Supporting Information

Mejia et al. 10.1073/pnas.1421067112

SI Materials and Methods

Stall Complex Formation. The polymerases used in this study were purified from *E. coli* by the Nudler Laboratory at the New York University Medical Center (1). All three preparations, the two mutants and the WT enzyme, contain a biotin in the β′ subunit. Stalled elongation complexes are made by incubating the polymerase holoenzyme (core enzyme + σ initiation factor) with the template DNA that contains a λ_{PR} promoter. Directly upstream of the promoter, the DNA has a 70-bp sequence that lacks GMP (C-less cassette). The polymerase is walked from the promoter sequence to the stall site by further incubation of the open complexes with a low concentration of three of the four nucleotide triphosphates (ATP, GTP, and UTP at 150 μM; Fermentas). The addition of 2.5 μM of a ribo-dinucleotide (ApU; Dharmacon) complementary to the first two bases of the DNA template, facilitates polymerase initiation. The downstream end of the DNA used in this experiment is ligated to a digoxigenin handle. This handle is a 380-bp PCR product of a reaction supplemented with Dig-labeled deoxy-nucleotides (labeled dNTPs, Roche Diagnostics; nonlabeled dNTPs, Fermentas). The stalled elongation complexes are then bound to anti-Dig polystyrene beads (Spherotech) with a radius of 2.1 μm. These beads are suspended in transcription buffer containing 20 mM Tris, pH 7.9, 20 mM NaCl, 10 mM $MgCl_2$, and 20 mM DTT and adjusted to a pH of 7.9. The beads are then introduced into a glass chamber where the single molecule experiment will take place. Once in the chamber, a tether is made between one of the anti-Dig beads containing the complexes and another 2.1-μm streptavidin bead (Spherotech). The starting force on the tether was 3–4 pN. A buffer containing transcription buffer and a full set of nucleotide triphosphates at a saturating concentration of 1 mM (Fermentas) and 1 μM PPi (Fluka Biochemica) was used to reinitiate transcription. For experiments with nucleotide analogs, this same buffer was supplemented with 200 μM of either ITP (Sigma-Aldrich) or BrUTP (Sigma-Aldrich). In this passive mode experiment, the distance between the traps is kept constant so that when transcription restarts and the DNA between the beads shortens, the force on the polymerase increases. The initial contour length from the polymerase stall site to the other end of the DNA was 11,731 bases.

Data Analysis. Force data were converted into number of nucleotides transcribed using the extensible Worm-Like-Chain model of DNA elasticity (2) using a persistence length of 53 nm, a stretch modulus of 1,200 pN, and the initial DNA contour length. Data collected were filtered both with a Gaussian filter (1,200 ms) and a Savitzky-Golay filter (2.5 ms). Velocity was calculated by differentiating the number of bases transcribed. Pauses were removed by two methods that yielded equivalent results. (*i*) A histogram of all velocities consisted of two peaks, one around zero corresponding to pauses and the second one around the average velocity of the enzyme. A threshold of 2 or 3 SDs from the mean velocity was set to differentiate between elongation and pause. Velocities below the set threshold were set to be a pause, whereas everything above it was considered to be the pause-free velocity. (*ii*) Dwell times, which is the time spent at a particular position on the template, were calculated using a spatial bin of two nucleotides. The average dwell time was calculated, and a threshold of 2–3 SDs was set to differentiate the paused and the elongation states as done for the velocity threshold above. These methods are able to reliably remove pauses longer than 2 s. The pause-free velocity information and the number of pauses and their duration were saved for further analysis.

SI Text

Constructing a Kinetic Model. As mentioned above, pause density decreases as the average velocity increases, both in the case of the mutant polymerases and for the experiments done in the presence of nucleotide analogs (Fig. 2*A*). By further analyzing these data, we can perhaps better understand the observed behavior and help answer questions such as the following. Which rate have the point mutations affected? Has more than one rate been modified? Why is it that nucleotide analogs seem to affect the enzyme in a similar way than the mutations? What is the rate-limiting step in transcription? What causes the enzyme to pause? The construction of a kinetic scheme that reproduces the experimental data provides a quantitative approach to answering these questions.

Several previous studies have found that the elongating and pausing pathways are in kinetic competition with each other (3–5). For the simplest kinetic scheme (shown in Fig. 2*B*), at each base pair, the enzyme either moves one step forward or pauses. In this case, the probability of entering a pause is given by $P_p = k_p/(k_p + v)$, where k_p is the pause entry rate, and v is the pause-free elongation velocity as shown in the diagram. As we already mentioned, the mutant polymerases present different pause-free elongation velocities, indicating that the mutations affect a rate-limiting step in transcription. That is, even if transcription is composed of many different steps like translocation, catalysis, and PPi release, velocity is determined by the rate affected by the mutations, and which, according to our interpretation of the single molecule data, is the folding of the TL, k_F. If the mutations affect the TL folding rate k_F and k_F completely determines transcription velocity so that $k_F \sim v$, then the simple model $P_p = k_p/(k_p + v)$ should be able to fit the experimental data.

The best fit to this model using the measured pause density data is shown by the dashed line in Fig. 2*A,* and as can be seen, this expression is not able to reproduce the experimental results. One possible explanation is that the folding rate k_F does not completely determine the velocity of transcription ($v \neq k_F$), but that instead there is more than one rate-limiting step, that is, the elongation velocity is determined by more than one rate of comparable magnitudes.

Based on this realization, a second kinetic scheme (Fig. 2*C*) was constructed in which the elongation pathway is now divided in two steps. The first one is the folding of the TL, prompted by the binding of the next NTP and characterized by a second-order rate coefficient k_F, with units of micromolar per second ($\mu M^{-1} \cdot s^{-1}$), and that is affected by the mutations. This step is assumed to be reversible with a reverse rate k_{FB}. Once in the folded state (F), the irreversible hydrolysis of the nucleotide and release of PPi occurs with rate k_{inc}. As already mentioned, pausing (P) is an off-pathway mechanism that stems from the main elongation pathway. It is unclear, however, from which state pauses originate. A previous study (3) has shown that the number of pauses decreases with increasing NTP concentrations. Therefore, pausing can only stem from a state before NTP binding occurs, which corresponds in this case to the unfolded TL state (U) of the proposed kinetic scheme [it can be shown that if pausing where to stem from the folded state (F), higher NTP concentrations would increase the number of pauses, contrary to experimental observations].

