Biophysical Journal, Volume *112*

Supplemental Information

Switching between Exonucleolysis and Replication by T7 DNA Polymerase Ensures High Fidelity

Tjalle P. Hoekstra, Martin Depken, Szu-Ning Lin, Jordi Cabanas-Danés, Peter Gross, Remus T. Dame, Erwin J.G. Peterman, and Gijs J.L. Wuite

Supplemental Information

Switching between Exonucleolysis and Replication by T7 DNA Polymerase Ensures High Fidelity

T. P. Hoekstra[1,2] , M. Depken[3], S. Lin[1,2,4], J. Cabanas-Danés[1,2], P. Gross[1,2], R. T. Dame[4], E. J. G. Peterman[1,2] and G. J. L. Wuite[1,2,*]

[1]Department of Physics and Astronomy and [2]LaserLaB Amsterdam, Vrije Universiteit, Amsterdam

[3]Department of Bionanoscience, Kavli Institute of Nanoscience, Delft University of Technology

[4]Leiden Institute of Chemistry, Leiden University

Exonucleolysis processively removes correct nucleotides

Based on the low error rates reported for T7 DNA polymerase (<1 per 1,000,000) we expect that DNAp inserts few incorrect nucleotides. In addition, the rate of polymerization after an incorrect nucleotide drops drastically. Tension on the template could possibly increase the rate of incorporating an incorrect nucleotide. We expect, however, that this increase is not sufficient to increase the rate of incorporation after an error to suggest that exonucleolysis is removing solely errors. For that the rate of error incorporation should increase more than 1000-fold, which seems improbable. Moreover, it is known that exonucleolysis by Klenow polymerase does not discriminate between a correct or incorrect nucleotide (1). The combination of these arguments leads to our conclusion that the exonuclease activity frequently removes correct nucleotides.

Resolution

The spatial resolution depends on the tension applied to the DNA: with increasing tension on the DNA the difference in length between ds- and ssDNA increases, resulting in a higher resolution. We have determined the resolution at 45 pN, by estimating the noise in the base pair over time signal. We used the spatial resolution at 45 pN as a reference to calculate the resolution as a function of tension by multiplication with changing length difference (Fig. S1).

The force resolution during constant force measurements depends on the quality of our feedback system. At 45 pN the standard deviation of the force during the force clamp is ~0.4 pN. Supplementary Figure 2 illustrates this using the force and distance data that corresponds to the trace presented in Fig. 1C.

Analysis of traces at constant tension

The extension of ds- and ssDNA differ for most tensions. Above ~6 pN the ssDNA is always longer while below that force it is shorter. We conduct all of our experiments above 6 pN. Thus, at constant tension, pol activity of DNAp will cause the DNA to become more dsDNA, resulting in a decreasing end-to-end length. In contrast, *exo* activity results in lengthening of the tether. Since the length of the DNA tether is known, the position at constant tension can

be converted to the amount of base pairs and nucleotide. As a result the amount of incorporation and exonucleolysis can be given in nucleotides and followed over time.

To extract rates and pauses the DNAp position vs. time traces were analyzed by finding break points—points where the trend changes, indicating that *pol* or *exo* starts or ends. Thereto, a window of ~2 sec was slid over the data. In each window, ~50 data points were fitted independently by a pair of lines (Fig. S3C) and a single line (not shown), cross points arisen where the sum of residues of the two-lines fittings is smaller than the sum of residues of a one-line fit (Fig. S3B Since each window slides through one data point each time, the break points repeatedly showed in several windows, therefore we defined the true break points as those with an occurrence of at least 10 windows (~0.36 sec) (Fig. S3D). Finally, the true break points were connected with linear fits to extract the rates and characteristic times of activities and pauses (Fig. S3A). Drift and noise in the system caused pauses to have a rate of not exactly zero and therefore a threshold of 25 nt s^{-1} was used to discriminate a pause from *pol* and *exo* activities. This threshold is for most conditions more than a standard deviation away from the mean rate of activity. This results in the loss of some low-rate activities of *pol* and *exo* if they occurred at a rate less than 25 nt s^{-1} (false-negatives), but prevents actual pauses to interfere with the determination of *pol* and *exo* activity rate (false-positives). As a processivity threshold, the spatial resolution given in Fig. S1 was used. Events shorter than 0.4 seconds were discarded from further analysis.

Polymerization and exonucleolysis rates

The binned histograms of the rates of polymerization and exonucleolysis for all different tensions are shown in Fig. S4. The polymerization and exonucleolysis rates were extracted from least-square fits of Gaussian functions to histograms of measured rates.

Off rates

The durations of *pol* and *exo* activities were fitted with single exponential functions. The resulting fits of both activities at different tensions are shown in Fig. S5.

