DNA primase acts as a molecular brake in DNA replication

Lee, Jong-Bong; Hite, Richard K.; Hamdan, Samir M.; Xie, X. Sunney; Richardson, Charles C.; van Oijen, Antonius

Published in:
Nature

DOI:
10.1038/nature04317

DNA substrate

The forked DNA substrate was made by annealing and ligating the following PAGE-purified oligodeoxynucleotides (Integrated DNA Technologies) to λ-phage DNA (New England Biolabs): '*3'Dig*' (5'AGGTCGCCGCCCA$_{12}$3'-digoxigenin), '*5'Biotin*' (biotin-5'A$_{16}$GAGTACTGTACG-ATCTAGCATCAATCACAGGGTCAGGTTCGTTATTGTCCAACTTGCTGTCC), '*Leading tail*' (5'GGGCGGCGACCTGGACAGCAAGTTGGACAATCTCGTTCTATCACTAATTCAC-TAATGCAGGGAGGATTTCAGATATGGCA), and '*Primer*' (5'TGCCATATCTGAAATCCT-CCCTGC).

Characterization of stretching force and bead displacements

The relation between the flow rate through the flow cell and the resultant force exerted on the DNA was established through two independent methods. First, the Brownian motion of the bead in the transverse direction, perpendicular to the flow and parallel to the surface, has a well-understood relation with the stretching force, F. According to the equipartition theorem, the bead's mean-square displacement $<\Delta x^2>$ in this transverse direction is related to F by $F=k_BTl/<\Delta x^2>$, where k_B is the Boltzmann constant, T is the temperature, and l is the length of the DNA.[1] Second, the force-extension profile of dsDNA (main text, Figure 2c) can be described accurately by the Worm-Like Chain model.[2] Fitting our flow-extension data with the known persistence length for dsDNA (53 nm)[2] provides a second, independent force calibration.

The force resolution in our experiment is limited by the small fluctuations in flow rate. At an acquisition rate of 5Hz and stretching force of 3 pN, the traces display noise with a standard deviation of 40 nm. From the slope of the dsDNA force-extension curve at 3 pN we can determine the associated force fluctuation to be 0.1 pN (3%). The small inaccuracy in the applied force is even further reduced in its impact on the conversion factor between ssDNA and dsDNA. The similarity of the slopes of the force-extension curves of ssDNA and dsDNA around 3 pN (0.48 versus 0.43 μm/pN) results in an error in the ssDNA-dsDNA conversion rate of 0.4%.

After determination of the time-dependencies of the lengths of the DNA molecules through particle tracking, the traces were corrected for residual instabilities in the flow by subtracting traces corresponding to tethers that were not enzymatically altered. The length difference between ssDNA and dsDNA at our experimental conditions resulted in a conversion factor between the bead displacement and amount of DNA synthesised of 3.4 nucleotides/nanometer.

Processivity of synthesis by T7 DNA polymerase

The DNA polymerase activity of individual gp5 proteins complexed with the thioredoxin (trx) processivity factor can be visualized by attaching completely single-stranded λ phage DNA molecules between the glass surface and the bead. The ssDNA was prepared by treating the dsDNA *in situ* with λ exonuclease.[3] We provide a primer for the DNA polymerase by annealing a short DNA oligonucleotide (20-mer) to the ssDNA template. The protein associates with the 3' end of the primer and extends it in the presence of dNTPs and Mg^{2+}.

Nucleotide incorporation leads to ssDNA to dsDNA conversion and was observed as a lengthening of the DNA. Short bursts of enzymatic activity are visible, corresponding to repeated cycles of association of the polymerase with the DNA, processive synthesis, and its dissociation from the DNA primer-template (Figure S1). The dwell time between length changes is inversely related to the concentration of protein used, supporting the notion that every burst corresponds to the association and processive enzymatic activity of a single enzyme.