Cleland (6) derived simple rules for the calculation of effective rates for a wide variety of kinetic schemes. Using these guidelines, it is possible to calculate expressions for velocity (effective forward rate), as well as the pausing probability, and then use them to fit the experimental data.

In this way, we obtain for the scheme in Fig. 2*C* that $v = 1/(1/k_{inc} + 1/kf_{eff})$ and $kf_{eff} = (k_F N k_{inc})/(k_{FB} + k_{inc})$. Then a calculation of the effective forward rate renders

$$v = \frac{k_F N k_{inc}}{k_F N + k_{inc} + k_{FB}}, \quad \textbf{[S1]}$$

where N is the nucleotide concentration.

Now the pausing probability is

$$P_p = \frac{kp}{kp + \frac{k_F N k_{inc}}{k_{FB} + k_{inc}}}. \quad \textbf{[S2]}$$

To write the pause density as a function of velocity, we can solve for k_F from Eq. **S1** and get

$$k_F = \frac{v(k_{inc} + k_{FB})}{N(k_{inc} - v)}. \quad \textbf{[S3]}$$

Substituting this expression back into Eq. **S2** we get

$$P_p = \frac{k_p}{k_p + \left(\frac{v k_{inc}}{-v + k_{inc}}\right)}. \quad \textbf{[S4]}$$

This final expression relates the measured pause-free velocity, V, with the enzyme's pause density. In fact, as shown in Fig. 2*A* (solid line), this model is able to reproduce the experimental data very well ($R^2 = 0.97$) with two fitting parameters and renders the values

$$k_p = 0.69 \pm 0.13\ \mathrm{s}^{-1}$$

$$k_{inc} = 25.3 \pm 4.5\ \mathrm{s}^{-1}.$$

This value of k_{inc} is very close to the one obtained by Abbondanzieri et al. (7) ($k_{inc} = 24 \pm 2\ \mathrm{s}^{-1}$) and by Bai et al. (8) ($k_{max} = 24.5 \pm 1.9\ \mathrm{s}^{-1}$) through very different methods. The k_p values are also in relatively good agreement with $k_p \approx 0.42\ \mathrm{s}^{-1}$ calculated from the temperature dependence of transcription by Mejia et al. (5).

It must be pointed out that in this previous calculation k_{inc} has been assumed to be constant; however, this assumption has not been strictly justified yet. As shown below, if k_F and k_{FB} are taken to be constant and k_{inc} is allowed to change, the calculated expression for pause density as a function of velocity is not able to fit the experimental measurements (Fig. S2, dotted line). Therefore, we can conclude that the observed changes in velocity and pause density are not determined by changes in k_{inc}.

In addition, note that the pausing probability P_p does not explicitly depend on k_F, the rate of TL folding, or on k_{FB}. One way to learn more about how k_F changes due to the point mutations is to calculate the changes in equilibrium energy for the k_F/k_{FB} transition compared with the WT enzyme, as given by $\Delta\Delta G_{x-wt}$, where x represents either the fast or slow mutant and *wt* refers to the WT enzyme.

We know that

$$\Delta G = -k_B T \ln \frac{k_F N}{k_{FB}}.$$

We define

$$\Delta\Delta G_{x-wt} \equiv \Delta G_x - \Delta G_{wt},$$

therefore

$$\Delta\Delta G_{x-wt} = -K_B T \ln k_{Fx} - K_B T \ln N + K_B T \ln k_{FBx} + K_B T \ln k_{Fwt} + K_B T \ln N - K_B T \ln K_{FBwt}.$$

For a fixed value of N, we can simplify to

$$\Delta\Delta G_{x-wt} = -K_B T \ln k_{Fx} + K_B T \ln k_{FBx} + K_B T \ln k_{Fwt} - K_B T \ln K_{FBwt}. \quad \textbf{[S5]}$$

In this model, the rate of TL folding k_F is the rate responsible for the observed changes in velocity, and the other rates, including k_{FB}, are considered constant, we demonstrate that changes in k_{FB}, while keeping all other rates constant, are not sufficient to generate the observed velocity changes. Furthermore, because, as it has already been shown, variations in k_{inc} cannot account for the observed changes in pause density as a function of velocity, this analysis leaves k_F as the rate most likely responsible for the velocity changes caused by the mutations or the nucleotide analogs). Therefore, taking $k_{FBx} = k_{FBwt}$, Eq. **S5** now becomes

$$\Delta\Delta G_{x-wt} = -K_B T \ln \frac{k_{Fx}}{k_{Fwt}}. \quad \textbf{[S6]}$$

By using Eq. **S3**, it is possible to write k_F in terms of velocity and therefore

$$\frac{k_{Fx}}{k_{Fwt}} = \frac{v_x(k_{inc} - v_{wt})}{v_{wt}(k_{inc} - v_x)}.$$

Eq. **S6** then becomes

$$\Delta\Delta G_{x-wt} = -K_B T \ln \frac{v_x(k_{inc} - v_{wt})}{v_{wt}(k_{inc} - v_x)}, \quad \textbf{[S7]}$$

where again, x represents either the fast or the slow mutant, and *wt* represents the WT enzyme. Using the measured velocities for the fast and the slow mutant and the k_{inc} value of 25 s^{-1} that was previously obtained from the fit in Fig. 2*A*, we get

$$\Delta\Delta G_{SlowM-WT} = +2.2\ \mathrm{pNnm} = +0.31\ \ \mathrm{kcal/mol}$$

$$\Delta\Delta G_{FastM-WT} = -5.1\ \mathrm{pNnm} = -0.73\ \ \mathrm{kcal/mol},$$

for the changes in equilibrium energy for the mutant's k_F/k_{FB} transition with respect to the corresponding transition in the WT enzyme. Significantly, this result is in excellent agreement with the estimates of the changes in equilibrium free energy for the mutant TL folding based on differences in TL helix propensity. From the helix propensity scale shown in Fig. S4, we obtain

$$\Delta\Delta G_{SlowM-WT} = +0.2\ kcal/mol$$

$$\Delta\Delta G_{FastM-WT} = -0.5\ kcal/mol$$

Thus, despite the simplicity of the helix propensity calculation, it is reassuring that it yields $\Delta\Delta G$ values similar to those obtained

from the direct comparison of mutant and WT rates. The diagram in Fig. 2*D* illustrates how the TL mutations are thought to change the transcription energy landscape. Note that for the fast mutant, k_F is faster, whereas for the slow mutant, k_F is slower compared with the WT enzyme. All other rates are unchanged.

