Supporting Information

Lisica et al. 10.1073/pnas.1517011113

SI Text

Stochastic Model of Backtrack Recovery. We describe the stochastic motion of the backtracked polymerase as a 1D continuous time random walk between different states $i \in [1,2,\ldots,n-1,n,n+1,\ldots,\infty)$. Each state represents the backtracked distance in nucleotides, with zero being the elongation-competent state. Because the backtrack recovery was experimentally probed at low forces (with a mean value of 1.9 pN), we assume that no external forces are exerted on the polymerase; therefore, its stochastic motion can be described by an unbiased diffusion process. The polymerase is represented by a Brownian particle that jumps between adjacent states with rate k and uses cleavage to return to zero with rate k_c. For Pol I and Pol II, the cleavage rate is not constant but depends on the backtrack depth (Fig. 2*B*), λ being the cleavage cutoff at which k_c drops to zero. The statistics of backtrack recovery for a finite value of λ can only be obtained numerically.

Because in the case of Pol II–TFIIS, we do not observe a threshold depth in backtrack recovery (Fig. 3*A*), we consider a constant cleavage rate and $\lambda \to \infty$. In the case of Pol I A12.2Δ C and Pol II Δ Rpb9, there is no cleavage reaction, and the backtrack recovery can be described by a 1D diffusion process ($k_c = 0$). Therefore, we derive exact analytical expressions for the statistics of backtrack recovery for both Pol II–TFIIS and the mutants as shown below.

For simplicity and because of the fast elongation rate from $i=0$ [~10 times faster than the backtrack diffusion rate (10)], we neglect the backward diffusion rate from zero to one in our model. Therefore, we consider that polymerases that reach the elongation-competent state ($i=0$) will most likely elongate. With $p_i(t)$, we denote the probability of the particle to be at state i at time $t \geq 0$. The dynamics of the system can be described by the following set of master equations:

$$\frac{dp_1}{dt} = kp_2 - (2k+k_c)p_1, \qquad \textbf{[S1]}$$

$$\frac{dp_2}{dt} = kp_3 - (2k+k_c)p_2 + kp_1, \qquad \textbf{[S2]}$$

and

$$\frac{dp_i}{dt} = kp_{i+1} - (2k+k_c)p_i + kp_{i-1} \text{ for } i \geq 3. \qquad \textbf{[S3]}$$

Here, $p_i \equiv p_i(t)$ for convenience. The elongation-competent state at $i=0$ is not explicitly considered.

The recovery time τ_{rec} of a polymerase initially backtracked n nucleotides equals the first-passage time to the elongation state $i=0$. The calculation of τ_{rec} can be simplified considering a continuous model, where the position of the polymerase, x, is a continuous random variable. We define $\rho(x,t)dx$ as the probability for a polymerase to be in the interval $[x, x+dx]$ at time t. The probability density $\rho(x,t)$ evolves in time according to the following Fokker–Planck equation:

$$\frac{\partial \rho(x,t)}{\partial t} = k\frac{\partial^2 \rho(x,t)}{\partial x^2} - k_c\ \rho(x,t). \qquad \textbf{[S4]}$$

We assume that the polymerase is initially located at $x=n$ [that is, we set the initial condition to $\rho(x,0) = \delta(x-n)$ with an absorbing boundary at the elongation-competent state $\rho(0,t)=0$]. The solution of Eq. **S4** in the half-plane $x \in [0,\infty)$ can be expressed as

$$\rho(x,t;n) = \frac{e^{-k_c\, t}}{\sqrt{4\pi kt}}\left[e^{\frac{-(x-n)^2}{4k\,t}} - e^{\frac{-(x+n)^2}{4k\,t}}\right]. \qquad \textbf{[S5]}$$

