Biophysical Journal
Article

Biophysical Society

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

Tjalle P. Hoekstra,[1,2] Martin Depken,[3] Szu-Ning Lin,[1,2,4] Jordi Cabanas-Danés,[1,2] Peter Gross,[1,2] Remus T. Dame,[4] Erwin J. G. Peterman,[1,2] and Gijs J. L. Wuite[1,2,*]

[1]Department of Physics and Astronomy and [2]LaserLaB Amsterdam, Vrije Universiteit, Amsterdam, the Netherlands; [3]Department of Bionanoscience, Kavli Institute of Nanoscience, Delft University of Technology, Delft, the Netherlands; and [4]Leiden Institute of Chemistry, Leiden University, Leiden, the Netherlands

ABSTRACT DNA polymerase catalyzes the accurate transfer of genetic information from one generation to the next, and thus it is vitally important for replication to be faithful. DNA polymerase fulfills the strict requirements for fidelity by a combination of mechanisms: 1) high selectivity for correct nucleotide incorporation, 2) a slowing down of the replication rate after misincorporation, and 3) proofreading by excision of misincorporated bases. To elucidate the kinetic interplay between replication and proofreading, we used high-resolution optical tweezers to probe how DNA-duplex stability affects replication by bacteriophage T7 DNA polymerase. Our data show highly irregular replication dynamics, with frequent pauses and direction reversals as the polymerase cycles through the states that govern the mechanochemistry behind high-fidelity T7 DNA replication. We constructed a kinetic model that incorporates both existing biochemical data and the, to our knowledge, novel states we observed. We fit the model directly to the acquired pause-time and run-time distributions. Our findings indicate that the main pathway for error correction is DNA polymerase dissociation-mediated DNA transfer, followed by biased binding into the exonuclease active site. The number of bases removed by this proofreading mechanism is much larger than the number of erroneous bases that would be expected to be incorporated, ensuring a high-fidelity replication of the bacteriophage T7 genome.

INTRODUCTION

Life depends on the accurate transfer of genetic information between generations. With genomes consisting of up to several gigabasepairs (1), it is essential for DNA polymerase (DNAp), the principal enzyme catalyzing the replication of DNA, to work with high fidelity. Part of the selectivity for incorporating correct rather than incorrect nucleotides results from the compounded effects of nucleotide discrimination and the slow incorporation of erroneous bases into the nascent DNA strand. This selectivity limits the number of errors incorporated into the product to $10^{-5} - 10^{-7}$ bp^{-1} (2,3). For the 40 kb genome of bacteriophage T7, this would mean ~0.04 mutations per replication. Erroneous nucleotides that do get incorporated into the nascent strand cannot form correct Watson-Crick basepairs with the template, and therefore destabilize the primer-template structure (PTS) in the polymerase active site (pol) (4). For family A DNAp enzymes (such as T7 DNAp), this triggers the transfer of the 3′ end of the nascent DNA strand from pol to a second active site, the exonuclease active site (exo). In exo, the 3′ end is actively digested, removing erroneously incorporated bases (5). This postincorporation proofreading mechanism brings the error rate down two additional orders of magnitude, to $10^{-7} - 10^{-9}$ (6).

The presence of two thermally connected catalytic sites raises the question of how their kinetic balance is tuned to mitigate the competing demands of replication speed and fidelity. It has been shown that T7 DNAp starved of deoxynucleotides (dNTPs) processively removes bases from double-stranded DNA (dsDNA) while its exo active site interacts with the DNA. Furthermore, exonucleolysis by Klenow polymerase, a fragment of the related family A DNAp I from *Escherichia coli*, does not discriminate between correct and incorrect nucleotides (7). Both observations indicate that access to the exo site is relatively easy, and thus frequent. Yet, processive exonucleolysis by DNAp has not been observed in the presence of dNTPs in bulk. Bulk approaches measure time-averaged activities, and processive exonucleolysis runs could thus be masked

Submitted August 22, 2016, and accepted for publication December 27, 2016.

*Correspondence: g.j.l.wuite@vu.nl

Peter Gross's present address is Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany.

Editor: Antoine van Oijen.

http://dx.doi.org/10.1016/j.bpj.2016.12.044

by more frequent runs of polymerization activity. Hence, using bulk experiments, it is hard to quantify directly the dynamics between polymerization and exonucleolysis. For this reason, investigators have typically measured the nucleotide-incorporation kinetics of T7 DNAp in bulk on an exonuclease-deficient mutant, to make sure that potential exo activity would not reduce the measured polymerization rate (3).

Single-molecule force-spectroscopy experiments, on the other hand, do not require time or ensemble averaging. Researchers have performed such experiments to elucidate the mechanochemistry of DNAp, using wild-type DNAp to study pol and exo activities and the kinetic balance between them. Tension on the template strand of the DNA destabilizes the PTS and triggers transfer of the DNA 3′ end to the exo site. The destabilizing effect of force on the PTS appears to be similar to that of incorporating an erroneous base, and has previously been used to study exonucleolysis by DNAp (8–10). It also has been shown that the apparent rate of polymerization goes down as tension increases up to ~35 pN (8), after which DNAp seems to exclusively perform exonucleolysis (10,11). In a single-molecule Förster resonance energy transfer study using fluorescent base analogs, Klenow polymerase was shown to bind relaxed, matched PTS with both its pol and exo active sites (12,13). Moreover, the replicative T4 and T7 DNAp (families B and A, respectively) were previously shown to exhibit similar behavior in preferentially performing exonucleolysis when polymerization was obstructed by dsDNA ahead of the DNAp (14). In addition, processive exonucleolysis was favored over polymerization, only at tensions below 9 pN. However, exo activity of T7 DNAp in the absence of a destabilized PTS would demonstrate that replication is highly stochastic also in the absence of incorporated errors.

Here, we studied the kinetics of T7 DNAp using high-resolution optical tweezers. Our results reveal several previously unresolved kinetic states of DNAp, and provide direct access to state-occupation probabilities and rates of transition between them. We show that both pol and exo activities coexist over the full range of tensions examined. Moreover, we show that the apparent slowdown with tension in the net replication rate results from a gradually increased bias for binding into exo over pol due to tension-induced destabilization of the PTS. Based on our high-resolution data, we propose a kinetic model for T7 DNAp that differs in a few crucial ways from previous schemes, leading to a markedly different interpretation of the proofreading mechanism (2,15). Our analysis challenges the notion of DNAp proofreading as a precisely controlled mechanism that engages only when an error is encountered (16). Instead, replication appears to be prone to false positives (i.e., the removal of correctly incorporated bases) as a means to ensure the elimination of false negatives (erroneously incorporated bases that are not removed). This "paranoia" of DNAp might be a general strategy employed by high-fidelity molecular machines to deal with thermal noise at the smallest scales of living systems.

MATERIALS AND METHODS

Optical tweezers

A detailed description of the optical-tweezers instrument can be found elsewhere (17,18). In short, an Nd:YAG laser (3 W continuous wave, 1064 nm; Ventus 1064, Laser Quantum, Konstanz, Germany) was used to generate two optical traps. The laser beam was split into two orthogonally polarized beams using a polarizing beam-splitter cube. The two beams were expanded with a 1:2.67 telescope system. One beam was steered by laterally displacing a telescope lens. Two traps were produced with a high-numerical-aperture water-immersion objective. Two microspheres (1.87 μm, streptavidin coated; Spherotech, Lake Forest, IL) were held in these traps and the distance between the two microspheres was determined by a LabVIEW program (National Instruments, Austin, TX), applying template-directed pattern matching. The displacement of the trapped microsphere with respect to the trap center was detected using a position-sensitive detector. Force and distance were simultaneously recorded by a LabVIEW program operating at a sampling rate of 25 Hz.

DNA construct

The pKYBI vector was restricted with KpnI and *Eco*RI (Fermentas, Thermo Fisher Scientific, Waltham, MA). This resulted in a fragment of 8364 bp. The 5′-end overhang was filled in with biotin-14-dATP (Invitrogen, Waltham, MA), dTTP (Fermentas), and Klenow exo- polymerase (Fermentas). A 5′-biotinylated 29-mer (5′-CTCTCTCTCTCTTC TCTCTTCTCTTGTAC-3′) was ligated to the 3′ end, resulting in a DNA construct labeled with biotins on one strand, with a 25 nt 5′-end overhang. Therefore, exonucleolysis is required to generate single-stranded DNA (ssDNA) on which polymerization can subsequently take place. Switching between low and high tensions allows for repeated cycles of exonucleolysis on a single DNA molecule.

Fluidic system

Before experiments were started, the multichannel laminar flow cell was passivated with Roche blocking reagent (Roche Applied Sciences, Almere, the Netherlands) to prevent nonspecific binding of enzyme to the surface. DNA molecules were captured in buffer (10 mM Tris-HCl (pH 7.7) and 50 mM NaCl) between two optically trapped microspheres using a multichannel laminar flow cell. Tension to the DNA was applied by increasing the distance between the optical traps. Experiments were started by exchanging the buffer to buffer supplemented with 10 mM $MgCl_2$, T7 DNAp (New England Biolabs, Brookfield, WI), and dNTPs (0.6 mM each; Fermentas).