Pauses during *pol* and *exo*

In our current model, there should be no distinguishable difference in characteristic pause durations depending on the flanking activity. Fig. S6 shows similar distributions for the pause durations coming from or binding in *pol* or *exo* (solid circles, black and red, and empty triangles respectively). Data is combined for all concentrations and tensions.

Fraction of long pauses

The fraction of long pauses (of the total amount of pauses) does not significantly depend on the DNAp-concentration. At 5 nM the fraction is 0.46 ± 0.08 (mean ± S.D.), 0.31 ± 0.05 at 15 nM and 0.22 ± 0.05 at 25 nM after 1000 bootstraps.

Net replication rate

In previous studies the reported rates of polymerization and exonucleolysis probed at different tension; below 35 pN there was tension-dependent polymerization; as the tension increased further, exonucleolysis started to occur with no or only small dependency on

tension (2, 3). We observe both *pol* and *exo* over the full range of applied tensions (15-55 pN). However, we are able to reconstruct the previously reported behavior by weighing together both rates with respect to their relative occurrences (Fig. 2C). We do not take the off rates of both activities into account since they are similar for all tensions. Pauses are not taken into account, as their duration is dependent on the concentration of DNAp used.

Direct switches

In our assay we cannot resolve pauses below 0.4 seconds. However, direct switches (switches from *pol* to *exo* or vice versa) could still be resolved. The different directions of *pol* and *exo* in our traces results in clear, distinguishable trend change. Fig. 1D shows all possible trend changes. As long as the activities are above the threshold direct switches would still be observed.

Modeling and Maximum likelihood fitting results

We use continuum-time random walk theory to determine the four probability distributions for the time it takes to reactivation into *exo* or *pol* when having entered the pause from exo or pol. The results are

$$P_{\mathrm{EE}}(t) = \frac{k_{\mathrm{EE}}}{k_{\mathrm{EP}} + k_{\mathrm{ES}}} \frac{k_{\mathrm{SE}}}{k_{\mathrm{SE}} + k_{\mathrm{SP}}} P_{\mathrm{sol}}(t)$$

$$P_{\mathrm{EP}}(t) = \frac{k_{\mathrm{ES}}}{k_{\mathrm{EP}} + k_{\mathrm{ES}}} \frac{k_{\mathrm{SP}}}{k_{\mathrm{SE}} + k_{\mathrm{SP}}} P_{\mathrm{sol}}(t) + \frac{k_{\mathrm{EP}}}{k_{\mathrm{EP}} + k_{\mathrm{ES}}} \delta(t)$$

$$P_{\mathrm{PE}}(t) = \frac{k_{\mathrm{PS}}}{k_{\mathrm{PS}} + k_{\mathrm{PI}}} \frac{k_{\mathrm{SE}}}{k_{\mathrm{SE}} + k_{\mathrm{SP}}} P_{\mathrm{sol}}(t)$$

$$P_{\mathrm{PP}}(t) = \frac{k_{\mathrm{PS}}}{k_{\mathrm{PS}} + k_{\mathrm{PI}}} \frac{k_{\mathrm{SP}}}{k_{\mathrm{SE}} + k_{\mathrm{SP}}} P_{\mathrm{sol}}(t) + \frac{k_{\mathrm{PI}} k_{\mathrm{IP}}}{k_{\mathrm{PS}} + k_{\mathrm{PI}}} e^{-k_{\mathrm{IP}} t},$$

where the notation for the rates are as indicated in Figure 4, and the subscript on the probability densities correspond starting in *exo*/*pol* (E/P) followed by ending in *exo*/*pol* (E/P). Further, the solution escape distribution is given by

$$P_{\mathrm{sol}}(t) = \frac{k_{\mathrm{SE}} + k_{\mathrm{SP}}}{\omega_+ - \omega_-} \left(e^{-\omega_- t} (k_{\mathrm{LS}} - \omega_-) - e^{-\omega_+ t} (k_{\mathrm{LS}} - \omega_+) \right)$$

with the characteristic rates

$$\omega_\pm = \frac{1}{2} \left(k_{\mathrm{LS}} + k_{SL} + k_{\mathrm{SE}} + k_{\mathrm{SP}} \pm \sqrt{(k_{\mathrm{LS}} + k_{SL} + k_{\mathrm{SE}} + k_{\mathrm{SP}})^2 - 4 k_{\mathrm{LS}} (k_{\mathrm{SE}} + k_{\mathrm{SP}})} \right).$$

From these distributions we construct the maximum likelihood function

$$ML(k_{\mathrm{XY}}) = - \sum_{X \in \{\mathrm{E,P}\}} \sum_{Y \in \{\mathrm{E,P}\}} \sum_{i=1}^{I^{XY}} \ln P_{\mathrm{XY}}(t_i^{\mathrm{XY}}),$$

where the sums run over the observed pauses of each type, with experimentally measured pause durations t_i^{XY}. The rates are now estimated by numerically minimizing *ML* over all the rates k_{XY} of the model (see Fig. 4A). The errors are calculated by bootstrapping the data 300 times and we report the one-sigma confidence interval.