The average enzymatic rate amounts to 220 ± 80 nucleotides per second, with a processivity of 800 ± 300 nucleotides (Fig S2). These values are in agreement with those observed in a previously published single-molecule experiment[4] and ensemble studies.[5, 6]

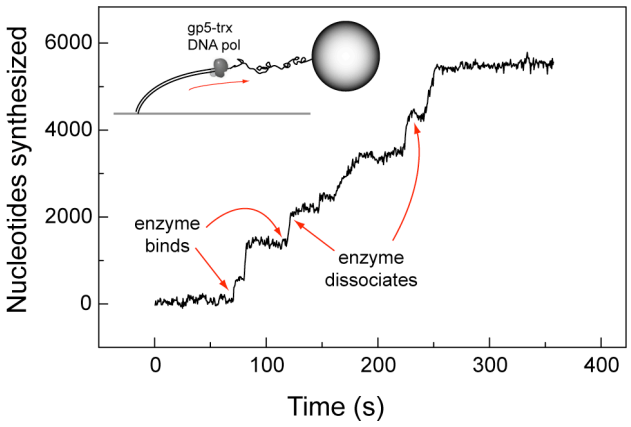

Figure S1

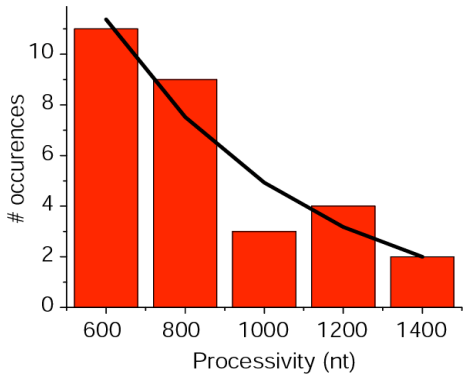

Figure S2

Threshold velocity for pause detection

Pauses in the single-molecule traces were detected as follows: the data was smoothed over a 2-second window and fitted with a series of line segments. A pause was scored when the slope of a line segment was lower than a certain threshold velocity. All line segments obtained from all recorded traces (n=64) are analyzed to determine a threshold velocity for the pause detection. The segment slopes (velocities) are distributed in a gaussian manner (Figure S3) with a mean of 154 ± 9 bp and a standard deviation of 32 ± 3 bp. We chose a threshold velocity of 50 bp/s, which is 3 standard deviations from the mean velocity between pauses, giving rise to a 99% confidence interval of properly assigning pauses. Note that this provides a high level of confidence that a detected pause is indeed a pause instead of a position or rate fluctuation. However, this stringent threshold will result in a fraction of pauses not to be detected (see 'Duration of pauses').

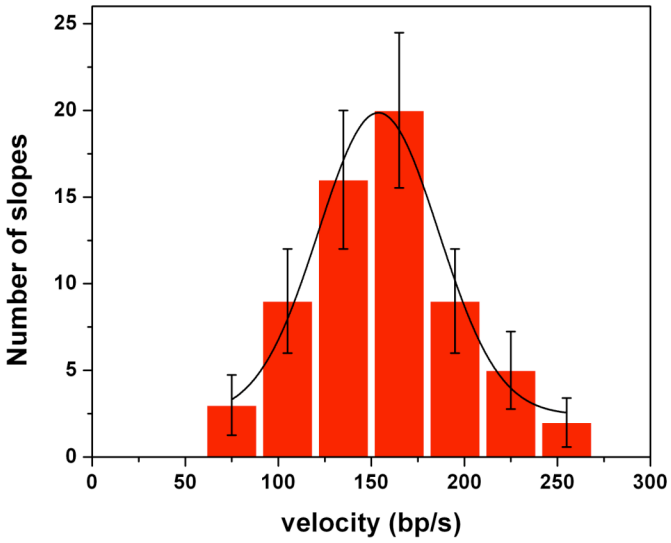

Figure S3

Duration of pauses

The pause duration is characterized by constructing a histogram of all measured pause durations (Figure S4). The stringent threshold that ensures that detected pauses are indeed pauses, as opposed to position or rate fluctuations, causes short pauses (<2 seconds) to be missed. This results in an underestimation of the first data bin. The histogram is fit with an exponential decay

not taking into account the first bin. The pause duration obtained from this fit is 5.6 ± 0.7 s (n=62; at 3 pN stretching force). We measured a pause duration of 5 ± 2 seconds (n=8) at 1 pN. Extrapolating the fit to zero pause length provides us a value of 40% for the fraction of pauses not observed due to their short length.