RESULTS

To investigate the kinetic balance between replication and proofreading by T7 DNAp, we used a high-resolution, double-trap, optical-tweezers assay that allows direct observation of polymerization and exonucleolytic activity (8). A dsDNA molecule of 8 kb with a 25-nucleotide (nt) 5′ overhang was attached between two optically trapped polystyrene microspheres (Fig. 1 *A*). At the start of each experiment, we applied high tensions (above 45 pN) to deliberately further extend the ssDNA overhang by means

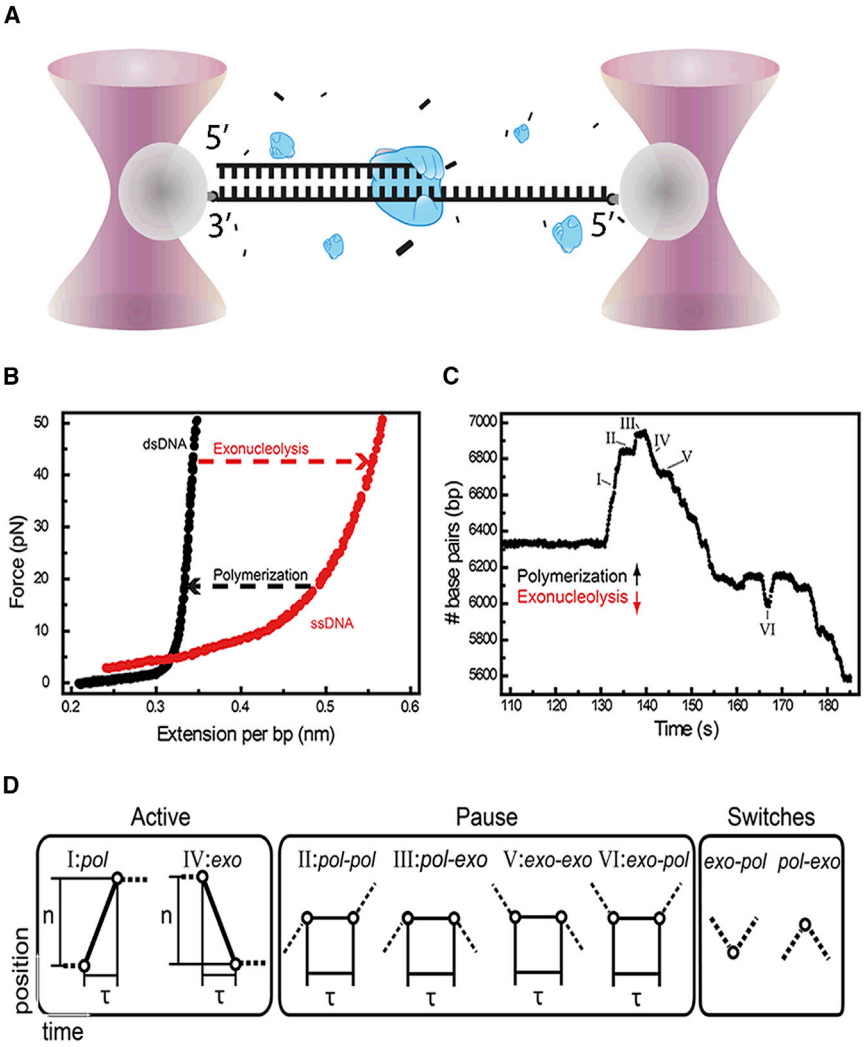

FIGURE 1 Optical-tweezers assay to monitor T7 DNAp activity. (*A*) Schematic of the experimental layout. A DNA molecule with a 5′ overhang (*right*) is held between two optically trapped microspheres. In the presence of T7 DNAp and nucleotides, activity is measured by determining DNA length changes while clamping the force. (*B*) Force-extension curves of dsDNA and ssDNA. At tensions above 6 pN, polymerization by DNAp at constant tension results in shortening of the DNA construct, whereas exonucleolysis leads to lengthening (*dashed arrows*). (*C*) Typical time trace of T7 DNAp activity at a force of 45 pN, with six different behaviors indicated. (*D*) Schematic of observed behaviors: polymerase, exonucleolysis, four different types of pauses, and two direct switches in activity. In these diagrams, solid lines between circles represent current behavior, and dashed lines indicate preceding and subsequent behaviors. The axes are equivalent to those in (*C*).

of the intrinsic T7 DNAp exonuclease activity, which is favored in the high-tension regime. Because the end-to-end lengths of ds- and ssDNA differ significantly at tensions above ~6 pN, the conversion of ss- to dsDNA by the replication reaction results in a decreased length of the DNA tether, whereas exonucleolysis results in tether lengthening (Fig. 1 *B*). From the DNA tether length and the known tension, the position of the ss-/dsDNA junction and the DNAp can be calculated. Our instrument can detect events as short as 0.4 s at a spatial resolution of 20 bp at 45 pN (Fig. S1 in the Supporting Material). To ensure sufficient resolution, we focused our experiments on data obtained with applied forces above 15 pN (see Fig. 1 *B*). This allowed us to study the different kinetic states of T7 DNAp in great detail. A representative trace of T7 DNAp activity is presented in Fig. 1 *C*, which shows bursts of polymerization and exonucleolysis activity interspersed by pauses (Fig. S2). We analyzed the traces by determining the transition points between trend changes (Fig. S3) and thereby segmenting the traces in periods of distinct activities (polymerization, exonucleolysis, and four kinds of pauses; Fig. 1 *D*). From the slope of bursts of polymerization ($n = 525$) and exonucleolysis ($n = 4382$) activity, we could directly determine their rates.

Polymerization and exonucleolysis at different tensions

We first determined the pol and exo rates as a function of DNA tension during bursts of activity (Fig. 2 *A*; Fig. S4). In accordance with previous studies, we observe that the rate of polymerization decreases with DNA tension up to 35 pN. The polymerization rates we find (up to 500 nt s^{-1} at a tension of 15 pN) are significantly higher than those previously obtained in bulk (300 nt s^{-1}) (6) and optical-tweezers (100 nt s^{-1}) experiments (8), but are close to those reported in a recent magnetic-tweezers study (14). The difference between this study and previous optical-tweezers work is likely a consequence of our improved temporal resolution, which allows the identification (and exclusion) of

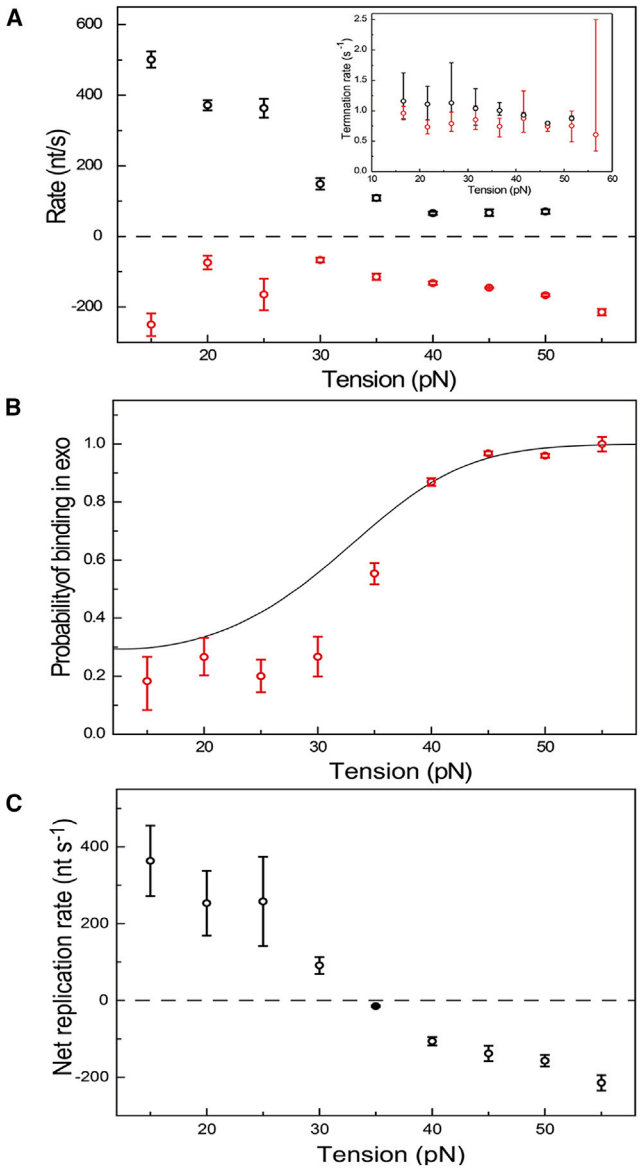

FIGURE 2 Polymerization and exonucleolysis occur over the full range of tensions applied to the DNA. (*A*) The rate of polymerization (*black*) decreases with tension, whereas that of exonucleolysis is relatively constant (*red*; 145 ± 1 nt s^{-1}, mean ± SE, $n = 4382$). Inset: the off-rates of both polymerization (*black*) and exonucleolysis (*red*) activities do not depend on tension. (*B*) The relative probability of DNAp binding with its exo active site depends strongly on tension (mean ± SD). The solid line is not a direct fit to these data, but is calculated using the rates fitted to the complete time statistics of pauses (see Supporting Material). (*C*) Reconstruction of the net replication rate, obtained by adding the polymerization and exonucleolysis rates weighted for their relative occurrences, showing a switch in net activity from polymerization to exonucleolysis at 30 pN. Data obtained at all concentrations studied are combined.