With the purpose of obtaining an estimate of k_F and k_{FB}, the expression for velocity shown in Eq. **S1** can be further analyzed. After some rearrangement

$$v = \frac{k_F N k_{inc}}{k_F N + k_{inc} + k_{FB}} = \frac{k_{inc}}{1 + \frac{k_{inc}}{k_F N} + \frac{k_{FB}}{k_F N}}, \qquad \textbf{[S8]}$$

where, as we described, for our experimental data, all rates are constant except k_F, which varies.

Because $\Delta G_F \leq 0$ is required for the reaction to go forward, we will impose the condition $k_{Fslow}/k_{FB} = 1$, that is, $\Delta G_{Fslow} = 0$ for the slowest reaction (slow mutant) and so $\Delta G_F < 0$ for all other reactions.

Starting from Eq. **S5**, we have

$$\Delta\Delta G_{slow-x} = -K_B T ln k_{Fslow} + K_B T ln k_{FBslow} + K_B T ln k_{Fx} - K_B T ln k_{FBx},$$

where we now placed the slow mutant as reference and x represents the fast and WT enzymes. After some rearranging we have

$$\Delta\Delta G_{slow-x} = +K_B T ln \frac{k_{FB}}{k_{Fslow}} - K_B T ln \frac{k_{FB}}{k_{Fx}}.$$

Here we used the fact that k_{FB} is constant. Applying the condition that $k_{Fslow}/k_{FB} = k_{FB}/k_{Fslow} = 1$

$$\Delta\Delta G_{slow-x} = 0 - K_B T ln \frac{k_{FB}}{k_{Fx}}.$$

Inverting the natural logarithm, we have

$$\frac{k_{FB}}{k_{Fx}} = e^{\frac{\Delta\Delta G_{slow-x}}{K_B T}} \equiv A_x. \qquad \textbf{[S9]}$$

Substituting back in V (Eq. **S8**)

$$v = \frac{k_{inc}}{1 + \frac{1}{N}\left(\frac{k_{inc}}{k_{Fx}} + A_x\right)}, \qquad \textbf{[S10]}$$

A_x can be calculated from Eq. **S9** and using Eq. **S7,** resulting in

$$v = \frac{k_{inc}}{1 + \frac{1}{N}\left[\frac{k_{inc}}{k_{Fx}} + \frac{v_{slow}(k_{inc} - v_x)}{v_x(k_{inc} - v_{slow})}\right]}. \qquad \textbf{[S11]}$$

Because in this expression, k_{inc} is a known quantity and N is the nucleotide concentration, the only element that remains unknown is k_{Fx}, which can then be used as a fitting parameter. Fig. S3 shows some of the published data for the nucleotide dependence of transcription velocity for the WT enzyme [blue triangles from Abbondanzieri et al. (7); pink diamonds from Forde et al. (3); green triangles from Davenport et al. (9); and red circle from this study]. As the plot shows, this nucleotide dependence data can be well fit by the expression for velocity derived in Eq. **S11** with only one fitting parameter k_{Fwt} with values of 0.15 (blue curve), 0.05 (pink curve), 0.04 (black curve), and 0.025 s^{-1} (red dashed curve). We believe that this variability is most likely due to differences between experimental conditions, including but not limited to the template sequence used (10–12), the data processing methodology, and the pause removal algorithms.

Even if the experimental data do not allow the exact determination of the value of k_F for the WT, this analysis does at least provide bounds for k_{Fwt} of 0.15–0.025 s^{-1}. From Eq. **S7**, we see that $k_{Fwt} = k_{FB}/A_{wt}$ and therefore these boundaries correspond to values for k_{FB} that range from 0.09 to 0.015 s^{-1} (remember that k_{FB} is assumed to be constant for all data sets; only k_F changes). Table S2 shows the values of $\Delta\Delta G_{x-wt}$, $\Delta\Delta G_{slow-x}$, and k_F for all datasets, using the known value of $k_{inc} = 25\ s^{-1}$ and choosing $k_{FB} = 0.09\ s^{-1}$ because it is the value that best fits our experimental data (Fig. S3).

It is interesting to note that under saturating nucleotide concentrations ($N = 1{,}000/\mu M$), $k_F N$ is of the same order of magnitude as k_{inc}; thus, under these conditions, two rate-limiting steps determine the transcription velocity: the TL folding rate ($k_F N$) and nucleotide hydrolysis/PPi release (k_{inc}). Having two rate-limiting steps means that modifying either of them will result in a change in the pause-free velocity. Fig. 3*B* (red line) illustrates how the pause-free transcription velocity as given by Eq. **S1** varies as a function of k_F for the known values $k_{inc} = 25\ s^{-1}$, $k_{FB} = 0.09\ s^{-1}$, and $N = 1{,}000\ \mu M^{-1} \cdot s^{-1}$. Fig. 3*B* also shows the experimental velocity data and the corresponding values of k_F calculated using $k_F = [v(k_{inc} + k_F B)]/[N(k_{inc} - v)]$ (blue circles). Note that even for a constant k_{inc} and nucleotide concentration, changes to k_F caused by the mutations or the nucleotide analogs will have an effect on the pause-free transcription velocity. This model also accurately predicts the velocity saturation for high values of k_F as seen in the data, wherein k_{inc} then becomes the rate-limiting step.