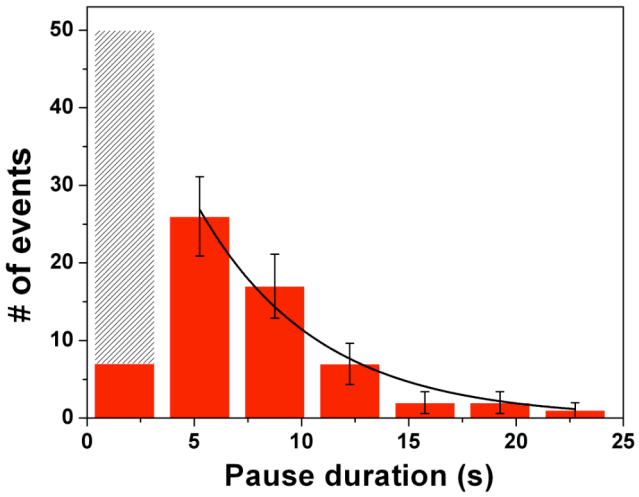

Figure S4

Yield of primase recognition site utilization

The λ-phage DNA substrate used for the single-molecule experiments is constructed such that the T7 gene 4 protein translocates along the strand whose sequence starts with 5'AGGTCGCCGCCCCGT…3' (see also section 'DNA substrate'). There are 34 primase recognition sites (5'-GGGTC-3', 5'-TGGTC-3', 5'-GTGTC-3', 5'-TTGTC-3') in the first 14 kb (the complete λ-phage genome sequence can be found at NCBI (http://www.ncbi.nlm.nih.gov); access code NC_001416). Since the limited processivity of the replication complexes prevented us from obtaining statistically meaningful experimental data beyond the first 14 kb, we restricted our analyses to this region. We determined both the average distance between adjacent sites (obtained from the λ-phage sequence) and the average distance between adjacent pauses (obtained from the experiments) by fitting their histograms with single exponential decays (Figure S5 for distances between sites (n=34) and Figure S6 for distances between pauses (n=107)).