short, previously undetected pauses in-between replication events. Surprisingly, we now also observe pol activity at high tension, including tensions close to the dsDNA overstretching transition at ~65 pN (19). The polymerization rate in these particular events is relatively low (~80 nt s^{-1}) and does not appear to depend on a further increase in tension, in agreement with Manosas et al. (14). Our data on T7 DNAp exonucleolysis suggest that it removes correct bases (see also Supporting Materials and Methods) at a constant rate of ~150 nt s^{-1} ($n = 4382$) over the full range of tensions studied. Previously, processive force-induced exonucleolysis measured using direct-observation methods like the one used here was only reported at tensions above 35 pN, at a rate of ~25 nt s^{-1} (8,10). Again, the difference with regard to our data is likely the result of the better temporal and spatial resolution of our instrument.

Next, we determined the time DNAp stays active in both pol and exo. At each tension, the durations of polymerization and exonucleolysis events are exponentially distributed (Fig. S5). Fits to the durations yield an off-rate of ~1 s^{-1} for both pol and exo activities that does not depend strongly on DNA tension (*inset*, Fig. 2 *A*). For pol exchange, this value is consistent with rates found in bulk experiments, whereas for exo on dsDNA it is three orders of magnitude lower than the previously reported off-rate on ssDNA (1000 s^{-1}) (15). To our knowledge, the off-rate for exo on dsDNA has not been directly measured before.

Another important measure to determine the balance between pol and exo is the relative probability of either pol or exo activity occurring at each tension. To determine this, we counted the number of exo events and divided it by the total number of pol and exo events per tension bin. Fig. 2 *B* shows that the probability of binding into exo increases drastically above 35 pN. This has consequences for fidelity: the incorporation of erroneous bases destabilizes the PTS and triggers their removal by predominant binding into exo. Using the relative probability of pol and exo binding as a weighting factor to calculate (see Supporting Materials and Methods) the averaged (net) polymerization and exonucleolysis rates at different tensions yields values very similar to those obtained in previous, lower-resolution experiments (Fig. 2 *C*) (8,10). This supports the notion that averaging can obscure important features that determine the balancing dynamics between the two active sites.

Pausing reveals multiple kinetic states

Although the biological functions of pol and exo in ensuring proper DNA replication are straightforward, the role of pausing remains unclear. A pause indicates that DNAp either is not present at the ss-/dsDNA junction or is present and not actively bound, preventing catalysis by another DNAp. To unravel the origin of replication pauses, we systematically investigated the pause statistics and how it depends on tension and DNAp concentration. With our high-resolution experiments, we not only can distinguish pauses that occur during transitions between pol and exo, but we are also able to detect pauses during pol and exo (Fig. 1 *D*).

First, we considered the distribution of pause durations. This distribution is not exponential, but is well filled by a double exponential with two well-separated average pause durations (Fig. 3 *A*). Hence, the pauses do not represent a single state, but rather a composite of at least two distinct states. A comparison of different concentrations of DNAp (5, 15, and 25 nM) reveals that the shorter pauses depend significantly on concentration (Fig. 3 *B*). Therefore, the shorter pauses most likely are dominated by events that represent the binding of DNAp from solution to the DNA in either pol or exo, since these processes are concentration dependent. In contrast, long pauses presumably are caused by a state in which DNAp is bound to the DNA but replication is blocked, since their average duration is mostly independent of DNAp concentration. Blocked replication could be caused by DNAp being bound tightly to ssDNA close to the junction, or bound to the junction in a wrong configuration. We further observe that the duration and relative occurrence of long pauses do not depend significantly on prior activity (Fig. S6).

An additional pause state during polymerization

More insight into the paused states is obtained by considering the statistics of activities that precede and follow pauses. Based on existing models (15), where exiting activity from both exo and pol takes the system to the same solution state, we would expect the statistics of binding into exo and pol to be independent of the activity before the pause. At low tensions this is indeed what we observe (Fig. 3 *C*). Surprisingly, when the PTS is destabilized by tensions of 30 pN or higher, the system retains a memory of the previous activity (*gray area* in Fig. 3 *C*): DNAp enzymes that pause during pol activity have a higher probability of resuming pol activity than proteins that pause during exo activity. We considered several different models to explain this observation (see Supporting Materials and Methods). However, the only model that seems to match our data is one in which the memory effect is due to an inactive state that is accessible only from pol, in which DNAp does not unbind from the DNA and thus retains a memory of where it originated. To account for the memory effect, the additional paused state has to be accessible from pol at high tensions and to return to pol with high probability. In our model, this pol pause could be part of the DNAp incorporation cycle: the state might be entered when DNAp checks the basepairing of the incoming nucleotide with the template strand, but is unable to incorporate it instantly because of tension-induced deformation of the PTS. DNAp could escape from this paused state either by release of the incoming nucleotide or by the substantially slowed-down addition to the nascent strand. The lifetime of this pol pause is expected to be concentration independent, but as these pauses only constitute a small fraction of the total number of short

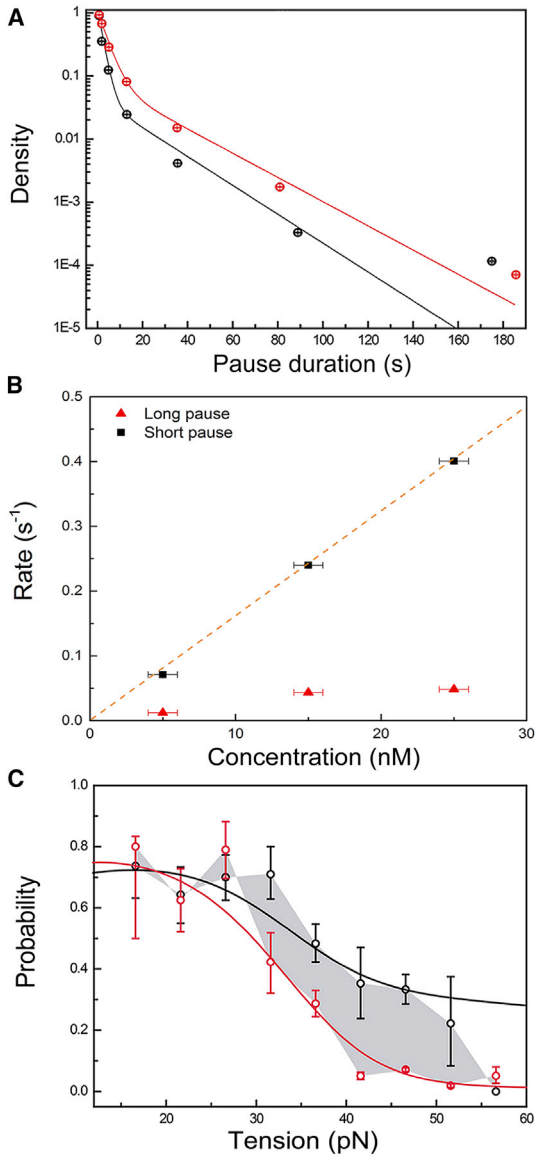

FIGURE 3 Distribution of pause durations. (*A*) Black, high concentration (25 nM, 1535 pauses, all tensions); red, low concentration (15 nM, 1430 pauses, all tensions); solid lines, double-exponential fits. For clarity, the data and fits have been normalized to 1 at 0.4 s. The error bars represent the 95% confidence interval of 10,000 bootstrapped data sets. (*B*) Binding rates for three concentrations (1000 bootstraps). The binding rate is the inverse of the characteristic pause duration. Black square, short pause; red triangle, long pause; 25 nM (1535 pauses, all tensions); 15 nM (1430 pauses, all tensions); 5 nM (313 pauses, 30 pN). The error bars in *x* reflect the error in concentration we expect due to absorption of protein on the walls of our flow chamber. The *y* error bars are smaller than the dots indicating the data. The orange dashed line is a linear fit of the short pause rates and is drawn through zero to guide the eye. (*C*) The probability of binding in the pol active site after a pause preceded by polymerization (*black*) or exonucleolysis (*red*) activity (mean ± SD, data for all concentrations and tensions are combined). The gray area marks the difference between the probabilities. The solid lines are not a direct fit to these data, but are calculated using the rates fitted to the complete time statistics of pauses (see Supporting Material).

pauses at all tensions, we still expect to see the concentration dependence observed in Fig. 3 *B*.