Distances to transition barrier

Assuming that the force dependence of any of the rates is dominated by one free-energy barrier, it possible to describe the force dependence in terms of a distance to a transition state:

$$k_{XY} = k_{XY}^0 e^{-d_{XY} f / k_B T}$$

This assumption is not always justified, which is the reason for not using it directly in our fits. Instead, we have binned the forces in three bins to be able to at least qualitatively judge the deviation from this basic form (which would result in straight lines in the log-plots of Fig. 4). Within the errors, all but the rate k_{SP} seem to be described by the above form. For k_{SP} we allow for next order variations with respect to force, according to

$$k_{SP} = k_{SP}^0 e^{-(\delta_{SP} f + \gamma_{SP}(f-\bar{f})^2 - \gamma_{SP}\bar{f}^2)/k_B T}$$

Here we have introduced the average force $\bar{f} = 33.1 pN$ to keep the interpretation of d_{SP} as the distance to the transition state in the center of the probed force range. The corresponding distances to the transition states and zero-force rates are given by

	k_{ES}	k_{EP}	k_{IP}	k_{LS}	k_{PI}	k_{PS}	k_{SP}	k_{SE}	k_{SL}
k^0	1.15	0.16	30.63	0.02	1.65	1.42	0.06	0.05	0.04
δ	0.02	0.19	0.42	-0.13	0.16	0.08	0.47	-0.15	-0.05
δ 68% conf *low*	0.01	-0.3	0.28	-0.18	0.12	0.05	0.43	-0.17	-0.12
δ 68% conf high	0.04	0.68	0.56	-0.07	0.21	0.11	0.49	-0.12	0.03

Table S1 The distance to the energy barrier for the different fitted rates.

In Table S1, we give the zero-tension rate for all fitted rates (k^0). Since the rates are extrapolated well beyond the probed tension regime, caution should be taken interpreting k^0.

Cost of proofreading

For above reasons on the extrapolation of the fitted rates, we use data at 15 pN to estimate the cost of proofreading at zero tension. At 15 pN, k_{Pol} is ~500 nt s^{-1} and k_{Exo} is ~150 nt s^{-1} (Fig. 2A), with the duration of both activities for ~1 sec (inset Fig. 2A). The probability of exonucleolysis at 15 pN is 0.2 (Fig. 2B). This would correspond to the removal of

approximately 150 nt for every 2000 nucleotides incorporated, and a cost of 7.5% for proofreading.

Alternative Models

Model A:

Several alternative models that could describe the data were considered. In this section we explicitly discuss three seemingly natural alternative models and their limitations.

The first model is one with four different states that are all interconnected (polymerization, exonucleolysis, unbound in solution and the long paused state) (Fig. S8A). Two reasons for discarding this model reside in the memory effect and the entrance probability of the long paused state. The memory effect in this model would be regulated through the long paused state, which would have to bias the binding at high tensions into the polymerization state. We have found, however, that the access (and exit) of long pauses does not depend on the activity before the pause. Therefore the long pauses state could not bias the binding into the polymerization state and describe the observed memory effect. Since the fractions of paused DNAp entering the long paused state are similar coming from or going into pol and exo, there is probably one common entry and exit from the long paused state.

Model B:

The second model that we discuss explains the memory effect not through a different state of the enzyme, but different states of the DNA (Fig. S8B). In this model, two unbound states exist, one in which the primer-template structure is aligned (Fig. S1) and one in which the PTS is frayed and several base pairs are opened up (Fig. S2). This model would be able to describe the data if the rates between the non-frayed and frayed DNA state are force-dependent. The effect of force could only be described, however, if the rates between the two states are on the order of seconds. In data previously acquired in our lab on the melting of DNA in the overstretching transition it was found that the melting and reannealing of tens of base pairs occurs multiple times per second (4). At this rate the fraying of a few base pairs at tensions below the overstretching transition, would equilibrate quickly and the force dependency would be lost. For that reason, this model was discarded.

Model C:

The last model we discuss here is similar to the model presented in the main text (Fig. S8C). In this case, an extra paused state is added to exonucleolysis, to justify the memory effect. This model, however, is not able to simultaneously capture the observed comparable amounts of *exo-pol* and *pol-pol* transitions at low forces, and the increased *pol-pol* transitions at high forces.