The distribution $\rho(\tau_{\text{rec}};n)$ of the recovery time for a polymerase initially at n to be within the interval $[\tau_{\text{rec}}, \tau_{\text{rec}} + d\tau_{\text{rec}}]$ is given by the probability density current into $x=0$, which contains a contribution of the diffusion and a contribution of the cleavage:

$$\begin{aligned}\rho(\tau_{\text{rec}};n) &= k\left.\frac{\partial \rho(x,t;n)}{\partial x}\right|_{x=0,t=\tau_{\text{rec}}} + k_c\ S(\tau_{\text{rec}};n)\\ &= \rho_{\text{diff}}(\tau_{\text{rec}};n) + \rho_c(\tau_{\text{rec}};n),\end{aligned} \qquad \textbf{[S6]}$$

where $S(\tau_{\text{rec}};n)$ is the survival probability or the probability of a polymerase initially at n to be in $x>0$ at time τ_{rec}. The probability density current into $x=0$ due to diffusion (8) equals to

$$\rho_{\text{diff}}(\tau_{\text{rec}};n) = k\left.\frac{\partial \rho(x,t;n)}{\partial x}\right|_{x=0,t=\tau_{\text{rec}}} = e^{-k_c\,\tau_{\text{rec}}}\frac{n}{\sqrt{4\pi k\,\tau_{\text{rec}}^3}}e^{\frac{-n^2}{4k\,\tau_{\text{rec}}}}. \qquad \textbf{[S7]}$$

At time τ_{rec}, the survival probability equals to

$$S(\tau_{\text{rec}};n) = \int_0^\infty \rho(x,\tau_{\text{rec}};n)\,dx = e^{-k_c\,\tau_{\text{rec}}}\text{erf}\left(\frac{n}{\sqrt{4k\,\tau_{\text{rec}}}}\right), \qquad \textbf{[S8]}$$

where erf is the error function. The probability density current into $x = 0$ by cleavage equals to

$$\rho_c(\tau_{\text{rec}};n) = k_c e^{-k_c\,\tau_{\text{rec}}}\text{erf}\left(\frac{n}{\sqrt{4k\,\tau_{\text{rec}}}}\right). \qquad \textbf{[S9]}$$

Similarly, the probability density $R(\tau_{\text{rec}};n)$ of recovery from an initial backtrack depth n in a time τ_{rec} or recovery probability is given by

$$R(\tau_{\text{rec}};n) = 1 - S(\tau_{\text{rec}};n) = 1 - e^{-k_c\,\tau_{\text{rec}}}\text{erf}\left(\frac{n}{\sqrt{4k\,\tau_{\text{rec}}}}\right). \qquad \textbf{[S10]}$$

For the case $k=0$, the recovery probability simplifies to

$$R(\tau_{\text{rec}};n) = \text{erfc}\left(\frac{n}{\sqrt{4k\,\tau_{\text{rec}}}}\right), \qquad \textbf{[S11]}$$

where erfc is the complementary error function. Eq. **S11** is used to fit the experimental data of recovery probability in Pol I and Pol II mutants (Pol I A12.2 Δ C and Pol II Δ Rpb9) in Figs. 1 *E* and *F* and 2 *C* and *D*.

The recovery time probability density can be obtained by summing Eqs. **S7** and **S9**:

$$\rho(\tau_{\text{rec}};n) = e^{-k_c\tau_{\text{rec}}}\frac{n}{\sqrt{4\pi k\,\tau_{\text{rec}}^3}}e^{-n^2/4k\,\tau_{\text{rec}}} + k_c e^{-k_c\,\tau_{\text{rec}}}\text{erf}\left(\frac{n}{\sqrt{4k\,\tau_{\text{rec}}}}\right). \qquad \textbf{[S12]}$$

In the presence of both diffusion and cleavage, the recovery time averaged over many polymerases initially backtracked n

nucleotides or mean recovery time, $\langle\tau_{\rm rec}\rangle_n$, is given by the following equation:

$$\langle\tau_{\rm rec}\rangle_n = \frac{1}{k_c}\left[1-\exp\left(-\frac{n}{\sqrt{\frac{k}{k_c}}}\right)\right]. \qquad \textbf{[S13]}$$

Eq. **S13** is used to fit the experimental data of mean recovery time of Pol II–TFIIS in Fig. 5 and Fig. S9.