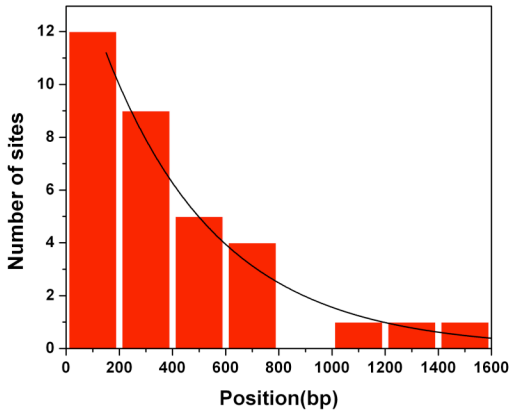

Figure S5

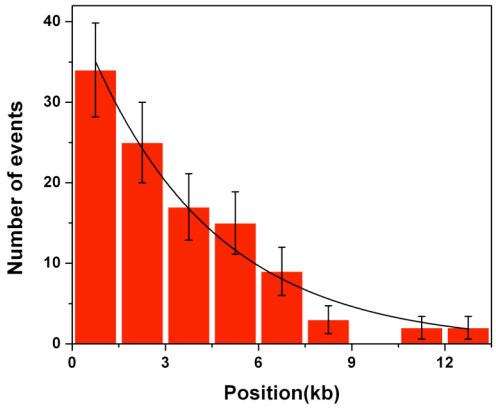

Figure S6

We obtain a distance of 4.1 ± 0.4 kb between observed pauses. Corrected for the fraction of pauses (40%) not observed due to their short length, this average inter-pause distance results in 2.5 ± 0.2 kb. We can determine the yield of site utilization by the primase by calculating the ratio between the average distance between the observed pauses (2.5 ± 0.2 kb) and the average distance between the primase recognition sites (430 ± 70 bp). The data reveal that only 17% of the primase recognition sites are used to synthesise a primer. The resulting average distance between two primase recognition sites utilised, 2.5 ± 0.2 kb, is in agreement with the average length of an Okazaki fragment as observed for T7 replication *in vitro*.[7]

We repeated these experiments at 1 pN (a three-fold reduction in force) and found an average distance of 3 ± 1 kb between observed pauses (compared to 4.1 ± 0.4 kb at 3 pN). Combined with the similarity in pausing duration between the two forces (5.6 ± 0.7 seconds, n=62 at 3 pN *vs*. 5 ± 2 seconds, n=8 at 1 pN), this provides further confidence that the observed primase activity is not adversely affected by the low forces we apply on the DNA.

Correlation between pause positions and recognition sites

Comparison of multiple leading-strand synthesis traces that display primase activity shows that the pauses tend to occur at reproducible positions (Figure S7a). These pause sites correspond to the positions of known primase recognition sites in the λ phage genome (Figure S7a, gray lines; Supplementary Materials). A strong correlation (Pearson's r = 0.63) of the relative frequency of occurrence of the different primase recognition sites in the λ phage genome (Figure

S7b, solid line) with the observed pause probabilities (Figure S7b, red bars) further confirms the notion that the observed pauses correspond to primase activity.