Acoustic Force Spectroscopy

We use an acoustic force spectroscopy (AFS) instrument (Lumicks BV) to apply stable low forces (<4 pN) which allows us to observe DNA end-to-end length change during *exo* activity. Details of the instrument and the technique have been described in recent publications (5, 6).In the experiments the DNA ends (pKYBI, 8.4 kbp) are bound between the flow cell

surface and a microsphere (4.5 µm, streptavidin-coated, Kisker). The tethered beads were imaged in an inverted bright-field image with a 40x microscope objective (Nikon 40x CFI Achromat Air Objective). A piezo translation stage generates a lookup table (LUT) to determine the z position of the microspheres with 50 nm resolution. 150 nM T7 DNA polymerase (New England Biolabs) in measuring buffer (10 mM Tris-HCl pH 7.5, 50 mM NaCl, and 2 mM $MgCl_2$) was flushed in flow cell after DNA tethers preparation and the end-to-end length change was tracked in real-time.

Supplementary Figures

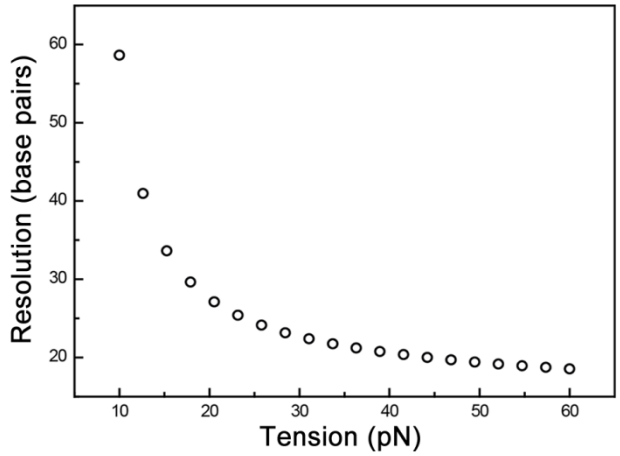

Fig. S1 Resolution in base pairs as a function of DNA tension. The resolution at 45 pN is 20 base pairs. The resolution decreases as the tension decreases because the difference between ds- and ssDNA decreases.

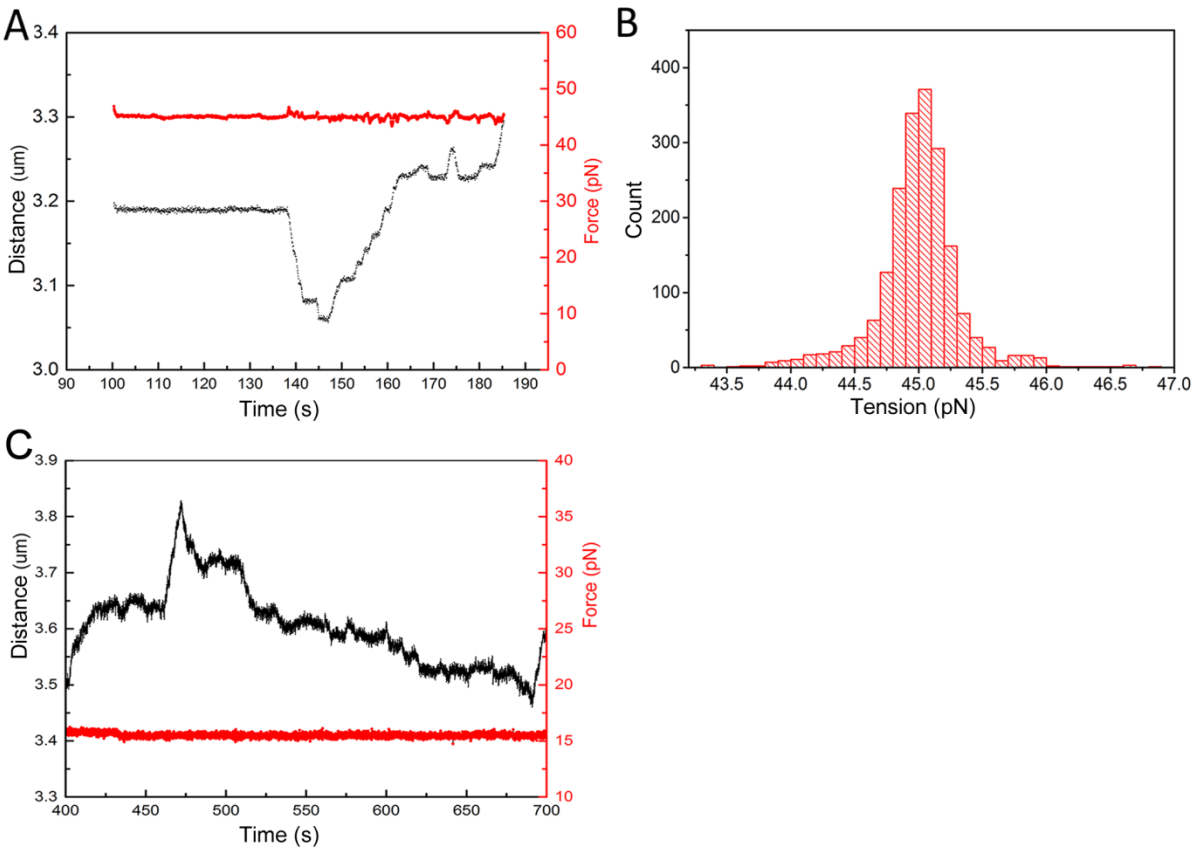

Fig. S2 Corresponding force and distance data for example trace. (A) Force and distance data over time for example trace given in Fig. 1B. (B) Histogram of the force during force clamp in panel A shows that the force is maintained constant with a standard deviation of ~0.4 pN. (C) Force and distance data over time with force kept at 16pN. The T7DNAp concentration is 15nM.