For a cleavage-deficient polymerase, where $k_c = 0$, recovery can only proceed by diffusion. In this case, the recovery time distribution has a power law tail $\sim\tau_{\rm rec}^{-3/2}$ with a diverging mean recovery time. Note that, even in the absence of cleavage, a finite value of the average is obtained when the average is done only over polymerases that recover in a given time (5 min in our experiments). An alternative statistic that can be considered is the median recovery time $\tau^\star_{{\rm rec},n}$, which can be obtained from the cumulative recovery time distribution $C(\tau_{\rm rec};n)$ that equals the survival probability at time $\tau_{\rm rec}$:

$$C(\tau_{\rm rec};n) = \int_0^{\tau_{\rm rec}} \frac{n}{\sqrt{4\pi k s^3}}\, e^{\frac{-n^2}{4ks}}\, ds = \mathrm{erfc}\left(\frac{n}{\sqrt{4k\,\tau_{\rm rec}}}\right). \qquad \textbf{[S14]}$$

At the median recovery time, the cumulative distribution equals $1/2$, which yields

$$\tau^\star_{{\rm rec},n} = \left[2\,\mathrm{erf}^{-1}\left(\frac{1}{2}\right)\right]^{-2}\frac{n^2}{k} \approx \frac{n^2}{k}, \qquad \textbf{[S15]}$$

because the prefactor $[2\,\mathrm{erf}^{-1}(1/2)]^{-2} \simeq 1.099 \approx 1$, with erf^{-1} being the inverse error function. The mode recovery time $\hat\tau_{{\rm rec},n}$ is also finite for a purely diffusive recovery and equal to

$$\hat\tau_{{\rm rec},n} = \frac{n^2}{6k}. \qquad \textbf{[S16]}$$

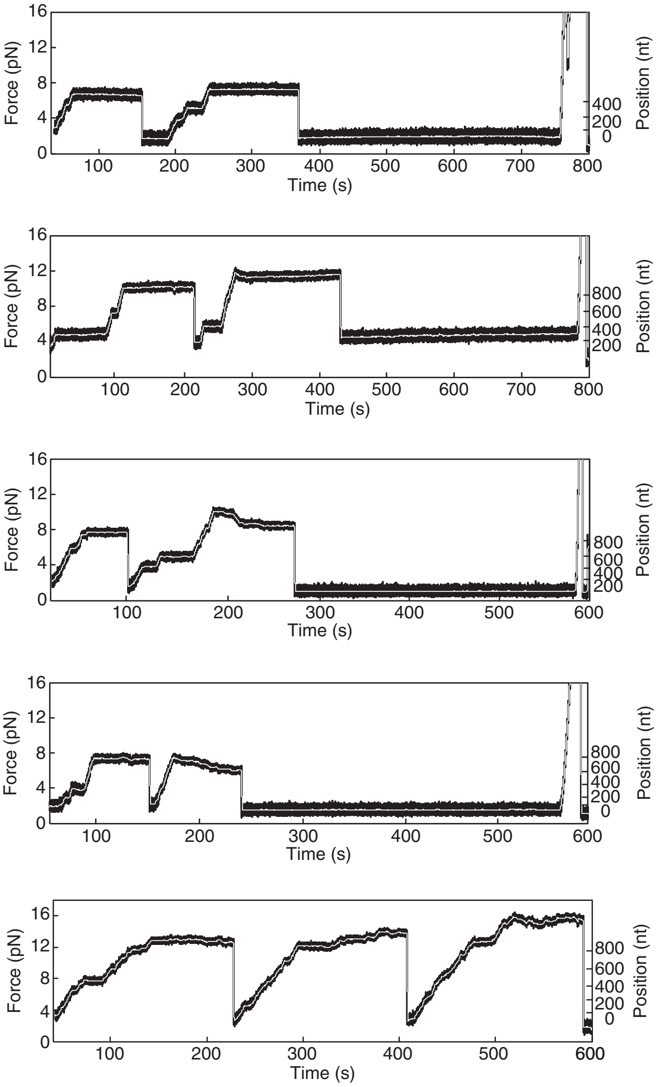

Fig. S1. Additional examples of force reduction experiments that resulted in at least one backtrack recovery.