Exo-to-pol direct switching

Bulk studies assume a direct switching back and forth between the exo and pol sites without DNAp unbinding from the PTS (15). Yet, recent magnetic-tweezers studies of T4 and T7 DNAp showed switching only from exo to pol (12). With our approach we can infer a direct switch in enzyme activity from a sudden change in the direction of the trace (see Fig. 1 *D*). We indeed observe direct switches, but these switches might (in part) be due to pauses that are too short to be resolved given the time resolution of our instrument. To estimate the fraction of switches that originate from subresolution solution pauses, rather than from direct switches, we extrapolate our fits of the solution pauses into the subresolution regime to calculate the number of expected subresolution pauses. For our data set, the expected number of subresolution pauses going from pol to exo was 19, compared with eight detected apparent direct switches. Therefore, with the current temporal resolution, we find no evidence for direct switching from pol to exo. In contrast, for exo to pol, 20 subresolution pauses were expected, which is far fewer than the 84 apparent direct switches detected. Taken together, these results provide clear evidence for a direct switch only from exo to pol (Fig. S7).

Kinetic model of T7 DNAp for replication, proofreading, and pausing

Previously, a three-state model for T7 DNAp activity was inferred from biochemical studies (15). For clarity, we start with this model (illustrated in *black* in Fig. 4 *A*), and expand it step by step to account for our new experimental observations. The original model suggests that processive polymerization and exonuclease activities are linked through two different pathways: DNAp can switch intermolecularly through solution by unbinding and rebinding of a different DNAp, or it can switch intramolecularly by switching to the other active site without unbinding. With the high resolution of our instrument and our ability to track the kinetic history of individual enzymes, we have demonstrated that this kinetic model does not describe our data, as it does not explain the observed memory of the state before the pause. Moreover, the observed long-pause state in our pause-duration distribution data is absent in the three-state system.

Hence, we considered several models: one in which all four states (polymerization (P), exonucleolysis (E), short pause (S), and long pause (L)) are interconnected (Fig. S8 *A*); one in which we consider that the memory effect is caused by fraying of the DNA (Fig. S8 *B*); and one that assumes that an extra paused state is present in the exonucleolysis state (Fig. S8 *C*). These models are discussed in Supporting Materials and Methods; all three models can be excluded, for different reasons. The model displayed in Fig. 4 *A* fits all of our observations. This model has three modifications with respect to the original three-state model. First, we observed evidence for direct exo-to-pol switching, but not for direct pol-to-exo switching. Therefore, we omit the latter transition in our model. Note that in previous studies, evidence was obtained for direct exo-to-pol switching, and it was inferred that the reverse transition also occurs (15). Second, an additional pause state is added to account for the long pauses (see Fig. 3 *A*). We have not observed any indication that these pauses occur predominantly before or after pol and exo activity (Fig. S6), and their relative occurrence does not significantly depend on concentration. Since entering the S state (short pause) is governed by the stability of the DNAp dwelling in either the pol or exo state, entering S is not concentration dependent, but its duration is strongly concentration dependent. The "L pause" (i.e., the long pause, illustrated in *green*) can be accessed directly by DNAp binding to the DNA from solution (S pause) similarly to exo and pol states. Therefore, the relative frequency of L states occurring with respect to the other states is independent of concentration. Additionally, the L-state duration is governed by a concentration-independent unbinding rate (Fig. 3 *B*). Third, a pause state that might be entered during incorporation is added (I, illustrated in *blue*) to account for the memory effect DNAp displays at high tension (Fig. 3 *C*). It is interesting to note that Manosas et al. (14) proposed a comparable model describing T4 DNAp kinetics. In their model, five states, including pol, exo, and different inactive states (pauses), are defined. Their model includes direct switching exclusively from exo to pol; otherwise, the transition occurs via an inactive state. However, the intermediate on-pathway pause they found, which is equivalent to our short-pause state, takes place intramolecularly, whereas we observe that the DNAp dissociates from DNA and rebinds.

On the basis of our new kinetic model, we calculated the theoretically expected distributions of the activation times for all types of detectable transitions (see Supporting Materials and Methods). We then used maximum-likelihood estimation with respect to the theoretically derived distributions to extract the rate constants. To fit the model without making assumptions for tension dependence, we binned the data into three tension bins. We further bootstrapped the data 200 times to estimate the sensitivity in fit parameters. The results of the maximum-likelihood estimates are shown in Fig. 4, *B–E*, together with the fits to a simple force-dependent rate of the form $k_{XY} = k^0_{XY} e^{-\delta_{XY} f / k_B T}$ (for k_{SP} a more complicated form was used; see below). Consistent with our direct observations, the binding rate from solution to the pol active site (k_{SP}) decreases with tension, whereas the binding rate from solution to the exo active site (k_{SE}) increases. At the same time, the binding rate into the long pause appears to be insensitive to tension (Fig. 4 *B*). The binding rate into pol is the only rate that cannot be fully

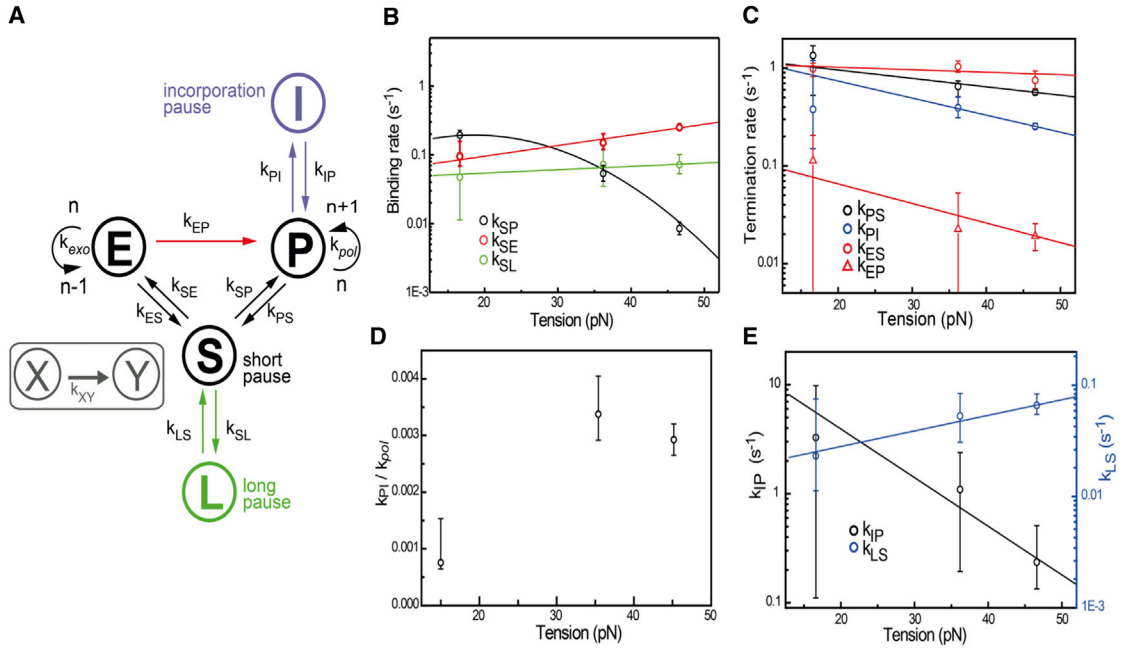

FIGURE 4 Kinetic model of T7 DNAp fitted to the experimental data. (*A*) Scheme of the kinetic model, with rates indicated. The original model (based on biochemical data (15)) is indicated in black. On the basis of our experimental results, three modifications were made. Green, long-pause state; blue, incorporation pause; red, direct switching occurring only from the exo to the pol active site. (*B–E*) Results of maximum-likelihood fitting of the model in (*A*) to the experimental data. Here we choose to bin our data in three forces (10, 35, and 45 pN) to minimize the error bars, allowing us to constrain our fits optimally. Using more bins would increase the error bars while not providing much additional information about the general trend of the data. Sensitivity in the parameters is estimated by bootstrapping the data (300 times), and the 95% confidence interval is shown. Except for k_{SP}, all rates are fitted with $k_{XY} = k^0_{XY} e^{-\delta_{XY} f / k_B T}$. $\boldsymbol{k_{SP}}$ is fit to a second-order force dependent rate. (*B*) Binding rates ($\boldsymbol{k_{SP}}$, $\boldsymbol{k_{SE}}$, $\boldsymbol{k_{SL}}$) out of solution as function of tension. (*C*) Off-rates of polymerization and exonucleolysis activities ($\boldsymbol{k_{PS}}$, $\boldsymbol{k_{PI}}$, $\boldsymbol{k_{ES}}$), and the direct-switching rate from exo to pol ($\boldsymbol{k_{EP}}$) as a function of tension. (*D*) Ratio of the rate into the incorporation pause ($\boldsymbol{k_{IP}}$) over the polymerization rate ($\boldsymbol{k_{Pol}}$) as a function of tension. (*E*) Recovery rates from incorporation pauses ($\boldsymbol{k_{PI}}$) and long pauses ($\boldsymbol{k_{LS}}$) as a function of tension.

described by the simple force dependence shown above. In this case, we expand the exponent to second order in the force, and fit the higher-order exponential function $k_{SP} = k^0_{SP} e^{-(\delta_{SP} f + \gamma_{SP}(f-\bar{f})^2 - \gamma_{SP}\bar{f}^2)/k_B T}$. Here we have included the average force $\bar{f} = 33.1$ pN of the experiments in the second-order force term to maintain the interpretation of δ_{SP} as a distance to a transition state in the experimentally probed range of forces.