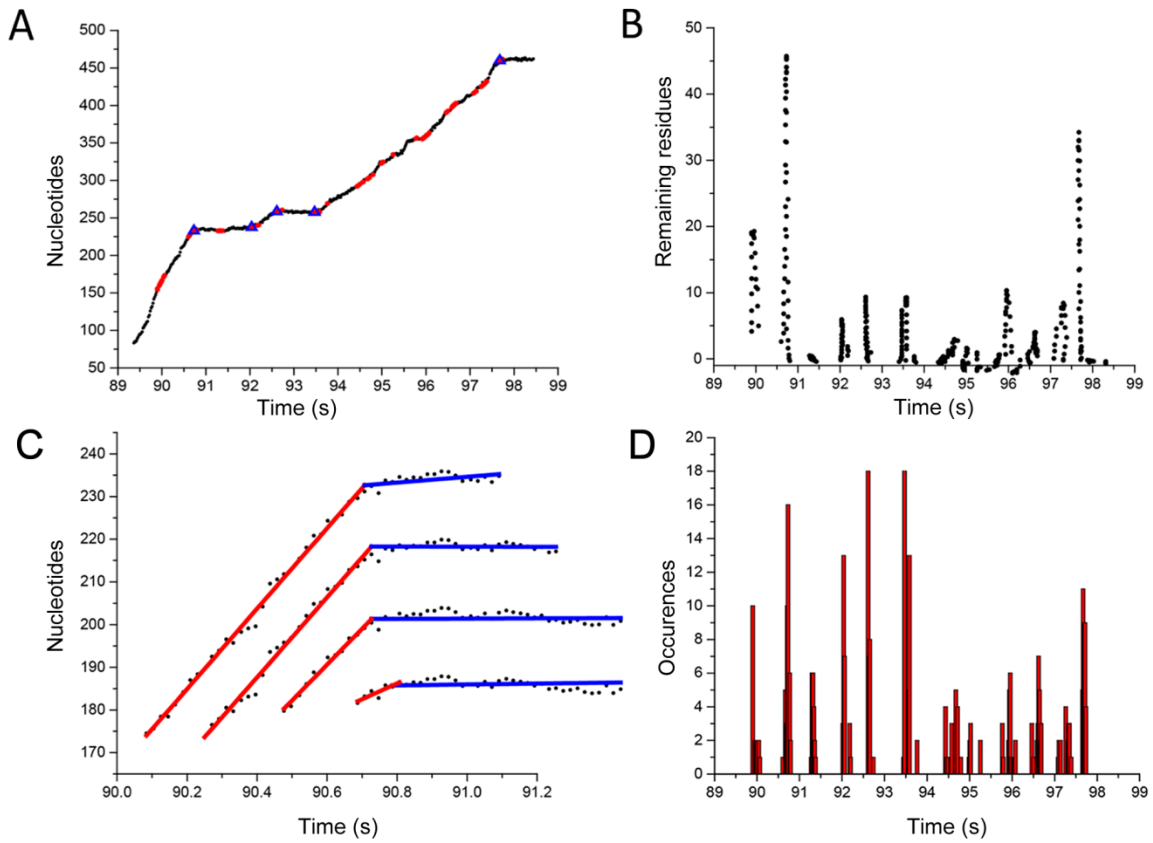

Fig. S**3 Finding breakpoints.** A) In red are all the break points found by the algorithm. B) Remaining residues after subtracting the sum of residues of the two-line fit from the sum of residues from the single linear fit from the data in the 2 sec window. C) Finding of break points in windows (several windows shown, displaced on y-axis for clarity). D) Occurrences of break point locations of data. Break points shown (blue triangles on A) are selected by a minimum of 10 occurrences of a breakpoint location.