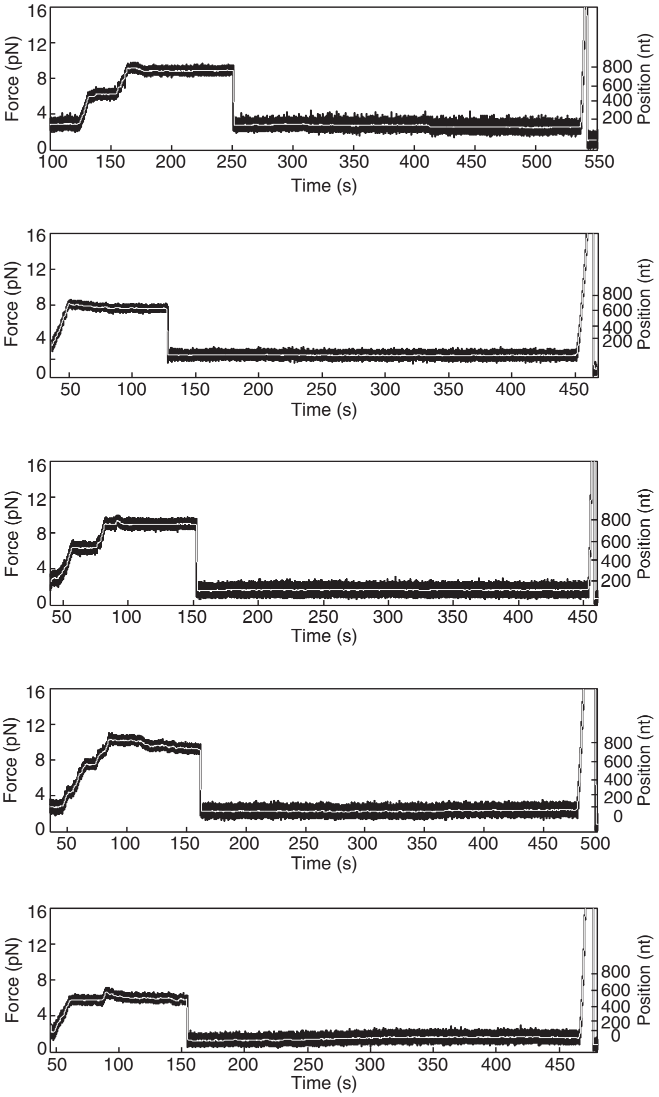

Fig. S2. Additional examples of force reduction experiments that did not result in backtrack recovery.

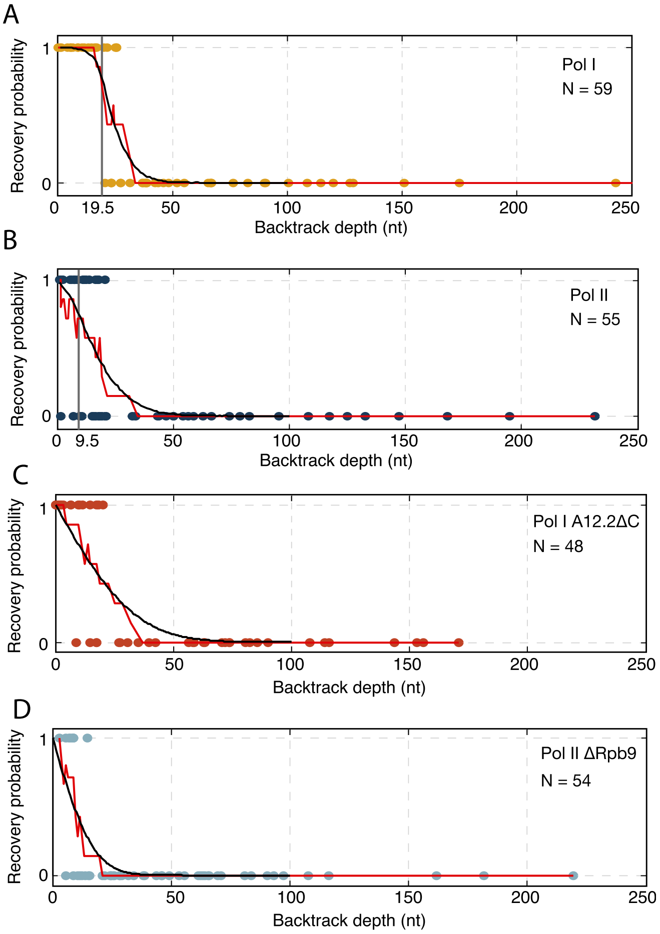

Fig. S3. Backtrack recovery differs between Pol I and Pol II (complete data). (*A* and *B*) Backtrack recovery as a function of backtrack depth for Pol I and Pol II: each data point represents one force reduction event (1, recovered; 0, not recovered), the red lines represent the smoothened data, and the black lines are fits to the model shown in Fig. 2*B*. The vertical gray lines represent the backtrack recovery threshold determined from the fit. (*C* and *D*) Backtrack recovery of Pol I A12.2Δ C and Pol II Δ Rpb9 as a function of backtrack depth: each data point represents one force reduction event. The red lines are the smoothened data, and the black lines represent the fits to the smoothened data (*Materials and Methods*) from which the diffusion rate (*k*) is extracted.