Direct switching from exo to pol activity predominantly occurs at low tension (k_{EP}), whereas all other rates for pol and exo activities (k_{PS}, k_{PI}, k_{ES}) are largely insensitive to tension (Fig. 4 *C*). The relative probability of entering an incorporation pause during a single nucleotide incorporation event (k_{PI} / k_{Pol}) increases with tension, since the polymerization rate decreases (Fig. 4 *D*), thereby increasing the relative probability of entering the pause state. We speculate that incorporation pauses are caused by distorted PTS. Over the full range of tensions tested, the escape rate from this incorporation pause back to pol activity (k_{IP}) is of similar magnitude as the binding rates from solution (k_{SP} and k_{SE}) (Fig. 4 *E*). The escape rate from the long-pause state to solution (k_{LS}) also remains constant and is considerably lower than all other rates. From the fits, we can also estimate the distances (DNA length changes) to the transition states for all transitions (see Supporting Materials and Methods). The DNA length changes are all in the subnanometer range. For example, the DNA length change that occurs between pol and exo activities is consistent with that of the melting of ~3 bp (0.62 nm), in agreement with structural data (20,21). Moreover, we can use the obtained rates to make predictions about the probability of binding in exo (Fig. 2 *B*), and the memory effect (Fig. 3 *C*). These predictions are plotted in Fig. 4, *B–E* (*solid lines*), and show that the obtained rates fit our data very well. In conclusion, our kinetic model reveals new, to our knowledge, pathways in the mechanochemistry of T7 DNAp, and is consistent with our single-molecule data as well as previous biochemical (6,15) and structural (20,21) studies.

DISCUSSION

Using high-resolution optical tweezers, we examined how the dynamics of DNAp changes with the amount of tension, destabilizing the PTS. Since our data have sufficient resolution to discriminate between different dynamical behaviors—keeping track of the origin, destination, and transition time for each transition—we are able to form a clear picture of the underlying dynamical process. In short, we find

evidence for a bound inactive state, an incorporation pause induced by a destabilized PTS, and direct switching from exo to pol activities, but we find no evidence for pol-to-exo direct switching. Here we focused on the effects of destabilizing the PTS, but we also probed a sufficient range of DNAp concentrations to detect concentration dependencies.

In our experiments, we observe both a slowdown of polymerization and an increased relative probability of being trapped in the incorporation pauses at increasing tension. The decrease of the polymerization rate can be explained by T7 DNAp undergoing force-sensitive structural changes during polymerization, when the finger domains have to close to correctly align and incorporate a nucleotide (22). In a physiological setting, our high-tension situation could correspond to an erroneous nucleotide that cannot properly basepair with the template strand, increasing both the probability of unbinding of the incoming nucleotide and the probability of unbinding of DNAp from its pol active site and rebinding with the exo active site (12,23,24). We find that the probability of entering an incorporation pause increases fivefold when the tension is increased to >35 pN. Consequently, these pauses compete with polymerization and lower the net polymerization rate of bound DNAp. It is remarkable that polymerization still occurs at tensions as large as 50 pN, when the PTS is perturbed. This relative indifference of polymerase activity to PTS perturbation might help the enzyme to overcome other obstacles during replication, such as proteins bound to the ssDNA (gp2.5) or DNA hairpins.

We observe exo-to-pol direct switching, consistent with biochemical and single-molecule data for T7 and T4 DNAp (14,15). In a physiological setting, it would be especially important to escape processive exo and continue replication, since an error is quickly removed and further exonucleolytic activity would be deleterious. At the same time, we find no evidence for direct switching from pol to exo. Hence, our data show that unbinding from pol and rebinding in exo is the main postincorporation error-correction pathway for T7 DNAp. The bias for exo at higher tension has a clear analog in proofreading: when the PTS is destabilized by an error added in pol, a bias toward exo binding ensures that such errors will be removed with high probability.

The biological function, if any, of the long pauses remains unclear. The length of the pauses indicates strong binding of the DNAp in the vicinity of the junction. The minimal tension dependence of the rate into the long pause (Fig. 4 *B*) suggests that DNAp binds to ssDNA close to the junction. It has been noted before that DNAp has a strong affinity for ssDNA (15,25). At the replication fork, the recruitment of DNAp is facilitated by gp4. The helicase/primase is capable of binding up to three DNAp molecules, increasing the local DNAp concentration to directly start replication of the Okazaki fragment when the primer is synthesized (26,27). We speculate that the long-pause state might represent inactive DNAp bound close to the junction, which could facilitate DNAp rebinding with its pol active site.

It was shown in a high-resolution optical-tweezers study of bacteriophage Φ29 DNAp (a member of the B family of polymerases) that a second polymerization state, with a slower, tension-dependent rate, is an obligatory intermediate to switch to proofreading activity (10). In the case of T7 DNAp, we have found no evidence for such a polymerization state. The model for the family B T4 DNAp that was proposed to be applicable to T7 DNAp entails a kinetic scheme that resembles the scheme presented here (14). This model contains a paused state during polymerization and a direct transition from exo to pol. For T7 DNAp, the transition to this incorporation paused intermediate is dependent on tension, in contrast to Manosas et al.'s (14) experiments. This particular difference might arise from the specific assay performed. In the hairpin assay used by Manosas et al. (14), DNA tension destabilizes the dsDNA ahead of DNAp. In contrast, in our assay the tension is directed along the template DNA strand. Although the exact mechanisms of achieving high-fidelity replication might differ slightly for different DNAps, they do seem to respond with remarkable similarities to DNA tension, suggesting a similar mechanistic functioning of these polymerases.

CONCLUSIONS

Taken together, our findings indicate that the removal of errors by DNAp should not be viewed as a precise control mechanism, but rather as a highly stochastic process that is prone to false positives but tuned to avoid false negatives. Our data indicate that at tensions of 15 pN, one out of five events is exonucleolysis, which suggests that the cost of proofreading is the removal of ~10% of correctly incorporated nucleotides. This result suggests that processive exonucleolysis is a robust process that is insensitive to the force on the DNA template. In fact, in experiments using acoustic force spectroscopy (28,29), exonucleolysis takes place at very low force (<4 pN), albeit with an apparent rate that is very low because pauses cannot be separate from activity at these forces (Fig. S9). Fersht et al. (30) showed similar results in a study of the cost of proofreading by *E. coli* DNAp III. The energy difference between a correct and an incorrect nucleotide is in the range of the thermal energy, and to nonetheless achieve the required high fidelity, DNAp needs to correct misincorporated nucleotides frequently. The "paranoid" proofreading behavior of T7 DNAp that we see evidence of might represent a general strategy of high-fidelity molecular machines to deal with thermal noise at the smallest scales of life (31,32).

SUPPORTING MATERIAL

Supporting Materials and Methods, nine figures, and one table are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(17)30038-3.

AUTHOR CONTRIBUTIONS

T.P.H., P.G., and G.J.L.W. designed the experiments. T.P.H., P.G., and S.-N.L. carried out the experiment. T.P.H., M.D., and G.J.L.W. analyzed data and built the model. T.P.H., M.D., S.-N.L., J.C.-D., E.J.G.P., and G.J.L.W. wrote the manuscript. The manuscript was approved by all of the authors. R.T.D., E.J.G.P., and G.J.L.W. supervised the project.

ACKNOWLEDGMENTS

The authors thank Els Kroezinga for biochemical assistance and Douwe Kamsma for assistance for the AFS experiment. This work is part of the research program of the Foundation for Fundamental Research on Matter (FOM) (R.T.D., E.J.G.P., and G.J.L.W.), which is part of the Netherlands Organization for Scientific Research (NWO). We acknowledge support by NWO VICI (R.T.D., E.J.G.P., and G.J.L.W.) as well as a European Research Council (ERC) starting grant (G.J.L.W.).

SUPPORTING CITATIONS

References (7–9,22,28,29) appear in the Supporting Material.

