

Single molecule measurement of the “speed limit” of DNA polymerase

Jerrold J. Schwartz and Stephen R. Quake¹

Department of Bioengineering, Stanford University and Howard Hughes Medical Institute, Stanford, CA 94305

Edited by William E. Moerner, Stanford University, Stanford, CA, and approved September 24, 2009 (received for review July 8, 2009)

Although DNA replication is often imagined as a regular and continuous process, the DNA polymerase enzyme is a complicated machine and can pause upon encountering physical and chemical barriers. We used single molecule measurements to make a detailed characterization of this behavior as a function of the template's secondary structure and the sequence context. Strand displacement replication through a DNA hairpin by single DNA polymerase molecules was measured in real time with near single base resolution and physiological concentrations of nucleotides. These data enabled the measurement of the intrinsic “speed limit” of DNA polymerase by separating the burst synthesis rate from pausing events. The strand displacement burst synthesis rate for *Escherichia coli* DNA Polymerase I (KF) was found to be an order of magnitude faster than the reported bulk strand displacement rate, a discrepancy that can be accounted for by to sequence specific pausing. The ability to follow trajectories of single molecules revealed that the burst synthesis rate is also highly stochastic and varies up to 50-fold from molecule to molecule. Surprisingly, our results allow a unified explanation of strand displacement and single strand primer extension synthesis rates.

FRET | pausing | DNA replication | DNA hairpin

Fast and accurate DNA replication is required to ensure the faithful transfer of genetic information to daughter cells. Replication forks can be slowed or paused by encountering DNA-binding ligands (1), genomic tRNA gene sites (2), stalled ternary complexes of RNA polymerase (3), and DNA-bound protein (4–5). DNA polymerase can also pause due to specific DNA sequences, including palindromic DNA capable of forming hairpin secondary structure (6–8), slow zones (9), and trinucleotide repeats of (CGG)_n/(CCG)_n or (CTG)_n/(CAG)_n (10–11). While these factors are thought to prevent fork movement along the template by steric hindrance and may be regulated by the replisome itself, other factors, including temperature, contaminants, nucleotide analogs (12), template tension (13–14), and nonclassic pause sites such as Pyr-G-C (15), may also interfere with steps in the DNA polymerase reaction pathway.

To measure the intrinsic “speed limit” of DNA polymerase itself, there must be sufficient spatial and temporal resolution to separate pausing events from burst synthesis. Although the kinetics of the Klenow fragment of *Escherichia coli* DNA Polymerase I (Pol I(KF)) have been studied using a variety of approaches (13, 15–20), there are conflicting reports of the synthesis rate and it has not yet been possible to dissect the interplay between pausing and burst synthesis. Rapid quenching and stopped-flow bulk techniques have determined that during primer extension synthesis, the first nucleotide is incorporated by Pol I(KF) with a rate ranging from 40–87 nt sec⁻¹ (16–17, 19) while subsequent nucleotides are incorporated at a slower rate of 13–15 nt sec⁻¹ (17, 20). This bulk processive rate is twice as fast as that found by single molecule primer extension synthesis rate measurements (7 nt sec⁻¹) (13). Primer extension synthesis is generally thought to be faster than strand-displacement synthesis, and this idea is supported by a bulk rapid quenching study which found that the strand displacement synthesis rate of Pol I(KF) was even slower at 1.2 nt sec⁻¹ (20).

We developed a FRET-based approach (21) to study single polymerase molecules undergoing strand displacement replication through a DNA hairpin to further investigate this conundrum between the various reported rates. This assay allowed us to measure polymerase activity in real time at saturating nucleotide concentrations with near single base resolution, thereby enabling separation of burst synthesis rates from pausing effects. Our results show that the only literature strand displacement synthesis rate for Pol I(KF) (20) is not the true burst synthesis rate because it likely includes sequence specific pausing effects due to the template used in the experiments. These pausing events are not unique to Pol I(KF) and were also observed with ϕ 29 DNA polymerase. The existence of the pausing effect is known from bulk studies (15), but without the characterization of its heterogeneity and kinetics enabled by the single molecule experiments described here, its central role in determining the speed limit of DNA polymerase was not previously appreciated. The burst synthesis strand displacement rates obtained from our single molecule measurements were also highly heterogeneous from molecule to molecule, with rates varying from 1 nt sec⁻¹ to over 50 nt sec⁻¹. This variation in rates was previously hidden due to the ensemble averaging of bulk measurements and could only be accessed with a single molecule approach. Surprisingly, our characterization of the pausing heterogeneity in strand displacement also allows us to explain the discrepancies in rates found in primer extension experiments and shows that strand displacement rates are not slower than primer extension rates.