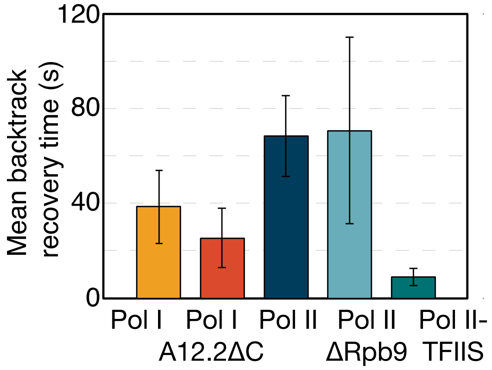

Fig. S4. Backtrack recovery times of indicated enzymes. Histogram of mean backtrack recovery times for force reduction events below 20.0 nt (which is the critical backtrack depth, λ, for Pol I cleavage activity).

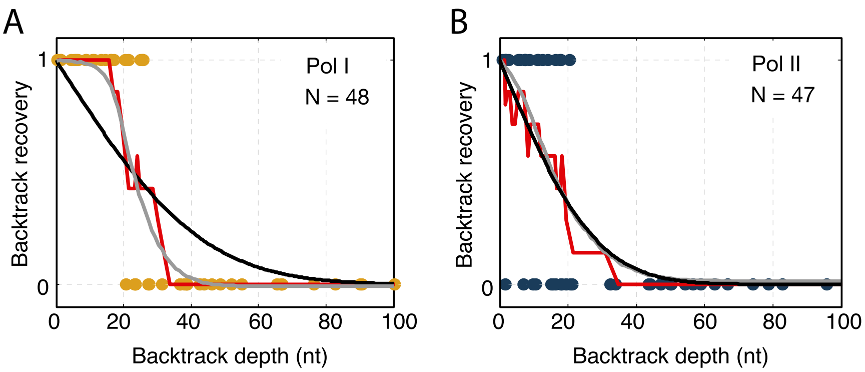

Fig. S5. Backtrack recovery of Pol I and Pol II fitted only with a diffusion rate. (*A* and *B*) Backtrack recovery as a function of backtrack depth for Pol I and Pol II: each data point represents one force reduction event (1, recovered; 0, not recovered), and the red lines represent the smoothened data. The black solid lines are fits to the model that considers only the diffusion rate parameter (k), whereas the gray solid lines are fits to the full model (k, k_c, and λ) (Fig. 2 *C* and *D*). Fitting of the Pol I data with the model with only diffusion gives significantly worse fit than the full model ($R^2 = 0.72$ compared with $R^2 = 0.96$ for the full model). However, fitting of the Pol II data gives equally good fit in both cases ($R^2 = 0.87$ for both). Note that the motivation to use the full model (with k, k_c, and λ) is provided by Fig. 2*F*, which shows that removing the cleavage pathway impacts the backtrack recovery probability.

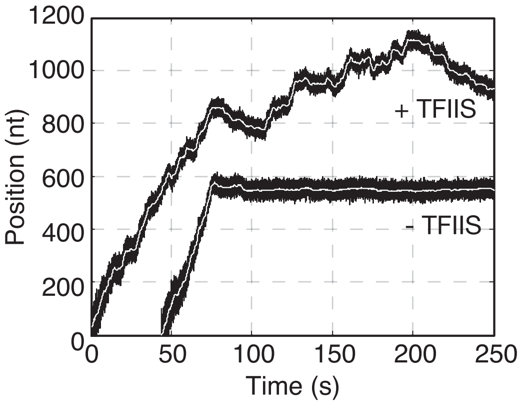

Fig. S6. Transcirption traces of Pol II in the presence and absence of TFIIS. Addition of TFIIS rescues backtrack Pol II and enables Pol II to transcribe against higher forces (also seen in ref. 7).