REFERENCES

1. Gregory, T. R. 2005. Genome size evolution in animals. *In* The Evolution of the Genome. T. R. Gregory, editor. Elsevier, Amsterdam, The Netherlands, pp. 3–87.
2. Johnson, K. A. 1993. Conformational coupling in DNA polymerase fidelity. *Annu. Rev. Biochem.* 62:685–713.
3. Patel, S. S., I. Wong, and K. A. Johnson. 1991. Pre-steady-state kinetic analysis of processive DNA replication including complete characterization of an exonuclease-deficient mutant. *Biochemistry.* 30:511–525.
4. Johnson, S. J., and L. S. Beese. 2004. Structures of mismatch replication errors observed in a DNA polymerase. *Cell.* 116:803–816.
5. Kunkel, T. A., and K. Bebenek. 2000. DNA replication fidelity. *Annu. Rev. Biochem.* 69:497–529.
6. Wong, I., S. S. Patel, and K. A. Johnson. 1991. An induced-fit kinetic mechanism for DNA replication fidelity: direct measurement by single-turnover kinetics. *Biochemistry.* 30:526–537.
7. Kuchta, R. D., P. Benkovic, and S. J. Benkovic. 1988. Kinetic mechanism whereby DNA polymerase I (Klenow) replicates DNA with high fidelity. *Biochemistry.* 27:6716–6725.
8. 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.
9. Morin, J. A., F. J. Cao, …, B. Ibarra. 2012. Active DNA unwinding dynamics during processive DNA replication. *Proc. Natl. Acad. Sci. USA.* 109:8115–8120.
10. Ibarra, B., Y. R. Chemla, …, C. Bustamante. 2009. Proofreading dynamics of a processive DNA polymerase. *EMBO J.* 28:2794–2802.
11. Maier, B., D. Bensimon, and V. Croquette. 2000. Replication by a single DNA polymerase of a stretched single-stranded DNA. *Proc. Natl. Acad. Sci. USA.* 97:12002–12007.
12. Datta, K., N. P. Johnson, and P. H. von Hippel. 2010. DNA conformational changes at the primer-template junction regulate the fidelity of replication by DNA polymerase. *Proc. Natl. Acad. Sci. USA.* 107:17980–17985.
13. Christian, T. D., L. J. Romano, and D. Rueda. 2009. Single-molecule measurements of synthesis by DNA polymerase with base-pair resolution. *Proc. Natl. Acad. Sci. USA.* 106:21109–21114.
14. Manosas, M., M. M. Spiering, …, V. Croquette. 2012. Mechanism of strand displacement synthesis by DNA replicative polymerases. *Nucleic Acids Res.* 40:6174–6186.
15. Donlin, M. J., S. S. Patel, and K. A. Johnson. 1991. Kinetic partitioning between the exonuclease and polymerase sites in DNA error correction. *Biochemistry.* 30:538–546.
16. Alberts, B., A. Johnson, …, P. Walter. 2002. Molecular Biology of the Cell, 4th ed. Garland Science, New York.
17. Candelli, A., G. J. L. Wuite, and E. J. G. Peterman. 2011. Combining optical trapping, fluorescence microscopy and micro-fluidics for single molecule studies of DNA-protein interactions. *Phys. Chem. Chem. Phys.* 13:7263–7272.
18. Gross, P., G. Farge, …, G. J. Wuite. 2010. Combining optical tweezers, single-molecule fluorescence microscopy, and microfluidics for studies of DNA-protein interactions. *Methods Enzymol.* 475:427–453.
19. Gross, P., N. Laurens, …, G. J. L. Wuite. 2011. Quantifying how DNA stretches, melts and changes twist under tension. *Nat. Phys.* 7:731–736.
20. Beese, L. S., V. Derbyshire, and T. A. Steitz. 1993. Structure of DNA polymerase I Klenow fragment bound to duplex DNA. *Science.* 260:352–355.
21. Doublié, S., S. Tabor, …, T. Ellenberger. 1998. Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 A resolution. *Nature.* 391:251–258.
22. Santoso, Y., C. M. Joyce, …, A. N. Kapanidis. 2010. Conformational transitions in DNA polymerase I revealed by single-molecule FRET. *Proc. Natl. Acad. Sci. USA.* 107:715–720.
23. Tsai, Y., and K. A. Johnson. 2006. A new paradigm for DNA polymerase specificity. *Biochemistry.* 45:9675–9687.
24. Joyce, C. M., O. Potapova, …, N. D. F. Grindley. 2008. Fingers-closing and other rapid conformational changes in DNA polymerase I (Klenow fragment) and their role in nucleotide selectivity. *Biochemistry.* 47:6103–6116.
25. Koehler, D. R., and P. C. Hanawalt. 1993. Digestion of damaged DNA by the T7 DNA polymerase-exonuclease. *Biochem. J.* 293:451–453.
26. Hamdan, S. M., and A. M. van Oijen. 2010. Timing, coordination, and rhythm: acrobatics at the DNA replication fork. *J. Biol. Chem.* 285:18979–18983.
27. Loparo, J. J., A. W. Kulczyk, …, A. M. van Oijen. 2011. Simultaneous single-molecule measurements of phage T7 replisome composition and function reveal the mechanism of polymerase exchange. *Proc. Natl. Acad. Sci. USA.* 108:3584–3589.
28. Sitters, G., D. Kamsma, …, G. J. L. Wuite. 2015. Acoustic force spectroscopy. *Nat. Methods.* 12:47–50.
29. Kamsma, D., R. Creyghton, …, E. J. G. Peterman. 2016. Tuning the music: acoustic force spectroscopy (AFS) 2.0. *Methods.* 105:26–33.
30. Fersht, A. R., J. W. Knill-Jones, and W. C. Tsui. 1982. Kinetic basis of spontaneous mutation. Misinsertion frequencies, proofreading specificities and cost of proofreading by DNA polymerases of Escherichia coli. *J. Mol. Biol.* 156:37–51.
31. Depken, M., E. A. Galburt, and S. W. Grill. 2009. The origin of short transcriptional pauses. *Biophys. J.* 96:2189–2193.
32. Larson, M. H., J. Zhou, …, S. M. Block. 2012. Trigger loop dynamics mediate the balance between the transcriptional fidelity and speed of RNA polymerase II. *Proc. Natl. Acad. Sci. USA.* 109:6555–6560.

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

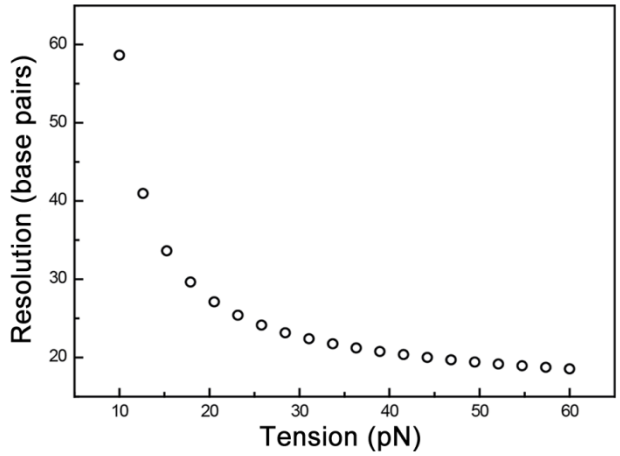

Fig. S1 Resolution in base pairs as a function of DNA tension. The resolution at 45 pN is 20 base pairs. The resolution decreases as the tension decreases because the difference between ds- and ssDNA decreases.