Results and Discussion

A 259-nt single-stranded DNA molecule was designed and synthesized containing an internal double-stranded 33-bp hairpin flanked by 94-nt single-stranded tails, one of which had a 3'-biotin group (Fig. 1A). The 3' base of the hairpin contained an internal Cy3 FRET donor and the 5' base contained an internal nonfluorescent FRET acceptor (Black Hole Quencher-2, BHQ-2) so that when the hairpin is fully folded, quenching of Cy3 by BHQ-2 prevents any fluorescence emission (Fig. 1B). The length of the hairpin was chosen to make full use of the dynamic range for this FRET pair: when the last base pair of the hairpin is broken due to strand displacement replication, the distance between fluorophores is just over twice the Förster radius ($R_0 = 5.02$ nm, 33 bp of dsDNA = 11.2 nm). An 80-nt primer was annealed to the 3' end of the template and the DNA constructs were attached to the surface of a glass coverslip via a 3' biotin-streptavidin linker. The coverslip was placed on an inverted Nikon TE-2000S TIRF microscope and images were collected with a 60 × 1.45 NA objective at 5 Hz with a Photometrics Cascade II EMCCD. Each field of view contained,

Author contributions: J.J.S. and S.R.Q. designed research; J.J.S. performed research; J.J.S. and S.R.Q. analyzed data; and J.J.S. and S.R.Q. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

¹To whom correspondence should be addressed. E-mail: quake@stanford.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0907404106/DCSupplemental.

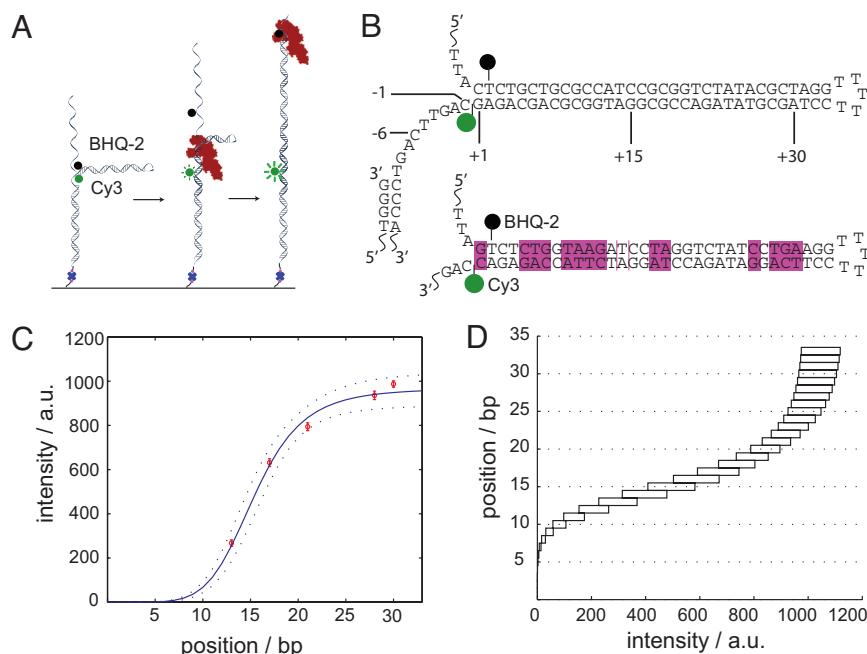


Fig. 1. FRET can be used to observe the real-time position of single DNA polymerase molecules. (A) Cartoon of the DNA template used to measure real-time polymerase kinetics. Primed 259 bp DNA molecules containing internal 33-bp hairpins and flanking 94 bp tails were immobilized on a glass surface through a biotin-streptavidin linker. DNA replication through the hairpin resulted in a reduction in FRET efficiency between Cy3 and BHQ-2, giving rise to an increase in Cy3 fluorescence. (B) Sequence and structure of the two DNA templates used in these experiments. The pink boxes highlight the differences in sequence between the sample (Top) and the control (Bottom) hairpins. Sequence numbering refers to the template position awaiting nucleotide addition. (C) Cy3 fluorescence recovery data for primer extension calibration experiments. The position along the x axis refers to the polymerase position along the stem of the hairpin and the error bars are the SEM. Dashed lines are 95% confidence intervals. (D) Intensity-to-position curve based on (C) to convert an arbitrary intensity to a template position. Each rectangle represents the 95% confidence intervals for the expected intensity at each polymerase position along the template.

on average, over 1,000 individually resolvable single molecules (for full experimental details, see *Materials and Methods*).