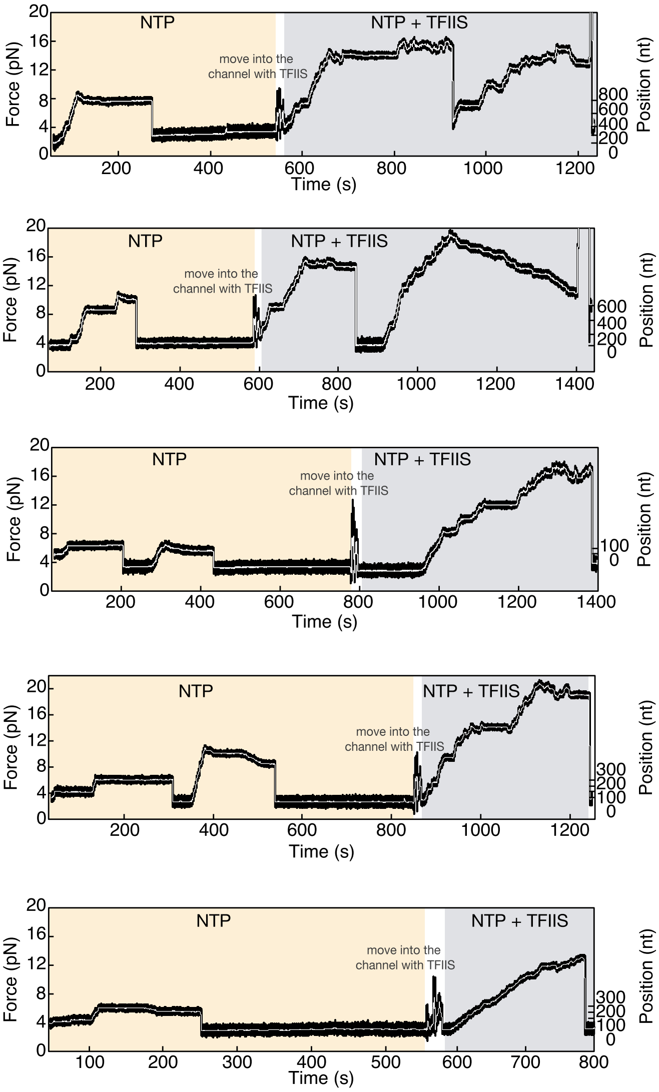

Fig. S7. Additional examples of backtrack recovery experiments of Pol II with adding TFIIS after the enzyme has already backtracked.

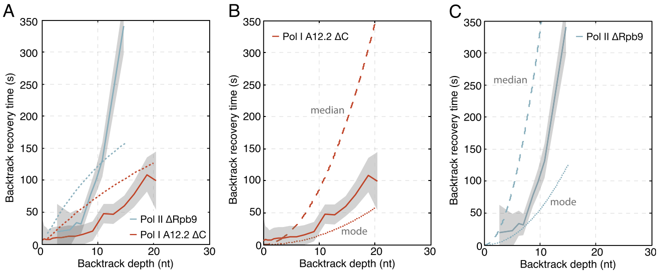

Fig. S8. Backtrack recovery times of Pol I A12.2Δ C and Pol II Δ Rpb9. Backtrack recovery times (solid lines) as a function of backtrack depth for Pol I A12.2 Δ C and Pol II Δ Rpb9, with SDs obtained by bootstrapping (gray) (*Materials and Methods*). (*A*) Dashed lines are predictions from the fit of the backtrack recovery probability data (Fig. 2 *E* and *F*). Note that the agreement between theoretical prediction and data is not very good here. We speculate that the reason for this is that, specifically for a diffusive first-passage process without cleavage, the recovery time distributions are very broad and not well-represented by an average from a finite and small dataset, like in our experimental data. Please note that the theoretical value of the average recovery time diverges here, because the distribution of the recovery time follows a $t^{-3/2}$ power law (7) (*SI Text*). (*B* and *C*) For this reason, we instead calculate the median (Eq. **S15**) and the mode (Eq. **S16**) of the recovery time distribution (*SI Text*) and compare these with the experimental data. Dotted lines represent modes, and dashed lines represent the medians of the backtrack recovery time. Note that the parabolic behavior now is correctly captured in *B* and *C*, and the measured values are between the mode and the median of the respective backtrack recovery time.