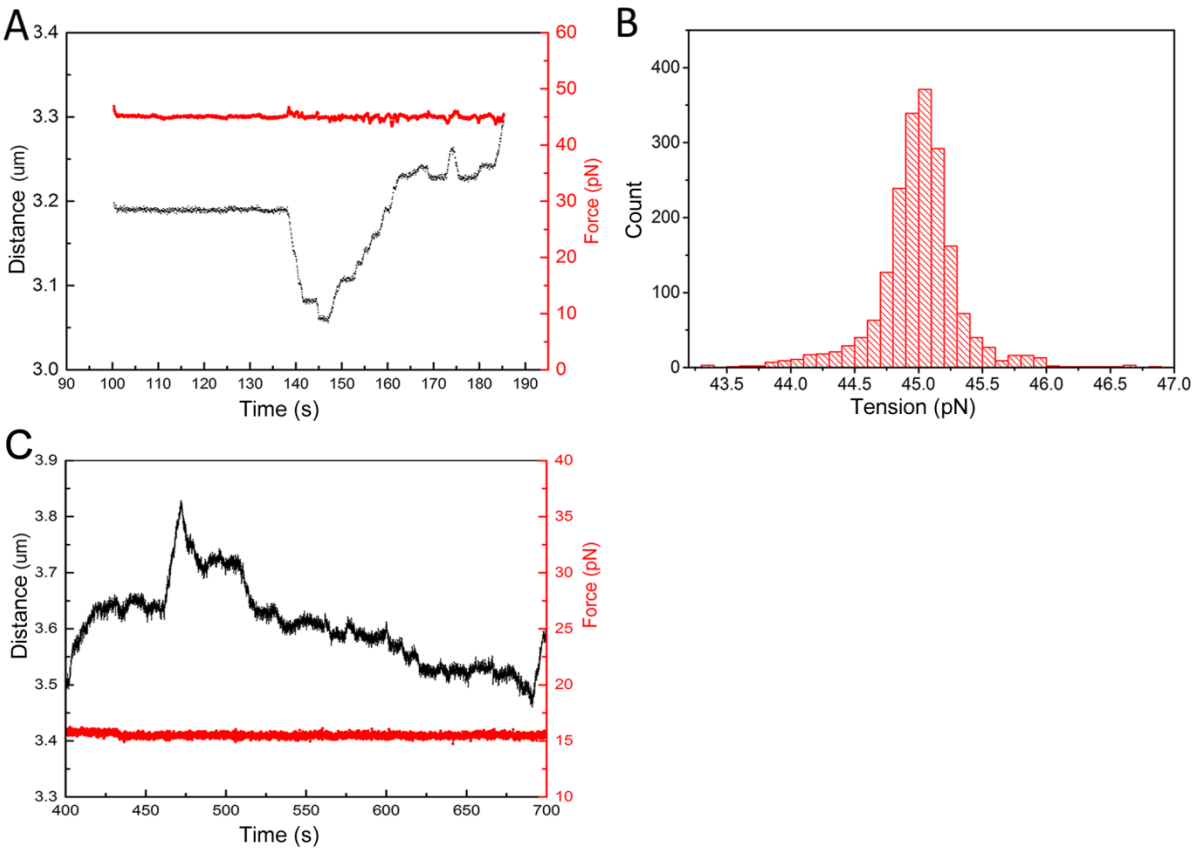

Fig. S2 Corresponding force and distance data for example trace. (A) Force and distance data over time for example trace given in Fig. 1B. (B) Histogram of the force during force clamp in panel A shows that the force is maintained constant with a standard deviation of ~0.4 pN. (C) Force and distance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

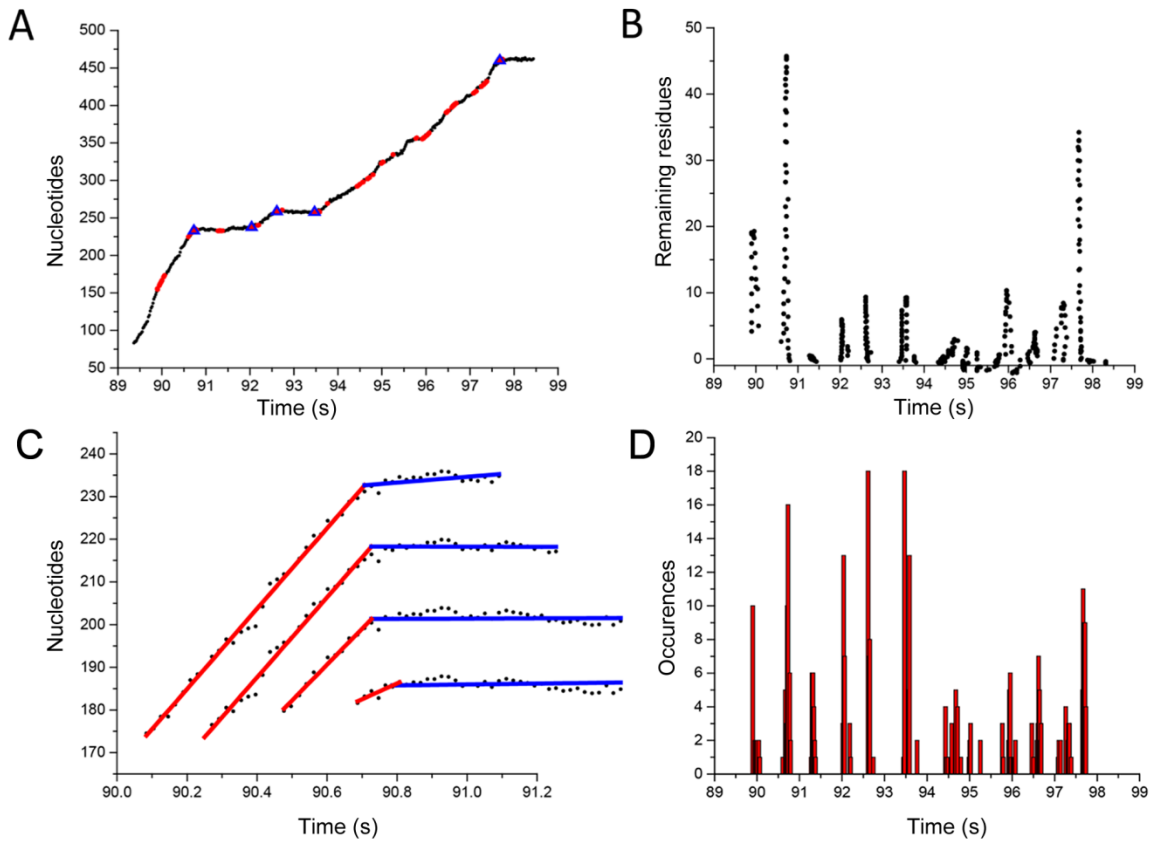

Fig. S**3 Finding breakpoints.** A) In red are all the break points found by the algorithm. B) Remaining residues after subtracting the sum of residues of the two-line fit from the sum of residues from the single linear fit from the data in the 2 sec window. C) Finding of break points in windows (several 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.

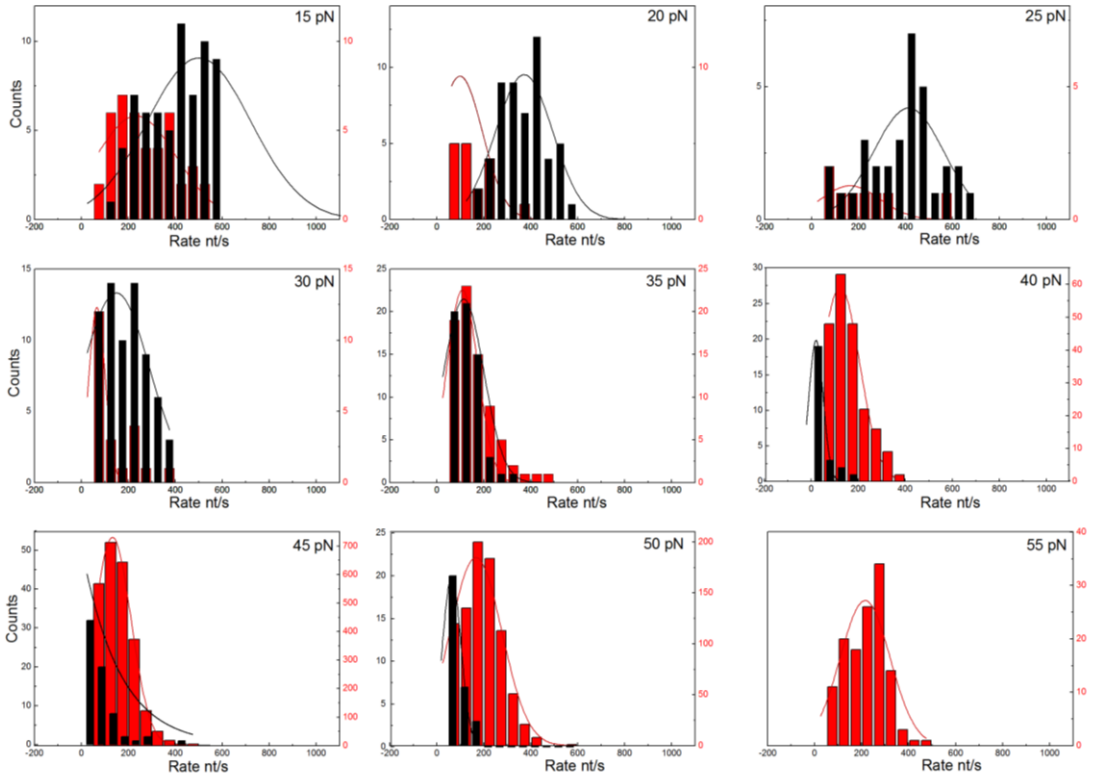

Fig. S4 Histograms of the rates of polymerization (black) and exonucleolysis (red) for different tensions. Data points in graphs: n = 98, 64, 13, 77, 71, 33, 129, 40 and 0 for polymerization in the respective panels. For exonucleolysis n = 42, 25, 8, 28, 88, 219, 2889, 949 and 141 respectively. The arithmetic 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.