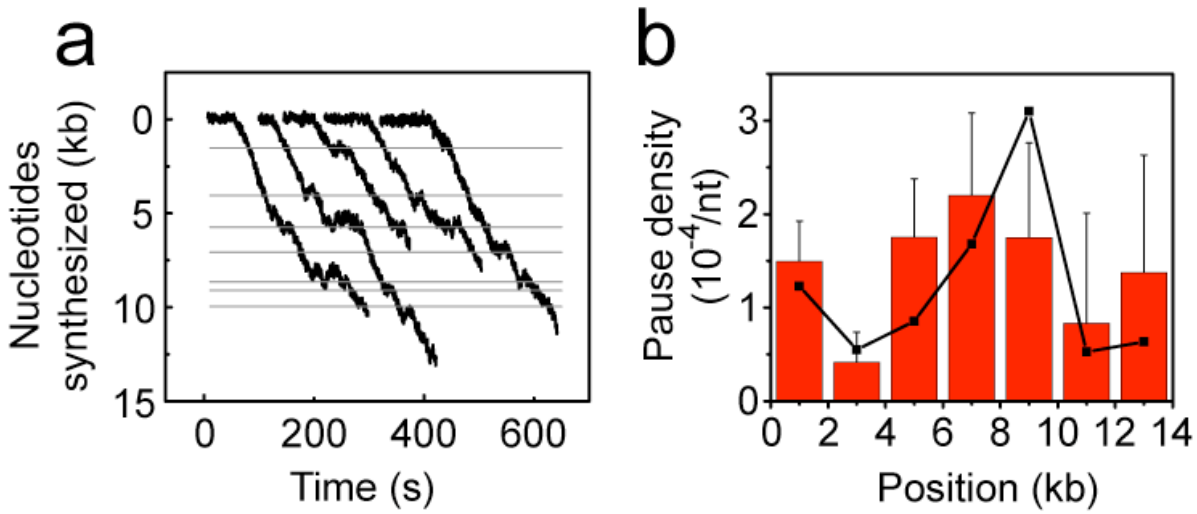

Figure S7

Dependence on ribonucleotide concentration

We performed the single-molecule leading-strand synthesis experiments in the presence of only rATP, rCTP, UTP, or rGTP (Figure S8). As expected, all pausing was abolished in the presence of only UTP or rGTP (300 μM in both cases). The first step of primer synthesis is condensation of the diribonucleotide p-p-p-A-C. For this process to take place, it does not matter whether the rATP binds the active site before the rCTP, or *vice versa*. The affinities of either rATP or rCTP alone for the empty active site is too low (~ 100 μM; Reference 8) to have it stall the primase long enough to be detected in our experiments. Therefore, omission of either rATP or rCTP is expected to prevent the initiation of primer synthesis, consistent with our observations (Figure S8). For comparison we show the observed background pausing density when the leading-strand synthesis experiments were performed leaving out all ribonucleotides from the reaction, or using the ZBD-less gp4 (Figure S8, bar marked with '-').

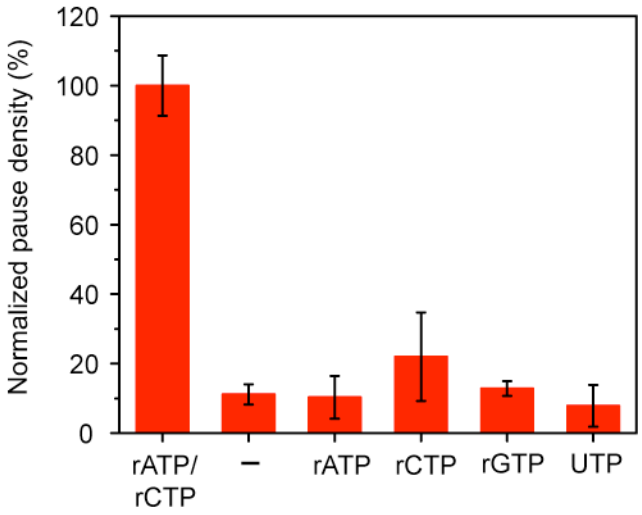

Figure S8

Furthermore, we performed the leading-strand synthesis experiments at four different rATP/rCTP concentrations (100, 300, 550, and 1000 μM). Both the duration and frequency of the pausing did not significantly change when increasing the ribonucleotide concentration (figure S9). This behaviour is consistent with previously reported experiments, where it is shown that the efficiency of lagging-strand synthesis saturates at rATP/rCTP concentrations of as low as 10 μM each.[7]

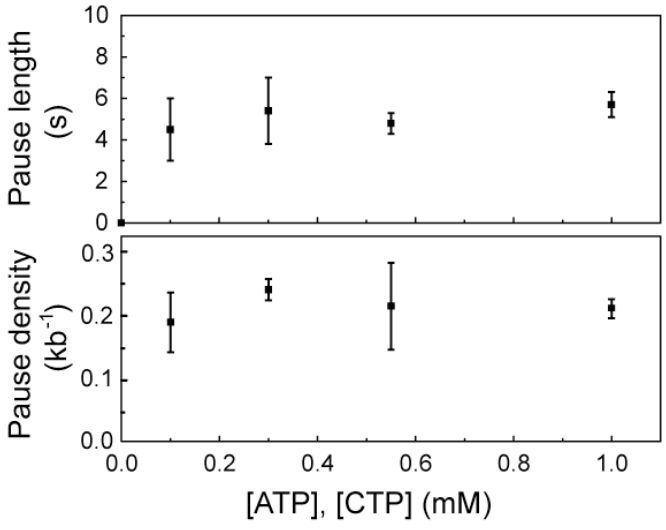

Figure S9

Replication loop formation

The events observed during the various replication reactions can be quantitatively interpreted as follows:

We denote the length of a single base pair dsDNA, as measured under our experimental conditions, as l_{ds}. Similarly, the length of a single nucleotide ssDNA is l_{ss}. Leading-strand synthesis effectively converts dsDNA into ssDNA (see main manuscript for details) and will give rise to the following length change per time unit:

$$k_{pol}\, l_{ss} - k_{pol}\, l_{ds}, \qquad \text{(Equation 1)}$$

where k_{pol} is the number of polymerized nucleotides per time unit.