Kinetic Scheme Comparison and Assumptions. As described in the main text, for the simplest kinetic scheme in which elongation and pausing compete (3–5), the probability of entering a pause is given by $P_p = k_p/(k_p + v)$, where k_p is the pause entry rate, and v is the pause-free elongation velocity. This model assumes that transcription velocity is completely determined by the TL folding rate k_F so that $v = k_F N$. However, we showed that this model is not able to reproduce the experimental data, suggesting that elongation velocity is not only determined by k_F but rather by a combination of more than one rate-limiting step or other reversible transitions.

As derived previously, the pausing probability in terms of the pause-free velocity, v, is

$$P_p = \frac{k_p}{k_p + \left(\frac{v k_{inc}}{-v + k_{inc}}\right)}. \qquad \textbf{[S12]}$$

As a comparison, note that this expression reduces to the pausing probability of the first kinetic scheme when k_{inc} is no longer rate limiting, i.e., in the limit of very large k_{inc}.

As part of this model, we assumed that only one rate is affected by the mutations, the TL folding rate k_F, whereas k_{inc} and k_{FB} are assumed constant. However, as shown below, the velocity and pausing probabilities were also calculated for the case in which either k_{inc} or k_{FB} is allowed to change while k_F is kept constant. This analysis demonstrated that these other models are unable to replicate the experimental data.

Another assumption of this model, as seen in the second scheme of Fig. 2, is that the binding of the nucleotide and the folding of the TL are simultaneous events. As a result, we thus assume that the duplex stability of an analog stems in part from the specific contacts made by the analog with the TL. This stability,

reflected in its binding rate, is in turn manifested in the value of the rate constant $k_F N$.

Supplementary Analysis on Kinetic Scheme. The kinetic scheme shown in the previous section is able to reproduce the experimental data; however, for some of the analyses, we assumed that k_F is the rate that changes, whereas all other rates are kept constant. Strictly speaking, we did not prove this idea. Now we will analyze the case in which one of the other rates involved in the kinetic scheme changes. For simplicity, our model will assume that only one rate changes at a time.

Assuming k_{inc} changes. First we will analyze the case where k_{inc} is the rate that varies. For this scenario, instead of solving for k_F from the velocity expression (Eq. **S1**), we now solve for k_{inc}

$$k_{inc} = \frac{v(k_F N + k_{FB})}{-v + k_F N},$$

and substitute it into the pause density (Eq. **S2**)

$$Pp = \frac{(k_{FB} + v)kp}{kpk_{FB} + kpv + vk_F N + vk_{FB}}.$$

After some rearrangement

$$Pp = \frac{kp}{kp + \frac{v(k_F N + k_{FB})}{k_{FB} + v}}.$$

This expression can now be used to fit the experimental data taking kp, k_F, and k_{FB} as the three fitting parameters. As shown by the dotted line in Fig. S2 (dashed line), this model is not able to fit the data. Therefore, we can conclude that the observed changes in velocity and pause density are not determined by changes in k_{inc}.

Assuming k_{FB} changes. Next, we will consider the case in which k_{FB} is the rate that changes, whereas the others are constant. Again, we solve for k_{FB} from Eq. **S1** and substitute into Eq. **S2**

$$k_{FB} = -\frac{vk_F N + vk_{inc} - k_F N k_{inc}}{v}$$

$$Pp = \frac{kp(v - k_{inc})}{kpv - kpk_{inc} - vk_{inc}}.$$

After some rearranging we get

$$Pp = \frac{kp}{kp - \frac{vk_{inc}}{v - k_{inc}}},$$

which is exactly the same expression we got in Eq. **S12** when we solved for k_F.

In addition, taking the expression for pause-free velocity in Eq. **S1**, but now leaving it in terms of k_{FB} instead of k_F as we had done before, we get

$$v = \frac{k_F N k_{inc}}{k_F N + k_{inc} + k_{FB}} = \frac{\frac{k_F}{k_{FB}} N k_{inc}}{\frac{k_F}{k_{FB}} N + \frac{k_{inc}}{k_{FB}} + 1}$$

$$v = \frac{\frac{N}{A_x} k_{inc}}{\frac{N}{A_x} + \frac{k_{inc}}{k_{FB}} + 1},$$

where in this last expression we used the previous definition that $A_x = k_{FB}/k_F$.

Using $k_{inc} = 25\ \text{s}^{-1}$ and A_{wt} that we calculated before, we can use this expression to fit the nucleotide dependence data for the WT enzyme. The same data of Fig. 3*A* can be now fit with this last expression of velocity with fitting parameter k_{FB}. For the Abbondanzieri et al. data, the fit renders $k_{FB} = 0.102\ (0.087, 0.1178)\ \text{s}^{-1}$, which in turn gives a value of $k_F = 0.166\ \text{s}^{-1}$, essentially the same value previously obtained from the fit to the Abbondanzieri et al. data as shown in Fig. 3*A*. The same can be shown for the data of Forde et al. and Davenport et al. This analysis demonstrates that the same results are obtained if either k_F or k_{FB} is used as the fitting parameter, serving as a good consistency check but not providing information about which of these two rates the mutations and nucleotide analogs are affecting.

We can also calculate the change in equilibrium energy $\Delta\Delta G_{x-wt}$ assuming that k_{FB} is the rate that changes. For the slow mutant

$$\Delta\Delta G_{s-wt} = -K_B T ln k_{Fs} + K_B T ln k_{FBs} + K_B T ln k_{Fwt} - K_B T ln k_{FBwt}.$$

If

$$k_{Fs} = k_{Fwt},$$

then

$$\Delta\Delta G_{s-wt} = -K_B T ln \frac{k_{FBwt}}{k_{FBs}}.$$

Using the expression for k_{FB} in terms of the velocity V

$$\frac{k_{FBwt}}{k_{FBs}} = \frac{v_s(v_{wt}k_F N + v_{wt}k_{inc} - k_F N k_{inc})}{v_{wt}t(v_s k_F N + v_s k_{inc} - k_F N k_{inc})}. \qquad \textbf{[S13]}$$