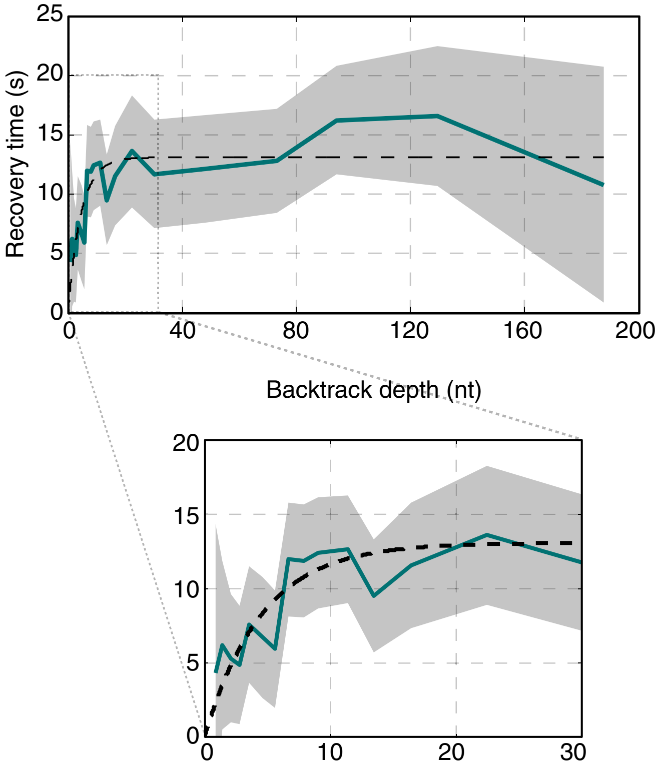

Fig. S9. Backtrack recovery times of Pol II with TFIIS. Backtrack recovery time as a function of backtrack depth for Pol II with TFIIS (solid green lines), with SDs obtained by bootstrapping (gray) (*Materials and Methods*). Dashed black lines are fits of the experimental data to the mean recovery time (*Materials and Methods* and *SI Text*).

Table S1. Summary of Pol I and Pol II transcription parameters in AF and OF mode experiments

	Assisting mode			Opposing mode		
Parameters	Pol I	Pol II	*P* value	Pol I	Pol II	*P* value
No. of traces analyzed	38	21	—	43	21	—
No. of detected pauses	60	82	—	61	69	—
Velocity (nt/s)	32.2 ± 2.5	18.7 ± 2.7	≤0.05	23.9 ± 1.7	11.7 ± 1.3	≤0.05
Pause-free velocity (nt/s)	39.2 ± 2.5	24.6 ± 2.6	≤0.05	31.4 ± 1.5	20.9 ± 1.2	≤0.05
Mean pause density (kbp^{-1})	3.8 ± 0.6	8.5 ± 2.6	0.04	6.5 ± 1.1	13.2 ± 2.0	≤0.05
Mean pause duration (s)	3.4 ± 0.6	3.5 ± 0.5	0.37	5.6 ± 1.4	6.9 ± 1.8	0.17
Arrest force (pN)	—	—	—	9.3 ± 0.5	6.2 ± 0.6	≤0.05*

*Note that all errors are SEMs. The *P* values are computed from the WRSTs performed between the corresponding values for Pol I and Pol II.

Table S2. Summary of fit parameters: Recovery by 1D diffusion and RNA cleavage

Enzymes	$k\,(1/s)$	$k_c\,(1/s)$	$\tau_c = 1/k_c\,(s)$	λ (nt)	R^2
Pol I	0.21 ± 0.13	0.019 ± 0.003	53 ± 8	20 ± 2	0.96
Pol II	0.54 ± 0.17	0.012 ± 0.003	83 ± 21	10 ± 2	0.87
Pol II–TFIIS	1.6 ± 1.2	0.076 ± 0.009	13 ± 2	∞	0.72

Table S3. Summary of fit parameters: Recovery by 1D diffusion

Enzymes	$k\,(1/s)$	R^2
Pol I A12.2 Δ C	1.16 ± 0.26	0.93
Pol II Δ Rpb9	0.30 ± 0.07	0.91