When both leading- and lagging-strand synthesis occur at equal rates, but no loop is formed, we will have two contributions that combine to a zero length change:

$$(k_{pol}\, l_{ss} - k_{pol}\, l_{ds}) + (-k_{pol}\, l_{ss} + k_{pol}\, l_{ds}) = 0 \qquad \text{(Equation 2)}$$

However, the ssDNA output of the helicase (first term) and the dsDNA output of the lagging-strand DNA polymerase (last term) will be transferred into a loop and will not contribute to the measured length of the DNA. Therefore, the length change during loop formation should be expressed as:

$$-(k_{pol}\, l_{ds} + k_{pol}\, l_{ss}) \qquad \text{(Equation 3)}$$

Comparison of the rate of length change during loop formation (Equation 3) to the rate of change during leading-strand synthesis (Equation 1) shows that rate of change during loop formation should be $(l_{ds}+l_{ss})/(l_{ds}-l_{ss})$ times higher than during leading-strand synthesis. We measured an average ratio of synthesis rates during and before loop formation of 1.3 ± 0.2 (n=5). This is consistent with the expected ratio of 1.2, based on l_{ds}=0.32 nm/nt and l_{ss}=0.03 nm/nt under our experimental conditions.

Upon replication loop release, the ssDNA and dsDNA regions in the loop will contribute again to the measured length of the DNA. According to Equation 2, the ratio of the length decrease during loop formation and increase during release should be equal to one. The measured ratio of 0.94 ± 0.1 (n=5) is consistent with this prediction. Furthermore, there is a strong correlation between the length decrease during loop formation and the length increase during loop release

(Pearson's correlation coefficient r=0.92).

These analyses tell us that during loop formation an equal amount of DNA is synthesized by the lagging-strand DNA polymerase as is produced by leading-strand synthesis. This symmetry in polymerization rates underlines the necessity of a transient halting of the fork to prevent slow primer synthesis and delivery causing an uncoupling of the replication reaction.

References

1. Strick, T.R., Allemand, J.F., Bensimon, D. & Croquette, V. Behavior of supercoiled DNA. *Biophys J* **74**, 2016-28 (1998).
2. Bustamante, C., Marko, J.F., Siggia, E.D. & Smith, S. Entropic elasticity of lambda-phage DNA. *Science* (Washington, DC, United States) **265**, 1599-600 (1994).
3. van Oijen, A. M. *et al*. Single-molecule kinetics of lambda exonuclease reveal base dependence and dynamic disorder. *Science* (Washington, DC, United States) **301**, 1235-9 (2003).
4. Wuite, G. J. L., Smith, S. B., Young, M., Keller, D. & Bustamante, C. Single-molecule studies of the effect of template tension on T7 DNA polymerase activity. *Nature* (London) **404**, 103-6 (2000).
5. Patel, S. S., Wong, I. & Johnson, K. A. Pre-steady-state kinetic analysis of processive DNA replication including complete characterization of an exonuclease-deficient mutant. *Biochemistry* **30**, 511-25 (1991).
6. Tabor, S., Huber, H. E. & Richardson, C. C. Escherichia coli thioredoxin confers processivity on the DNA polymerase activity of the gene 5 protein of bacteriophage T7. *J Biol Chem* **262**, 16212-23 (1987).
7. Lee, J., Chastain, P. D., II, Griffith, J. D. & Richardson, C. C. Lagging strand synthesis in coordinated DNA synthesis by bacteriophage T7 replication proteins. *J Mol Biol* **316**, 19-34 (2002).
8. Frick, D.N., Kumar S. & Richardson, C.C. Interaction of ribonucleoside triphosphates with the gene 4 primase of bacteriophage T7. *J Biol Chem* **274**, 35899-907 (1999).
9. Thompson, R. E., Larson, D. R. & Webb, W. W. Precise nanometer localization analysis for individual fluorescent probes. *Biophys J* **82**, 2775-83 (2002).