To correlate the FRET signal with template position, a series of primer extension experiments were performed by adding DNA polymerase and only 3 of the 4 deoxynucleotide triphosphates (dNTPs). After the first extension and imaging step, the flow cell was thoroughly washed to remove any unincorporated nucleotides, and the process was repeated multiple times with a different set of 3 dNTPs. This allowed us to extend the primer to known positions on the template and observe the corresponding reduction in FRET signal as the donor and acceptor were forced apart. Before extension begins, the distance R between the 2 dyes is small and the FRET efficiency is $E_{\text{FRET}} = 1/[1+(R/R_0)^6] \approx 1$. As the primer extends and the hairpin unwinds, R increases and E_{FRET} decreases, leading to recovery of Cy3 signal. The observed Cy3 intensity is thus given by $I_{\text{CY3}} = I_{\text{MAX}}(1 - E_{\text{FRET}}) = I_{\text{MAX}}[1 - (1/[1+(R/R_0)^6])]$, where I_{MAX} is the maximum Cy3 intensity achievable when the dyes are as far apart as possible (assuming 0.34 nm per base pair, the maximum dsDNA separation is $2 \times 33\text{bp} + 5\text{bp loop} = 71\text{ bp} \approx 24\text{ nm}$). A weighted least-squares fit of the data to this equation gave $I_{\text{MAX}} = 987\text{ a.u.}$ and $R_0 = 15.3\text{ bp} = 5.2\text{ nm}$, which is consistent with the vendor's reported Förster radius for Cy3 and BHQ-2. There is some width to the distributions of Cy3 intensities at each position due to signal to noise variation and Cy3-BHQ-2 orientation effects (supporting information (SI) Fig. S1 and Table S1). However, by performing this alignment procedure on an ensemble of records at each position, a calibration curve relating the observed fluorescent intensity and the polymerase position along the template was created (Fig. 1C). Using this calibration curve and its 95% confidence limits, any arbitrary fluorescent intensity could be converted into a corresponding polymerase location along the template (Fig. 1D).

The real-time kinetics of DNA replication were examined in the presence of all four dNTPs with two different DNA polymerases: the Klenow fragment of DNA Polymerase I (exo^-) from *E. coli* (Pol I(KF)) and the replicative polymerase from the *Bacillus subtilis* bacteriophage $\phi 29$. The trajectories for both polymerases exhibited heterogeneous behaviors that could be classified into 4 categories: fast replication without pausing (Fig. 2A), fast replication with a single pause (Fig. 2B), fast replication with multiple pauses (Fig. 2C), and slow replication. Slow replication occurred in less than 6% of the traces, and while its origin is unclear, it could stem from the template sticking to the surface or unfavorable polymerase-surface interactions. Approximately 66% of the DNA polymerases extended the entire template without a pause while about 28% of the polymerases paused at least once. The pauses occurred at highly stereotyped positions and we took advantage of the single molecule approach to measure both the distribution between pausing and nonpausing polymerase molecules and the relationship between the pauses and the burst synthesis speeds of the polymerase, data that would be impossible to observe with conventional bulk techniques.

The positional accuracy of individual trajectories depended on both template location and signal-to-noise with a maximum attainable resolution of approximately 2 bp. For example, a well-behaved trajectory that paused with a mean intensity of $562 \pm 47\text{ a.u.}$ (SD) was called at +15 or +16 bp; a noisier trajectory that paused with a mean intensity of $447 \pm 168\text{ a.u.}$ (SD) was called at $+14 \pm 2\text{ bp}$. By aligning an ensemble of trajectories for Pol I(KF) and $\phi 29$, polymerase pause positions were localized with near single base accuracy over the template from +8 bp to +18 bp (Fig. 3A and B). Outside of this window the overlapping confidence limits of the calibration curve prevented pause localization accuracy better than $\pm 3\text{ bp}$. For Pol

our results described above, we believe this change is more likely the replication transition than a dye-enzyme effect.

There was a surprising amount of heterogeneity in the strand displacement synthesis rates for both Pol I(KF) and $\phi 29$ from molecule to molecule over identical sequences (Fig. 5), with values ranging from 1 nt sec^{-1} to over 50 nt sec^{-1} . The stochasticity of the synthesis rates is hidden in time-averaged or bulk rates and can only be revealed with a single molecule approach. The source of this heterogeneity could be of great interest for future polymerase studies and has not received significant attention in the literature to date.