Using $k_{inc} = 25\ \text{s}^{-1}$ and $k_F = 0.166\ \mu\text{M}^{-1} \cdot \text{s}^{-1}$ that we got before, we get for the slow mutant

$$\Delta\Delta G_{s-wt} = +2.44\ \text{pNnm},$$

Which is very close to the value we got when substituting k_F. However, if we use $k_F = 0.025\ \mu\text{M}^{-1} \cdot \text{s}^{-1}$, which was the value that best fit our data, we now get

$$\Delta\Delta G_{s-wt} = +8.14\ \text{pNnm}.$$

If we plot k_{FBwt}/k_{FBs} as given by Eq. **S13** as a function of k_F, we see that it is a nonmonotonic function with a maximum at $k_F = 0.01$. In the case of the fast mutant, this instability is found near $k_F = 0.08$. In fact, there is no single value of k_F that gives a value for k_{FBwt}/k_{FBx} that is positive for the slow and the fast mutant. This analysis suggests that changing k_{FB} is perhaps not correct.

In addition, we can calculate the changes in velocity according to Eq. **S1** as a function of k_{FB} for $k_{inc} = 25\ \text{s}^{-1}$, $k_F = 0.025\ \text{s}^{-1}$, and $N = 1000\ \mu\text{M}^{-1} \cdot \text{s}^{-1}$. As Table S2 shows, because k_{FB} is in the denominator of Eq. **S1** and it is small compared with k_{inc} and $k_F N$, changes in k_{FB} do not result in significant changes to velocity.

Therefore, we conclude that k_F has to be the rate that changes and not k_{FB}.

Mechanical Modeling. As pointed out by Svetlov and Nudler (13), additional factors, like entropy, solvation, and internal strain, are also important for the correct determination of the energetic changes that occur after amino acid substitutions. These factors are all taken into account when using molecular mechanics algorithms. One such specialized algorithm that models structural changes and calculates changes in the folding free energy of mutants (CC/PBSA) (14) predicted changes in stability that were in good qualitative agreement with calculations using published helix propensity data.

Connecting Helix Equilibrium Energies with Folding Kinetics. The changes in stability of the folded TL presented here are calculated as changes in the ΔG or equilibrium energy of the transition. We propose that those energies are then related to the folding kinetics of the TL by assuming that only one of the rates, the folding rate, k_F, is modified by the mutations. The unfolding rate, k_{FB}, is not assumed to change (our data are not consistent with k_{FB} being the rate that changes). Thus, the change in free energy, $\Delta G = k_F/k_{FB}$ scales directly with the rate of folding of the loop. Therefore, we are able to limit the possible scenarios.

k_F—the forward rate—is given by $k_F = k_{F0}e^{-\Delta G^{\dagger}/K_bT}$, and k_{FB}—the backward rate—is given similarly by $k_{FB} = k_{FB0}e^{-\Delta G'^{\dagger}/K_bT}$, where $G^{\dagger}$ and $G'^{\dagger}$ are the energy of the respective transition states (Fig. S6*A*). To comply with the conditions that both ΔG and k_F change in the presence of the amino acid substitutions, but that k_{FB} is kept constant, there are only two options: (*i*) $\Delta\Delta G$ reflect changes in the stability of the unfolded state and involve no modifications of either the energy of the transition state $(\Delta G^{\dagger})$ or the energy of the folded state as shown in Fig. S6*B* or (*ii*) there is a simultaneous and identical shift of the energy of the folded and transition states relative to a fixed free energy value of the unfolded state as indicated in Fig. S6*C*. Although our experiments do not allow us to precisely determine which of these two possibilities actually takes place, we believe that the simultaneous and identical energy shift of the transition state and the folded state is rather unlikely (scenario 2), and therefore a change in the energy of the unfolded state is most likely (scenario 1).

1. Nudler E, Gusarov I, Bar-Nahum G (2003) *Methods of Walking With the RNA Polymerase*, eds Adhya S, Garges S (Academic, New York), pp 160–169.
2. Bustamante C, Marko JF, Siggia ED, Smith S (1994) Entropic elasticity of lambda-phage DNA. *Science* 265(5178):1599–1600.
3. Forde NR, Izhaky D, Woodcock GR, Wuite GJ, Bustamante C (2002) Using mechanical force to probe the mechanism of pausing and arrest during continuous elongation by Escherichia coli RNA polymerase. *Proc Natl Acad Sci USA* 99(18):11682–11687.
4. Herbert KM, et al. (2006) Sequence-resolved detection of pausing by single RNA polymerase molecules. *Cell* 125(6):1083–1094.
5. Mejia YX, Mao H, Forde NR, Bustamante C (2008) Thermal probing of E. coli RNA polymerase off-pathway mechanisms. *J Mol Biol* 382(3):628–637.
6. Cleland WW (1975) Partition analysis and the concept of net rate constants as tools in enzyme kinetics. *Biochemistry* 14(14):3220–3224.
7. Abbondanzieri EA, Shaevitz JW, Block SM (2005) Picocalorimetry of transcription by RNA polymerase. *Biophys J* 89(6):L61–L63.
8. Bai L, Shundrovsky A, Wang MD (2004) Sequence-dependent kinetic model for transcription elongation by RNA polymerase. *J Mol Biol* 344(2):335–349.
9. Davenport RJ, Wuite GJ, Landick R, Bustamante C (2000) Single-molecule study of transcriptional pausing and arrest by E. coli RNA polymerase. *Science* 287(5462): 2497–2500.
10. Bai L, Fulbright RM, Wang MD (2007) Mechanochemical kinetics of transcription elongation. *Phys Rev Lett* 98(6):068103.
11. Holmes SF, Erie DA (2003) Downstream DNA sequence effects on transcription elongation. Allosteric binding of nucleoside triphosphates facilitates translocation via a ratchet motion. *J Biol Chem* 278(37):35597–35608.
12. Rhodes G, Chamberlin MJ (1974) Ribonucleic acid chain elongation by Escherichia coli ribonucleic acid polymerase. I. Isolation of ternary complexes and the kinetics of elongation. *J Biol Chem* 249(20):6675–6683.
13. Svetlov V, Nudler E (2009) Macromolecular micromovements: How RNA polymerase translocates. *Curr Opin Struct Biol* 19(6):701–707.
14. Benedix A, Becker CM, de Groot BL, Caflisch A, Böckmann RA (2009) Predicting free energy changes using structural ensembles. *Nat Methods* 6(1):3–4.