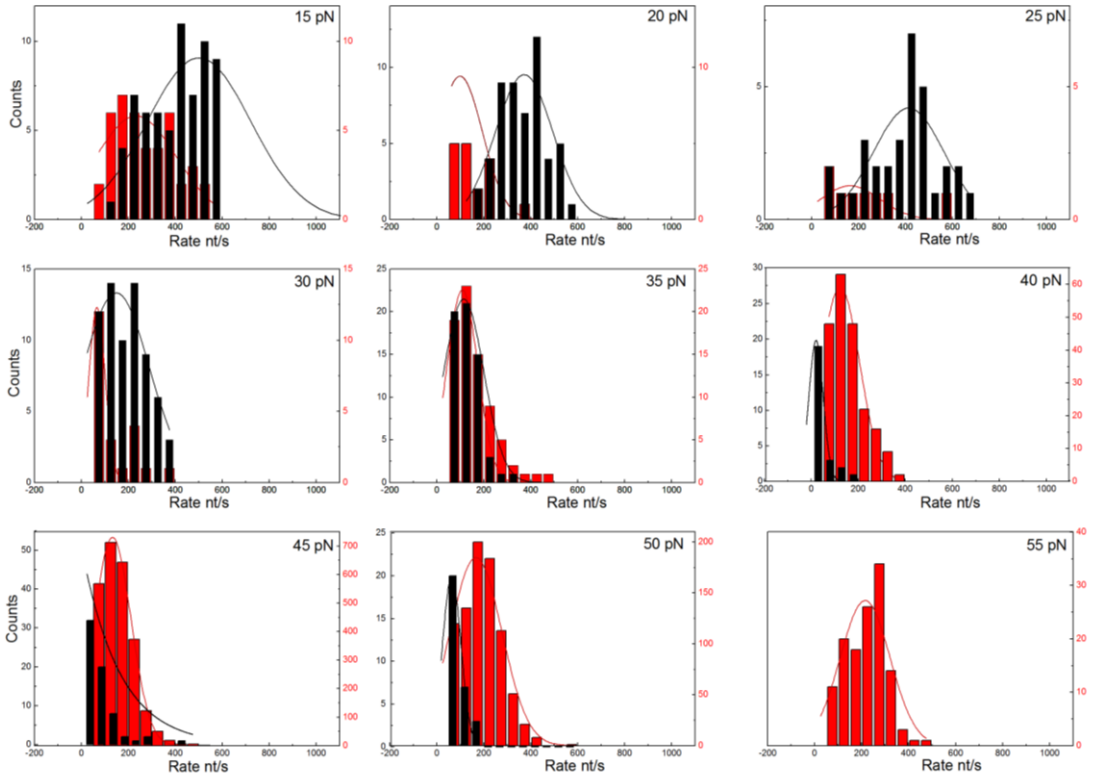

Fig. S4 Histograms of the rates of polymerization (black) and exonucleolysis (red) for different tensions. Data points in graphs: n = 98, 64, 13, 77, 71, 33, 129, 40 and 0 for polymerization in the respective panels. For exonucleolysis n = 42, 25, 8, 28, 88, 219, 2889, 949 and 141 respectively. The arithmetic mean values are respectively: 515, 436, 339, 173, 116, 78, 87, and 80 (nt s^{-1}) for polymerization and 272, 137, 229, 116, 141, 151, 139, 174, and 202 (nt s^{-1}) for exonucleolysis. The coefficient of determination (R^2) of Gaussian fitting for polymeraization: 0.89, 0.78, 0.4, 0.83, 0.98, 0.89, 0.87, and 0.96. For exonucleolysis: 0.64, 0.95, 0.36, 0.83, 0.98, 0.93, 0.98, 0.94, and 0.74.