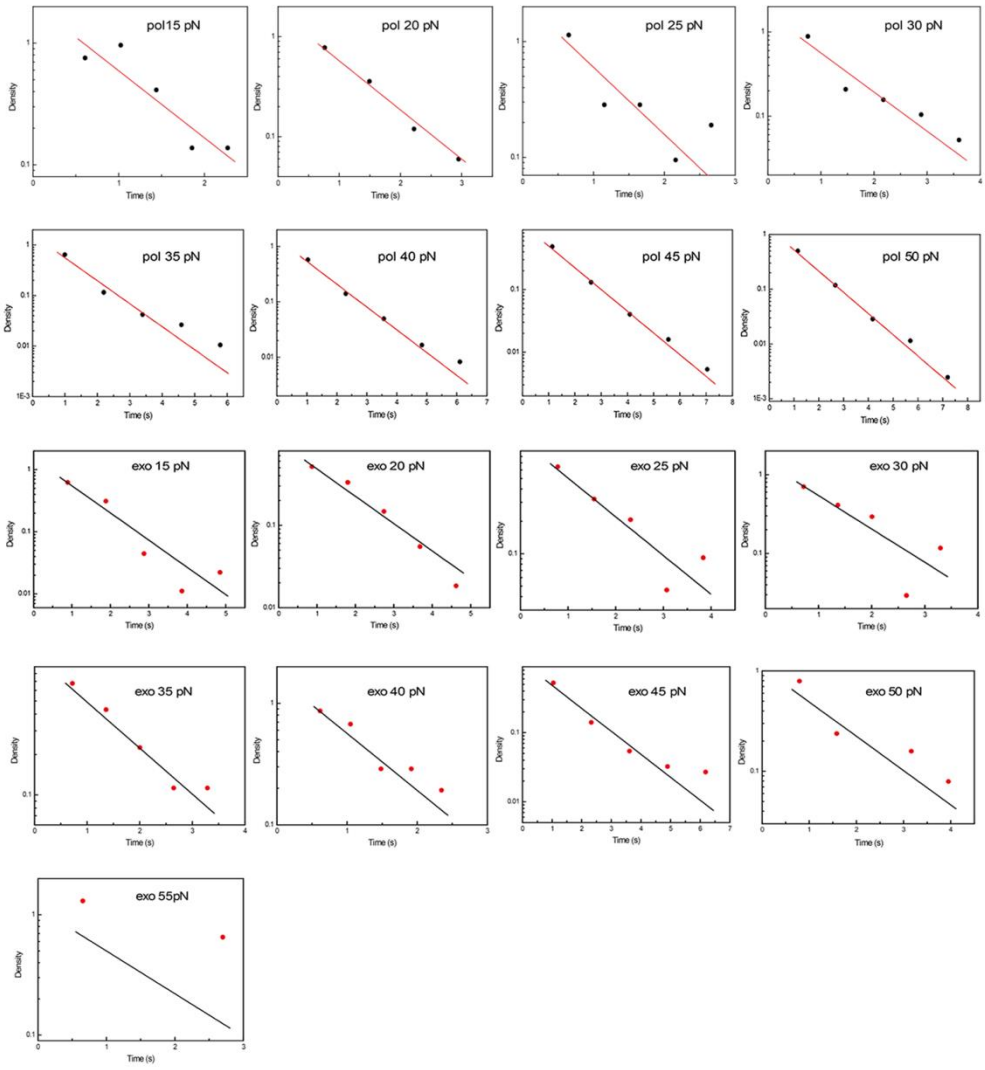

Fig. S5 Off rates of both *pol* and *exo* at different tensions. For all tensions the length of *pol* and *exo* events are described as a single exponential. The durations are log-binned.

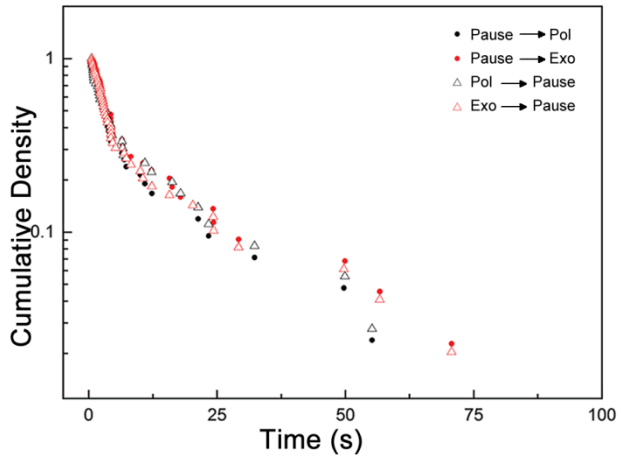

Fig. S6 Distributions of pause durations for different flanking activities show similar behavior for all conditions. Solid black circles: pauses after *pol* activity; black open triangles: pauses resulting in subsequent *pol* activity; solid red circles: pauses after *exo* activity; red open

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

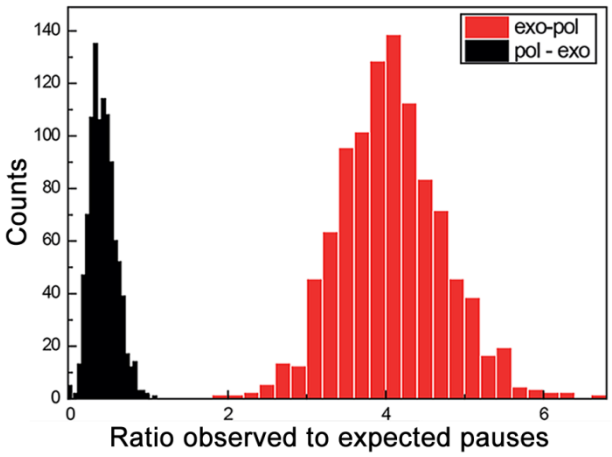

Fig. S7 Ratio of observed sub-resolution pauses to expected sub-resolution pauses for *pol-exo* (black. 8 observed vs. 19 expected) and *exo-pol* (red, 20 expected vs. 84 observed). The histogram is a result of 1000 bootstraps.

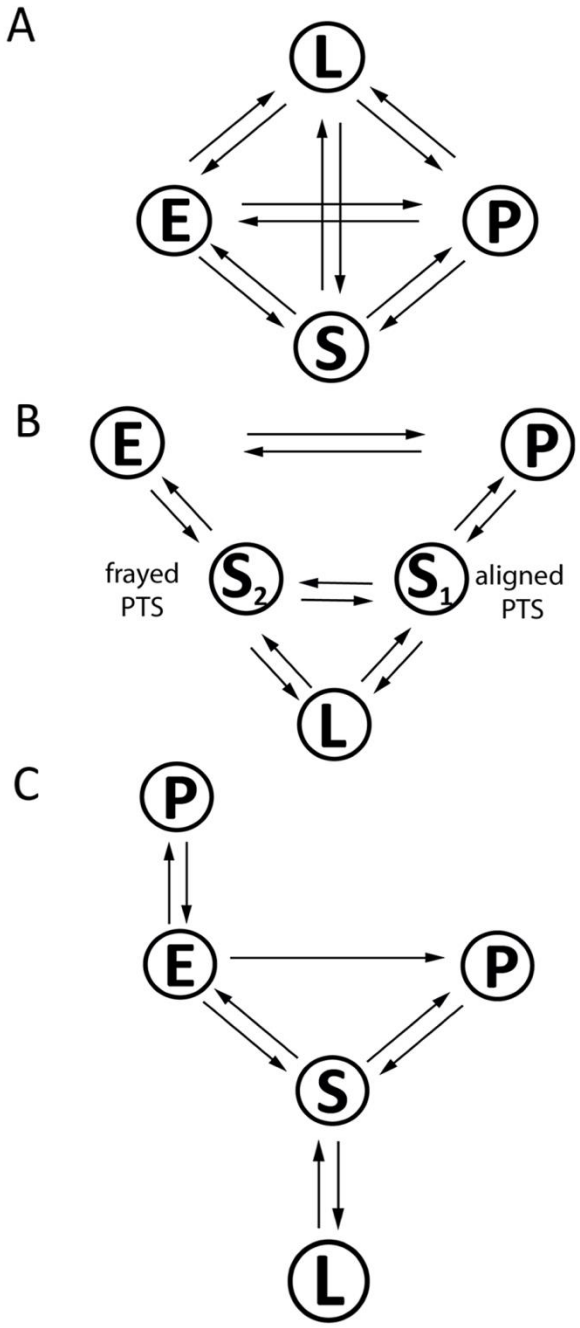

Fig. S8 Alternative models. See supplementary text for discussion.

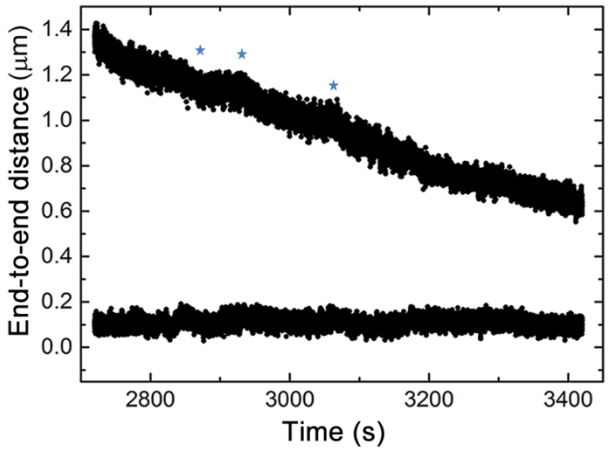

Fig. S9 Exo activity under 4 pN. AFS data was obtained at a constant of 1.1 pN. The top trace is shows *exo* activity interspersed with pauses and bursts. Stars indicate some of the clear bursts of exo activity. The bottom trace is a reference trace of a single-strand DNA tether. No activity is possible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