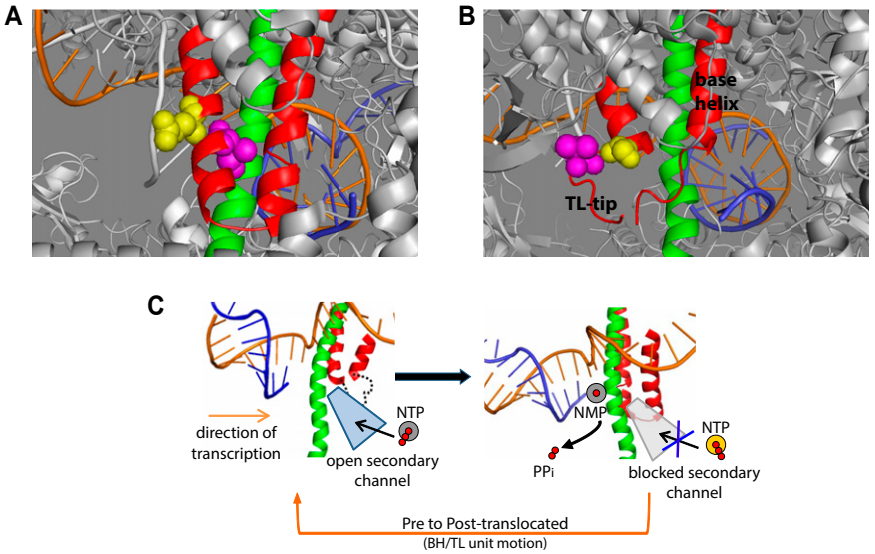

Fig. S1. Structural dynamics of the TL. (*A*) Crystal structure for the *T. Thermophilus* RNAP (1) (PDB ID code 2O5J) in the presence of the correct NTP shows a completely folded helical structure for the TL (red) and close contacts with the F-bridge (green). The mutated residue is shown for the slow mutant $(I \rightarrow V)$ in magenta and in yellow for the fast mutant $(G \rightarrow S)$. Mapping of the *E. coli* mutations was done based on the sequence alignment published by Tan et al. (2) and using the PyMOL molecular viewer software. Template DNA is shown in orange, RNA in blue. (*B*) Structure for a backtracked yeast Pol II complex (3) (PDB ID code 3GTJ) shows a partially folded TL. The remaining helical structures of the unfolded TL are known as the base helices. Mutated residues are located near the start of the unstructured amino acids on the TL-tip. Color code is the same as above. (*C*) When the TL is unfolded, the polymerase is in an open conformation allowing the next nucleotide to enter through the secondary channel. Once the TL folds on NTP binding, it blocks the secondary channel, serving as a pawl that rectifies the Brownian oscillations of the polymerase by not allowing backward motion. After hydrolysis is complete, the resulting PPi is released, prompting the TL to unfold and the polymerase to adopt the open conformation. Further conformations take the polymerase from the pre- to post-translocated state resetting the system for another cycle.

1. Vassylyev DG, Vassylyeva MN, Perederina A, Tahirov TH, Artsimovitch I (2007) Structural basis for transcription elongation by bacterial RNA polymerase. *Nature* 448(7150):157–162.
2. Tan L, Wiesler S, Trzaska D, Carney HC, Weinzierl RO (2008) Bridge helix and trigger loop perturbations generate superactive RNA polymerases. *J Biol* 7(10):40.
3. Wang D, et al. (2009) Structural basis of transcription: Backtracked RNA polymerase II at 3.4 angstrom resolution. *Science* 324(5931):1203–1206.

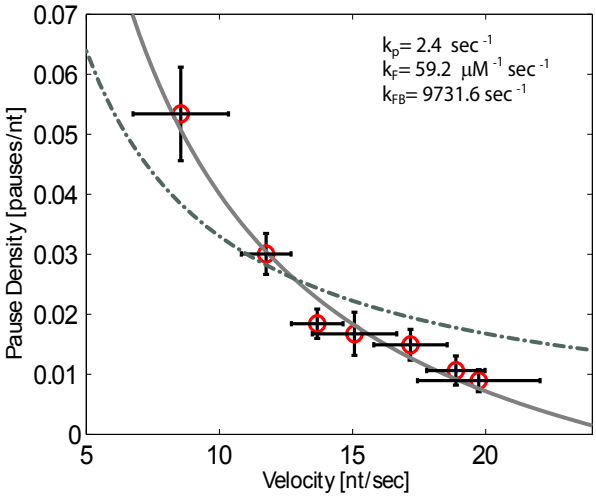

Fig. S2. k_{inc} as the affected rate. Fit to experimental pause density as a function of pause-free velocity with the expression derived assuming that k_p, k_F, and k_{FB} are constant and k_{inc} is the rate that varies (dotted gray line). As seen here, this model is not able to fit the data and therefore we can conclude that the observed changes in velocity and pause density are not determined by changes in k_{inc}.