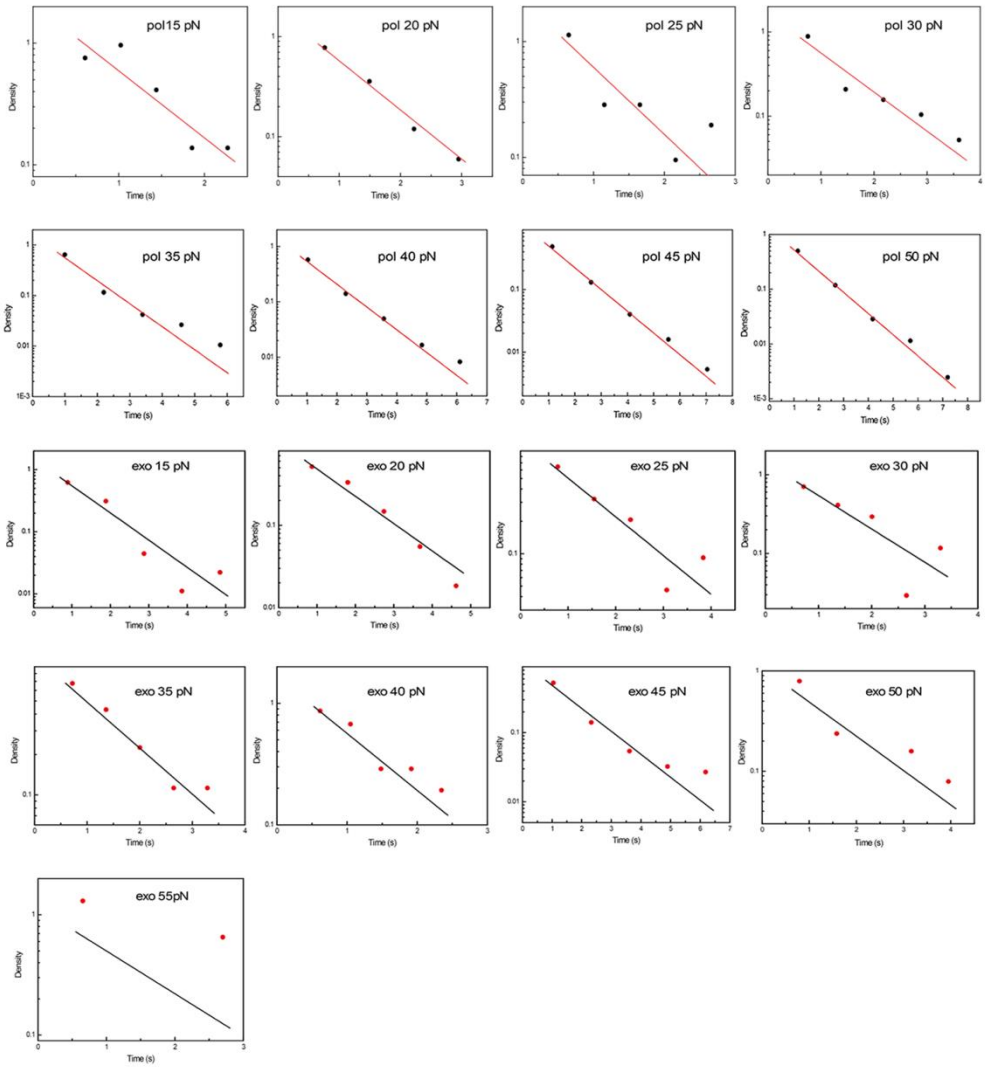

Fig. S5 Off rates of both *pol* and *exo* at different tensions. For all tensions the length of *pol* and *exo* events are described as a single exponential. The durations are log-binned.

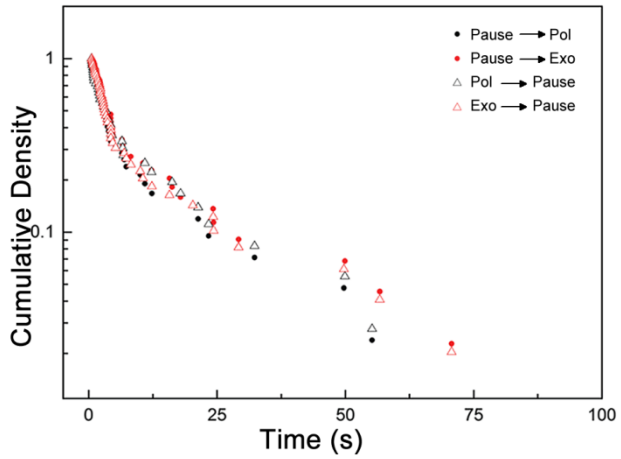

Fig. S6 Distributions of pause durations for different flanking activities show similar behavior for all conditions. Solid black circles: pauses after *pol* activity; black open triangles: pauses resulting in subsequent *pol* activity; solid red circles: pauses after *exo* activity; red open

triangles: pauses resulting in subsequent *exo* activity. Data points include data from both concentrations over 30-40 pN.

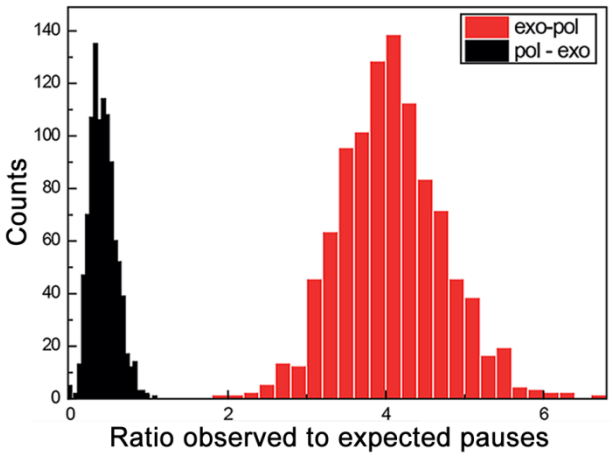

Fig. S7 Ratio of observed sub-resolution pauses to expected sub-resolution pauses for *pol-exo* (black. 8 observed vs. 19 expected) and *exo-pol* (red, 20 expected vs. 84 observed). The histogram is a result of 1000 bootstraps.