One potential drawback of a quencher-based approach is that blinking fluorophores might incorrectly be interpreted as real extension events. If a Cy3-Cy5 FRET pair were used, this would certainly be a concern as Cy5 is known to exhibit frequent blinking (30–31). In this FRET system, a false positive extension could be caused by either BHQ-2's quenching ability switching off before extension begins or a dark Cy3 fluorophore switching on after extension has finished. BHQ-2 blinking was not significantly observed while imaging folded hairpins in the absence of polymerase or dNTPs and has not been reported in the literature. Cy3 fluorophores have a dark-state lifetime of approximately 2 sec and the on-blink and off-blink rates have been reported to be 0.36 and 0.26 times per molecule over 120 sec, respectively (32). Given the infrequent blinking and short off times of Cy3, the probability of it being off during any real extension is very low. It is therefore not surprising that only approximately 1% of the trajectories exhibited identifiable blinking after full extension (signal going from on to off and then back to on). Importantly, for the kinetics experiments we observed intermediate intensity pauses in approximately 50% of the trajectories, a phenomenon that cannot be produced by fluorophore blinking. The calibration experiments (Fig. 1C) with polymerases stalled on the template did not show any significant deviation from the expected FRET behavior for a single acceptor system, so protein-induced quenching was not a significant source of error.

The real-time single molecule observation of Pol I(KF) revealed that the polymerase's intrinsic strand displacement "speed limit" is an order of magnitude faster than reported by previous bulk processive synthesis measurements. Replication was found to be interrupted by heterogeneous sequence-dependent pauses that were independent of template secondary structure, and knowledge of these pause frequencies and lifetimes would otherwise be impossible to obtain from ensemble studies. Importantly, the pause kinetics can account for the discrepancy between our time-averaged Pol I(KF) single molecule strand displacement rate and the reported bulk strand displacement rate. The pauses were highly localized to a short Pyr-G-C sequence motif and were relieved with the addition of heat or betaine, suggesting that they may be associated with a structural template rearrangement step during the polymerase reaction pathway. Pausing at Pyr-G-C motifs is thought to be fairly ubiquitous but the pause strength can be dependent on the surrounding sequence and other environmental factors. The assay described here could be used to explore how pause strength varies with surrounding sequence or to explore the pausing of other sequence motifs. Our single molecule assay also revealed up to 50-fold variation in the synthesis rates between polymerase molecules replicating identical templates. Future single molecule experiments should help fully elucidate the source of this rate stochasticity, the mechanism of DNA polymerase sequence-dependent pausing, and provide insight into how other endogenous and exogenous factors act to slow or stall replication forks.

Materials and Methods

DNA Hairpin Construction. DNA hairpins were constructed through the stepwise enzymatic ligation of 4 separate synthetic oligonucleotides (Integrated DNA Technologies and Operon). In the first step, oligo A (5'-Phos-TCA TAG

CCA GAT GCC CAG AGA TTA GAG CGC ATG ACA AGT AAA GGA CGG TT-Biotin-3') and oligo B (5'-Phos-CGG ATG GCG CAG CAG AG[Cy3]C AGT TCA GTC CCA CCG AGG TTT GGT CAG TT-3') were annealed to the complementary oligo Z (5'-AAC CGT CCT TTA CTT GTC ATG CGC TCT AAT CTC TGG GCA TCT GGC TAT GAT GTT GAT GGA ACT GAC CAA ACG TCG GTG GG-3') in 20 mM Tris, 100 mM NaCl, pH 8.0 by heating to 95 °C for 5 min and slowly cooling at 0.1 °C/s to 4 °C. 20 units of *E. coli* DNA ligase and 1X *E. coli* DNA ligase reaction buffer (New England Biolabs) were then added and the reaction was held at 16 °C for 3 h, after which the ligase was inactivated by holding the reaction at 65 °C for 20 min. Oligo C (5'-Phos-AGT GAC GCC AAC GCA ATT AC[BHQ2-dT] CTG CTG CGC CAT CCG CGG TCT ATA CGC TAG GTT TTT CCT AGC GTA TAG ACC G-3'), oligo D (5'-GCC CTG AGA GAG TTG CAG CAA GCG GTC CAC GCT GGT TTG CCC CAG CAG GCG AAA ATC CTG TTT GAT GGT GGT TCC GAA AT-3'), and oligo Y (5'-TGC GTT GGC GTC ACT ATT TCG GAA CCA CCA TCA AAC AGG ATT TTC GCC TGC TGG GGC AAA CCA GCG TGG ACC GCT TGC TGC AAC TCT CTC AGG GC-3') were added to the solution along with 100 units of Ampligase DNA ligase in 1X Ampligase Reaction Buffer (Epicentre Biotechnologies). The solution was then heated to 95 °C for 5 min to melt all of the strands, rapidly cooled to 89 °C, slowly cooled at 0.1 °C/s to 65 °C, held at 65 °C for 2 min, and then reheated to 95 °C for 1 min. This cooling and reheating process was repeated 20 times to ensure high stringency ligation. The DNA ligation mixture was then loaded on a 15% TBE-urea denaturing polyacrylamide gel (Invitrogen) and run at 175V for 1 h in a 65 °C water bath. The gel was stained with SYBR Green I (Invitrogen) for 20 min, the band corresponding to the 259 bp DNA product was excised from the gel, and the DNA was purified as described in Invitrogen's GeneTrapper manual. The control sequence hairpin was synthesized as above but with oligo B replaced with (5'-Phos-AGG ATC TTA CCA GAG AC/ICy3/CAGT TCA GTC CCA CCG ACG TTT GGT CAG TTC CAT CAA CA-3') and oligo C replaced with (5'-AGT GAC GCC AAC GCA ATT AC[BHQ2-T] CTC TGG TAA GAT CCT AGG TCT ATC CTG AAG GTT TTT CCT TCA GGA TAG ACC T-3').