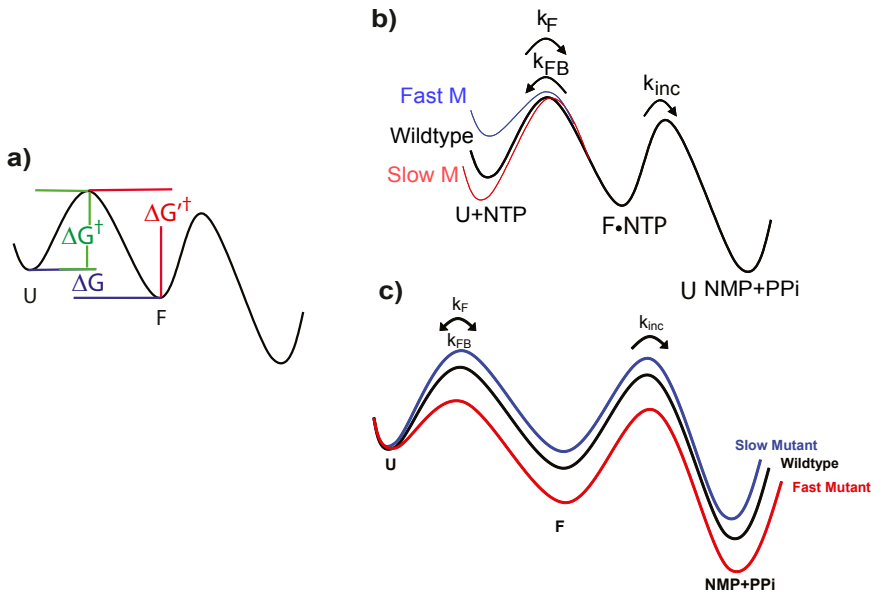

Fig. S3. Comparison of TL transition energetics. (*A*) Schematic of the transition between unfolded (U) and folded (F) TL states. ΔG is the equilibrium energy of the transition, whereas $\Delta G^{\dagger}$ is the energy of the transition state. (*B*) Scenario 1, in which changes in the equilibrium energy and forward rate k_F reflect changes in the energy of the unfolded state, with no change to the energy of the transition state or to the energy of the folded state. (*C*) Scenario 2, in which changes in equilibrium energy and forward rate k_F are caused by a simultaneous and identical shift of both the energy of the transition state and energy of the folded state. As discussed in the *SI Text*, we believe scenario 1 is more likely.

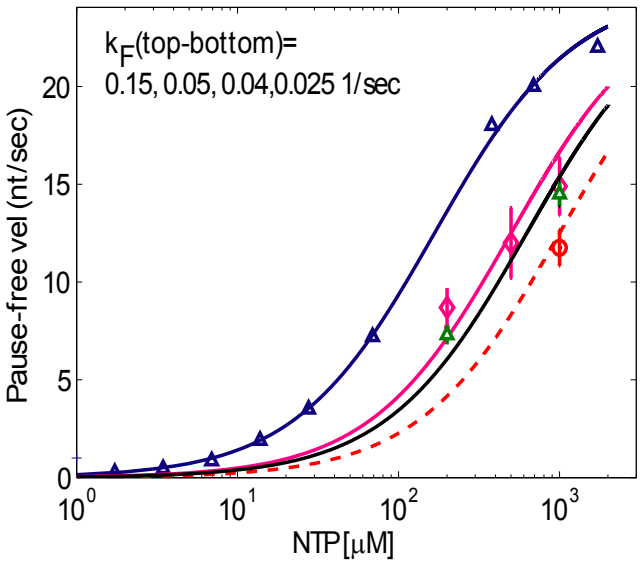

Fig. S4. Fit to nucleotide-dependent velocity data. Elongation velocity as a function of nucleotide concentration for the WT enzyme from several published studies [blue triangles from Abbondanzieri et al. (1), pink diamonds from Forde et al. (2), green triangles from Davenport et al. (3), and red circle from this study]. The data can be well fit by the model in Fig. 2*C* with values for k_{Fwt} of 0.15 (blue curve), 0.05 (pink curve), 0.04 (black curve), and 0.025 s^{-1} (red dashed curve). This variability is thought to arise from differences in data analysis or sequence-dependent effects.

1. Abbondanzieri EA, Shaevitz JW, Block SM (2005) Picocalorimetry of transcription by RNA polymerase. *Biophys J* 89(6):L61–L63.
2. Forde NR, Izhaky D, Woodcock GR, Wuite GJ, Bustamante C (2002) Using mechanical force to probe the mechanism of pausing and arrest during continuous elongation by Escherichia coli RNA polymerase. *Proc Natl Acad Sci USA* 99(18):11682–11687.
3. Davenport RJ, Wuite GJ, Landick R, Bustamante C (2000) Single-molecule study of transcriptional pausing and arrest by E. coli RNA polymerase. *Science* 287(5462):2497–2500.

Amino acid	Helix propensity Energy penalty (kcal/mol)
Ala	0.00
Glu ^{0}	0.16
Leu	0.21
Arg $^{+}$	0.21
Met	0.24
Lys $^{+}$	0.26
Gln	0.39
Glu $^{-}$	0.40
Ile	0.41
Asp ^{0}	0.43
Trp	0.49
Ser	0.50
Tyr	0.53
Phe	0.54
His ^{0}	0.56
Val	0.61
Asn	0.65
Thr	0.66
His $^{+}$	0.66
Cys	0.68
Asp $^{-}$	0.69
Gly	1.00
Pro	3.16

Slow Mutant

Fast Mutant

Fig. S5. Helix propensity of mutants. Helix propensity scale constructed using the values calculated by Pace and Scholtz (1). This table shows the energetic penalty contributed by each amino acid during helix formation. For example, glycine's increased flexibility gives this amino acid a high penalty when forming a helix. Therefore, changing a glycine to serine in the fast mutant increases the TL helix propensity (the energetic penalty for forming a helix decreases from 1 to 0.5 kcal/mol). On the other hand, the slow mutant substitution of isoleucine to valine decreases the TL helix propensity (the penalty increases from 0.41 to 0.61 kcal/mol). Even though there is some variability among helix propensity scales, they all place glycine as the amino acid with one of the lowest helix potentials and valine with a lower helix propensity than isoleucine. Therefore, we do in fact find a correlation between the helix forming propensity of the TL and the rate of catalysis as the crystal structures suggested.

1. Pace CN, Scholtz JM (1998) A helix propensity scale based on experimental studies of peptides and proteins. *Biophys J* 75(1):422–427.