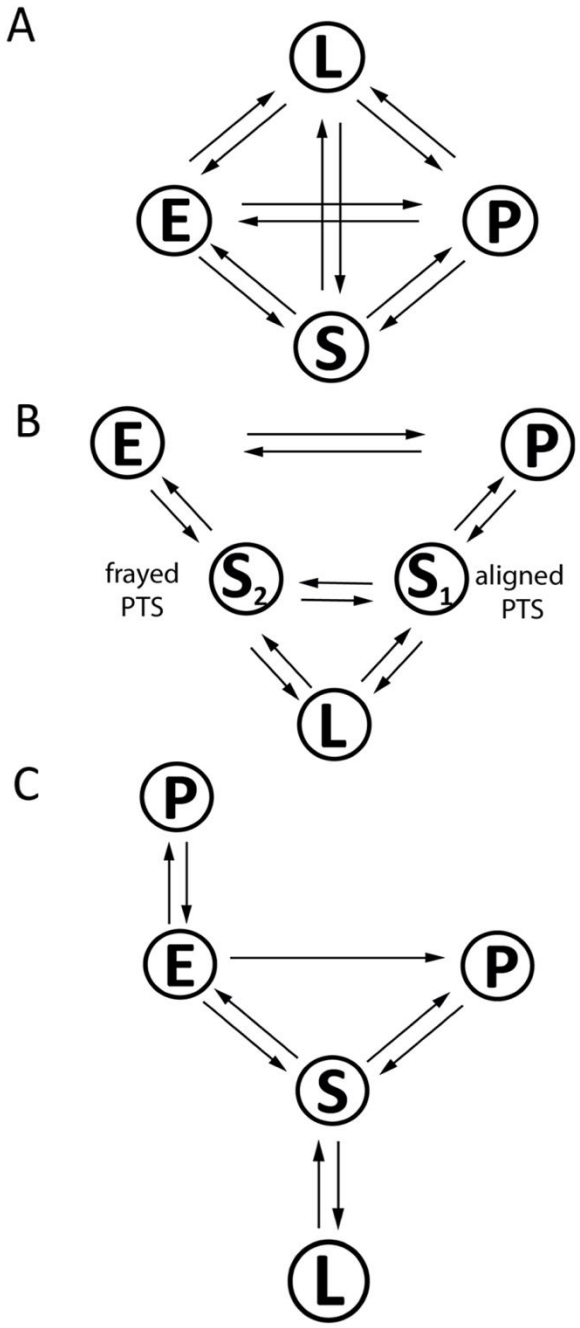

Fig. S8 Alternative models. See supplementary text for discussion.

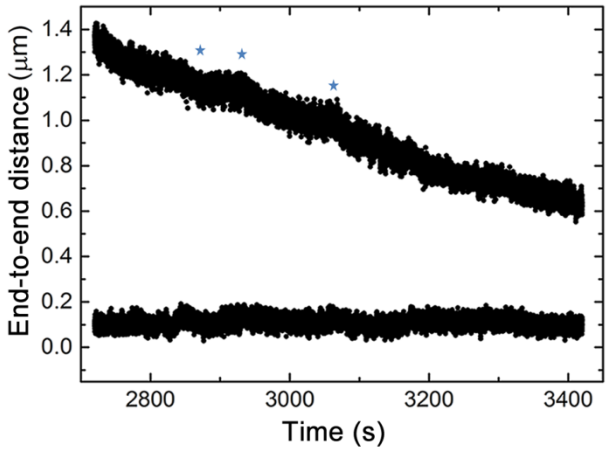

Fig. S9 Exo activity under 4 pN. AFS data was obtained at a constant of 1.1 pN. The top trace is shows *exo* activity interspersed with pauses and bursts. Stars indicate some of the clear bursts of exo activity. The bottom trace is a reference trace of a single-strand DNA tether. No activity is possible and as a result the end-to-end length stays constant during the measurement duration.

Supporting References

1. Kuchta, R. D., P. Benkovic, ,S. J. Benkovic. 1988. Kinetic mechanism whereby DNA polymerase I (Klenow) replicates DNA with high fidelity. *Biochemistry*. 27: 6716–6725.

2. Wuite, G. J., S.B. Smith,…C. Bustamante 2000. Single-molecule studies of the effect of template tension on T7 DNA polymerase activity. *Nature*. 404: 103–106.

3. Ibarra,B., Y.R. Chemla,…C. Bustamante. 2009. Proofreading dynamics of a processive DNA polymerase. *EMBO J.* 28: 2794–802.

4. Doublie, S., S. Tabor,…T. Ellenberger. 1998. Crystal structure of a bacteriophage T7 DNA replication. *Nature* 39: 251–258.

5. Sitters, G., D. Kamsma,…G.J. Wuite. 2015. Acoustic force spectroscopyNo Title. *Nat. Methods*, **12**, 47–50.

6. Kamsma, D., R. Creyghton,…E. J. Peterman. 2016. Tuning the Music: Acoustic Force Spectroscopy (AFS) 2.0. *Methods* 16: 30125–20125.