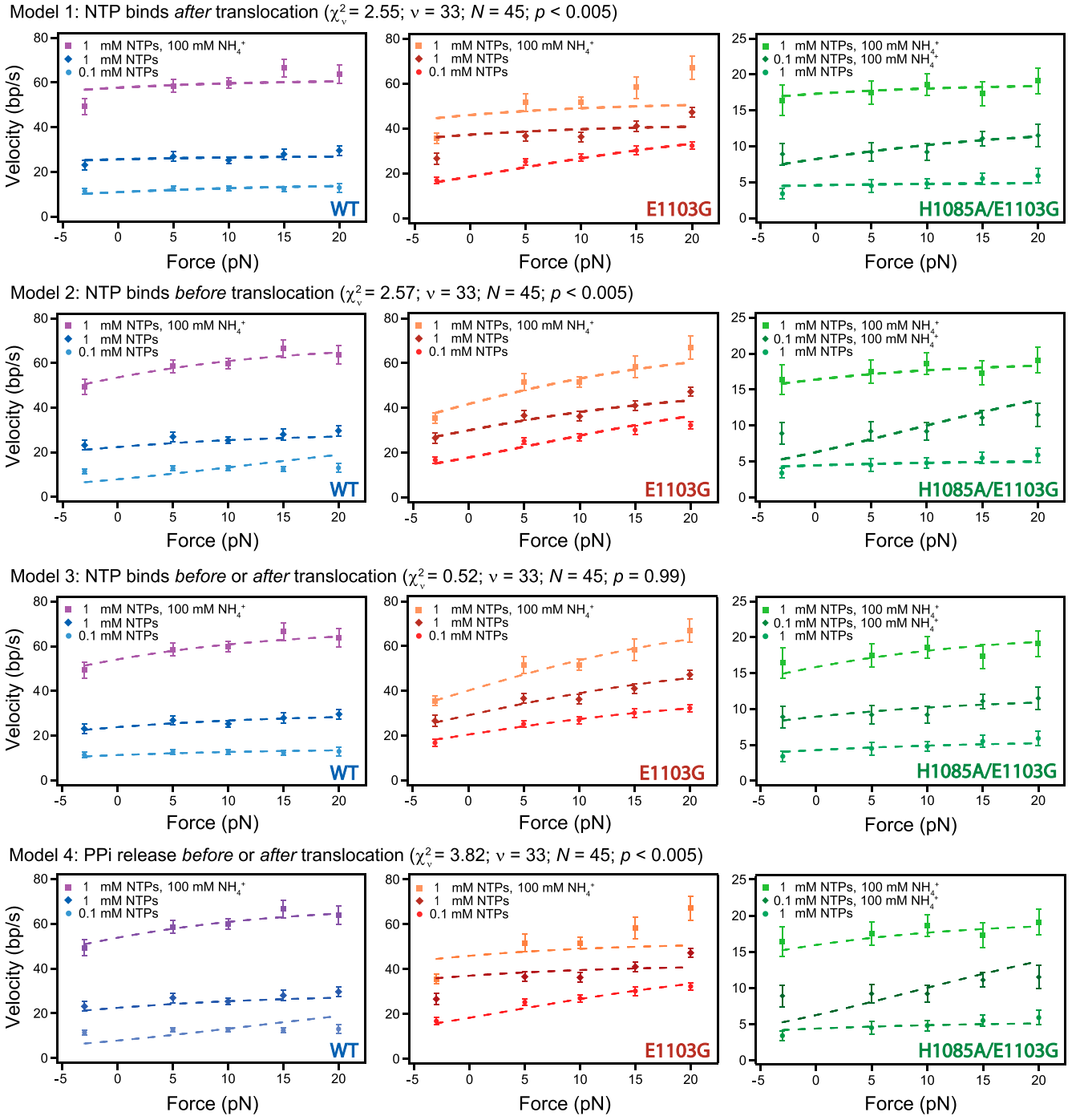

Fig. S1. Global fits to Brownian ratchet models of elongation. The single-molecule, pause-free elongation rates (error bars as SEM) measured as a function of external load for WT, E1103G, and H1085A/E1103G RNAPII at saturating (1 mM) and subsaturating (0.1 mM) NTPs, as well as in the presence of 100 mM ammonium, color-coded as indicated by the legends (*Inset*). The experimental datasets were fit globally to each the four Brownian ratchet reaction pathways shown in Fig. 2 (fits to Eqs. **S1**–**S4**, represented by dashed lines). There were 12 independent parameters fit for each model: specifically, the four kinetic parameters [k_{cat}, $k_{cat}(NH_4^+)$, K_δ, and K_D] for each of three RNAPII species (WT, E1103G, and H1085A/E1103G).