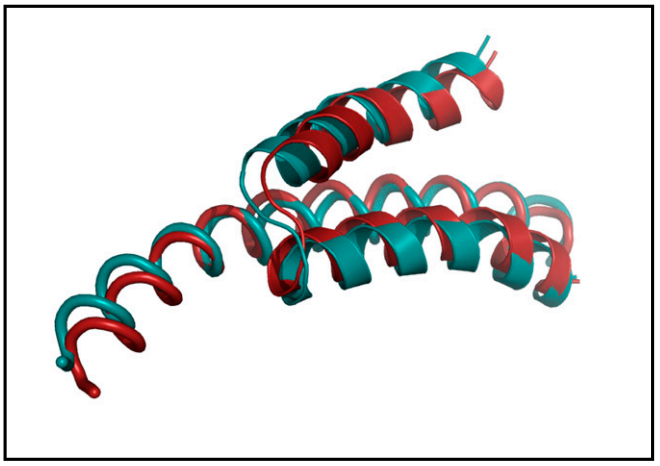

Fig. S6. Predicted structures of mutant TLs. The predicted lowest energy structures of the fast and slow mutants using the algorithm developed by Benedix et al. (1) appear to be nearly superimposable. It can also be shown that the positions of the conserved *E. coli* residues that contact the substrate for the WT and mutant enzymes are also superimposable with the solved *T. thermophilus* structure. Therefore, these two mutations are not expected to cause significant geometric alterations of the TL element at the catalytic center or to change the contacts between the TL and the NTP or other parts of the enzyme.

1. Benedix A, Becker CM, de Groot BL, Caflisch A, Böckmann RA (2009) Predicting free energy changes using structural ensembles. *Nat Methods* 6(1):3–4.

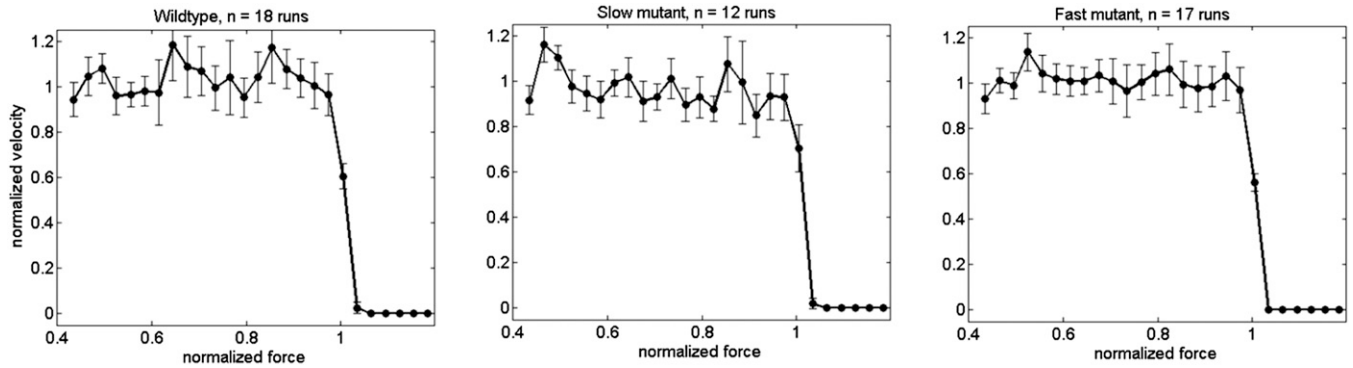

Fig. S7. Force-velocity curves. The normalized force-velocity curves for the WT and mutant polymerases show an independence of velocity to force, until the velocity suddenly drops to zero (see ref. 1 for discussion of the reason for this sudden stop). A force-velocity independence indicates that the translocation step of the enzyme is not the rate-limiting step of the cycle. The identification of both of the two rate limiting steps for elongation established in the present study can explain the force independence of velocity that has been observed for both mutants and the WT polymerases in this study and in earlier single molecule experiments (1–4). Because neither of the rate-determining processes identified here coincides with translocation and there is no a priori reason for the internal conformational changes involved in these processes to be force sensitive, it is to be expected that the measured velocity should be force independent.

1. Mejia YX, Mao H, Forde NR, Bustamante C (2008) Thermal probing of E. coli RNA polymerase off-pathway mechanisms. *J Mol Biol* 382(3):628–637.
2. Forde NR, Izhaky D, Woodcock GR, Wuite GJ, Bustamante C (2002) Using mechanical force to probe the mechanism of pausing and arrest during continuous elongation by Escherichia coli RNA polymerase. *Proc Natl Acad Sci USA* 99(18):11682–11687.
3. Wang HY, Elston T, Mogilner A, Oster G (1998) Force generation in RNA polymerase. *Biophys J* 74(3):1186–1202.
4. Galburt EA, et al. (2007) Backtracking determines the force sensitivity of RNAP II in a factor-dependent manner. *Nature* 446(7137):820–823.

Table S1. Changes in velocity as a function of k_{FB} using values of $k_{inc} = 25\ s^{-1}$ and $k_F = 0.025\ s^{-1}$

RNAP + nucleotide analog	$k_{FB}(s^{-1}) = k_F A$	Velocity (Eq. **S1**) (nt/sec)
Slow M	0.025	12.494
WT	0.015	12.496
WT + ITP	0.011	12.497
Fast M + ITP	0.009	12.498
WT + BrUTP	0.006	12.499
Fast M	0.004	12.499
Fast M + BrUTP	0.0035	12.499

As can be seen, k_{FB} values do not result in significant changes in velocity, proving that that k_F has to be the rate that changes and not k_{FB}.

Table S2. Calculated kinetic values for the proposed model using $k_{inc} = 25\ s^{-1}$ and $k_{FB} = 0.09\ s^{-1}$

RNAP + nucleotide analog	$\Delta\Delta G_{x-WT}$ (pNnm)	$\Delta\Delta G_{slow-x}$ (pNnm)	$k_F = \frac{v(k_{inc}+k_{FB})}{N(k_{inc}-v)}$ ($\mu M^{-1} \cdot s^{-1}$)
Slow M	+2.20	0	0.09
WT	0	+2.20	0.15
WT + ITP	−1.26	+3.45	0.21
Fast M + ITP	−2.20	+4.39	0.26
WT + BrUTP	−3.71	+5.09	0.39
Fast M	−5.11	+7.30	0.56
Fast M + BrUTP	−5.91	+8.10	0.69