Single Molecule Kinetics Measurement. Glass surfaces (Precision Glass & Optics, D-263T cut glass, 0.145 mm, 2"x1" 40/20 surface quality) were RCA cleaned and a plastic hybriwell (Grace Biolabs) was placed on the surface to create a fluidic chamber. The surface was incubated with 0.2 mg/ml biotinylated-BSA in 1X PBS for 20 min, washed extensively with 1X PBS, and then incubated with 0.5 mg/ml neutravidin in 1X PBS for 20 min. Purified DNA hairpin was annealed with a 5' biotinylated primer (oligo Z) as described above, diluted to 10 pM in 1X PBS, and incubated on the surface for 20 min. Surfaces were imaged on an inverted Nikon TE2000S microscope with a 60 × 1.45 NA PlanApo TIR objective, a low-fluorescence immersion oil ($n = 1.515$), and an automated precision stepper motor stage (Mad City Labs). A Nikon TIRF attachment in conjunction with a 532 nm diode-pumped solid-state laser (MeshTel) and a custom laser-launch system (ThorLabs) was used for illumination. A Cy3 filter set (HQ535/50, Q565LP, HQ610/75, Chroma) and a Photometrics Cascade II EMCCD were used for filtering and image acquisition. Immediately before enzyme addition, surfaces were imaged to ensure the hairpin was properly folded as evidenced by the absence of Cy3 emission. For the premature chain termination reactions, an enzyme mixture consisting of 1X polymerase reaction buffer (New England Biolabs), 100 μM dATP, dTTP, dCTP, 10 units of DNA polymerase (Klenow exo^- , New England Biolabs), 0.1 mg/ml glucose oxidase, 0.2 mg/ml catalase, 10% wt/wt glucose, and 1 mM Trolox was flowed into the chamber. After 5 min of incubation at room temperature, the surface was washed extensively with 1X PBS and 20 different fields of view were imaged. Fluorescent intensities at this stage represented a polymerase position at -6 bp with respect to the hairpin (i.e., the polymerase active site is 6 bp away from the first bp of the hairpin stem). This process was repeated 6 times with the dNTP mixes shown in Table S1.

For the real-time kinetics experiments, surfaces were prepared as described above. An enzyme mixture consisting of 1X polymerase reaction buffer (New England Biolabs), 100 μM dNTPs, 10 units of DNA polymerase (Klenow exo^- or $\phi 29$, New England Biolabs), 0.1 mg/ml glucose oxidase, 0.2 mg/ml catalase, 10% wt/wt glucose, and 200 μM Trolox was flowed into the chamber. Images were immediately acquired every 200 ms for approximately 20 min. Freshly prepared 1 M betaine was added to the above mixture for the betaine experiments. For the real-time temperature experiments, a commercial flow cell with internal heating elements (CFCS2, Biopetechs) was used in conjunction with an external temperature controller to hold the sample temperature at 37 °C.

Image Analysis. Custom software was written in MATLAB to automatically identify and track the XY coordinates of single molecules using a previously published particle tracking algorithm (33). A 3 × 3 pixel grid surrounding each molecule was integrated and the 5 × 5 pixel perimeter was used as to estimate the background. Raw and smoothed net intensity counts for each trajectory were then plotted and manually analyzed. Steps were identified