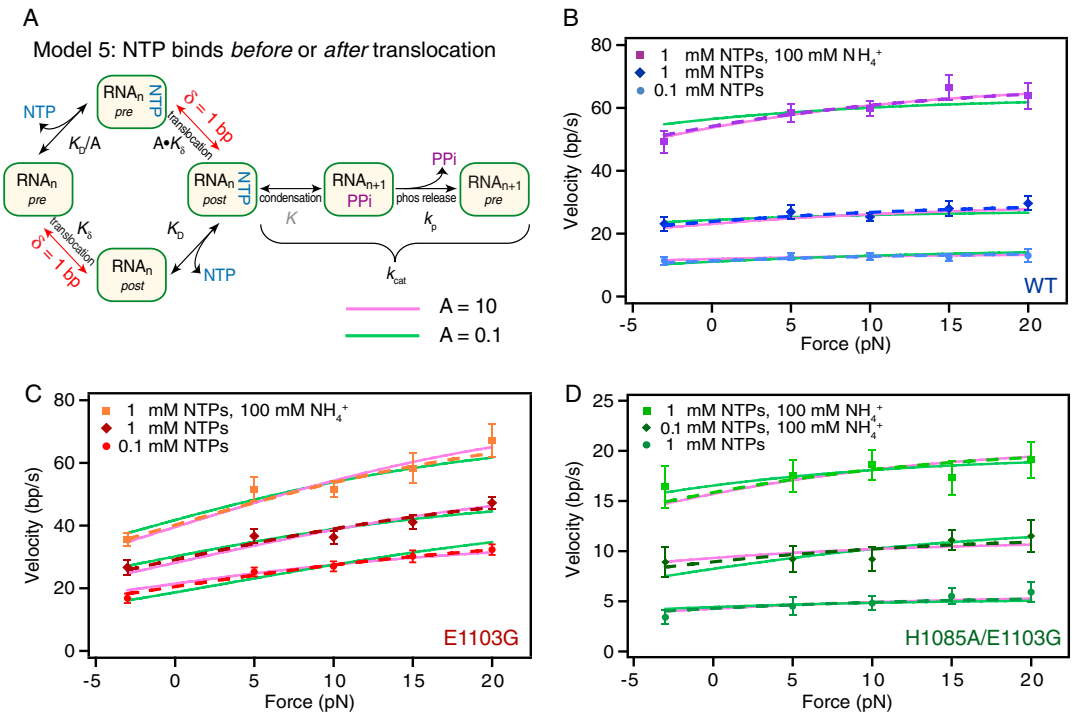

Fig. S2. Alternate fitting of the *F-v* data. (*A*) Model 5 is identical to Model 3, but with the addition of another free parameter, *A*, that allows for different rates of binding to the pre- and posttranslocated states. The *F-v* data were fit to Model 5 with fixed *A* values spanning an order of magnitude ($A = 10$, pink solid line; $A = 0.1$, green solid line). Dashed lines represent fits to Model 5 with $A = 1$, which causes this model to reduce to Model 3, depicted in Fig. 2*C*.

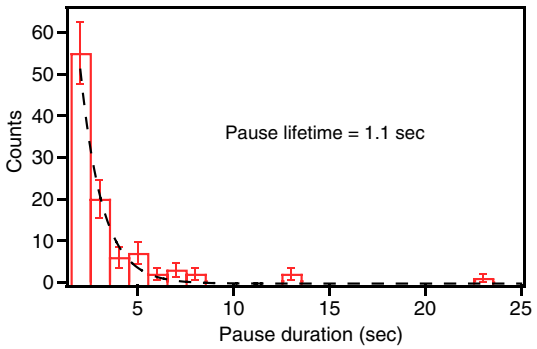

Fig. S3. Pause lifetime distribution for WT RNAPII. Histogram of the measured pause durations ($N = 98$, error bars as SD) from records of WT RNAPII transcribing at 10 pN assisting load and 1 mM NTPs. Dashed line represents a single-exponential fit to the data ($\tau = 1.1$ s).

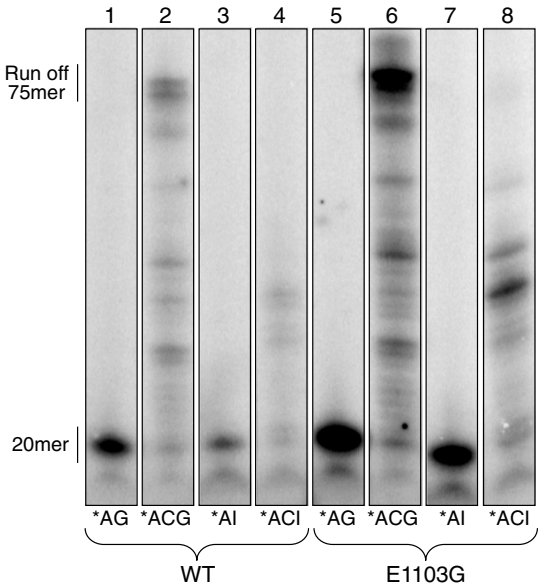

Fig. S4. ITP misincorporation assay. WT and E1103G ECs were incubated with four different NTP mixtures (see *SI Results and Discussion*), and the radiolabeled transcripts were separated on a 12.5% denaturing polyacrylamide gel. The positions of the A20 halt site and run-off product (75-mer) are indicated.