

# Supporting Information

Righini et al. 10.1073/pnas.1719906115

## SI Materials and Methods

All DNA-modifying enzymes were purchased from New England Biolabs. Oligonucleotides were purchased from IDT. Nucleotide triphosphates were purchased from Thermo Scientific, and standard salts and buffer components were purchased from Sigma Aldrich.

**Optical Trapping Instruments.** A split-path optical tweezers instrument (1, 2) and a time-shared instrument (3) were assembled essentially as described elsewhere. Briefly, an Nd:YAG 1,064-nm laser was used in both instruments. In the split-path instrument, the laser was split using a polarizing beam-splitter into two paths. In each path, the laser was reflected off a mirror, one of which was a piezo-actuated mirror to allow for steering. The two beams were later recombined in a second polarizing beam-splitter, imaged into an objective, and focused into homemade fluidics chambers. Detection of bead positions and forces was achieved by collecting the light on a second objective, splitting the beams again using a polarizing beam-splitter, and imaging the beams on quadrant photodiodes (QPDs).

In the time-shared instrument, the traps were generated by an AOD, with the frequency of the AOD controlled by a custom-made radio frequency board, and the position of the traps switched every 5  $\mu$ s. In this case, detection of the bead positions in both traps was achieved on the same QPD.

**Polystyrene Beads.** For power spectra measurements, we used 0.81  $\mu$ m-diameter streptavidin-coated polystyrene beads from Spherotech and 1  $\mu$ m-diameter carboxylated beads coated with anti-digoxigenin antibody (Roche, cat. no. 11333089001), as follows:

We washed 10  $\mu$ L of 10% bead suspension with coupling buffer (Mes 0.1 M, pH 4.7, 150 mM NaCl) four times, with centrifugation steps (5 min at 4,500 g) between the washes. The beads were dispersed in 500  $\mu$ L coupling buffer. We added 10  $\mu$ L of 3.5 mM 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC), followed by 65  $\mu$ L of 0.2 mg/mL of antibody solution. Reaction proceeded overnight at 4  $^{\circ}$ C. At this point, 10  $\mu$ L of 1 M glycine was added, and Tween 20 was added to 0.05%. The beads were centrifuged and washed five times with storage buffer (20 mM Tris, 8, 130 mM KCl, 0.05% Tween 20, 1 mM EDTA, 5 mM sodium azide) with 3-min sonication steps between washes. Beads were stored at 4  $^{\circ}$ C at a concentration of 1% (wt/vol) until use.

For preparing oligonucleotide-coated beads, the following oligonucleotides were ordered HPLC-purified and used as received:

Bead Amine:  
5'/5AmMC6/TTAATTCATTGCGTTCTGTACACG 3'

Bead CGGT:  
5'/5Phos/CGGTCGTGTACAGAACGCAATGAATT 3'

Bead ACCG:  
5'/5Phos/ACCGCGTGTACAGAACGCAATGAATT 3'.

To prepare a double-stranded oligo for coupling, Bead Amine oligo was hybridized to Bead CGGT oligo or to Bead ACCG oligo to generate a double-stranded oligo containing a phosphorylated 5' overhang. Annealing was performed by heating a 1:1 mixture of the oligos in water (0.25 mM each) to 95  $^{\circ}$ C for 10 min, followed by cooling to room temperature on the bench. This resulted in the following oligos:

CGGT Duplex:

5' NH<sub>2</sub>-TTAATTCATTGCGTTCTGTACACG 3'  
3' TTAAGTAACGCAAGACATGTGCTGGC/phos 5'

ACCG Duplex:

5' NH<sub>2</sub>-TTAATTCATTGCGTTCTGTACACG 3'  
3' TTAAGTAACGCAAGACATGTGCGCCA/phos 5'.

We coupled 1  $\mu$ m-diameter carboxylated polystyrene beads (Bangs Labs) to the prepared double-stranded oligos as follows: 10  $\mu$ L of 10% (wt/vol) beads were washed four times with 200  $\mu$ L coupling buffer (Mes 0.1 M, pH 4.7, 150 mM NaCl, 5% DMSO) and dispersed in 20  $\mu$ L coupling buffer. All centrifugations took place for 5 min at 4,500 g. We added 10  $\mu$ L of 20  $\mu$ M double-stranded oligo and 6  $\mu$ L of 2 M EDC, followed by vigorous shaking for 2 h at room temperature. At this point, another 10  $\mu$ L of 2 M EDC were added, followed by overnight shaking at room temperature.

The remaining active EDC was then quenched by adding 2.5  $\mu$ L of 1 M glycine, and the beads were washed five times with storage buffer (20 mM Tris, pH 8, 1 mM EDTA, 0.05% Tween 20, 5 mM sodium azide) with 3 min of sonication between washes. The beads were finally dispersed at a concentration of 1% (wt/vol) and stored at 4  $^{\circ}$ C.

**Bead Passivation.** The beads were passivated by diluting sixfold in TE (20 mM Tris, pH 8, 1 mM EDTA) and addition of  $\beta$ -casein to 1 mg/mL. The beads were vortexed for 10 min, washed once with TE, dispersed at a concentration of 0.2% in TE, and stored at 4  $^{\circ}$ C until the experiment.

**DNA Constructs for Power Spectra Measurements.** Measurement of power spectra used DNA constructs labeled with biotin and digoxigenin. The constructs were prepared by PCR using lambda DNA as the template, with biotinylated or digoxigenin-labeled oligonucleotides (IDT) as primers. The constructs were used after standard PCR purification. To perform the experiment, 1  $\mu$ L of 0.1% anti-Dig-coated beads were incubated with 1  $\mu$ L of 10 nM DNA for 15 min at room temperature, followed by dilution in the experimental buffer (20 mM Tris, pH 8, 130 mM KCl). Streptavidin-coated beads were used directly after passivation.

**DNA Constructs and Proteins for Transcription Experiments.**

**Plasmids and DNA templates.** Plasmids pIA1127 (for expression of sigma 70), pIA1234 (for expression of sortagged RNA polymerase), and pIA2-6 (used as a template for preparing DNA handles) were a gift from Irina Artsimovitch, Ohio State University, Columbus, OH. Plasmid for the expression of sortase was a gift from David Liu, Broad Institute, Cambridge, MA.

A plasmid containing the T7A1 promoter followed by eight repeats containing the His pause, originally described in ref. 4, was modified as follows: The 1,000 base sequence between the promoter and the repeats were removed by digestion with AgeI and BamHI. We annealed 0.7 pmol of the linearized template on a PCR machine with 10 pmol each of a forward (CCGGTGTGGCTCATCGATG) and a reverse (GATC-CATCGATGAGCCAAACA) primer, which were complementary and formed sticky ends complementary with the cut sites.

**Preparation of DNA template.** To prepare the DNA template, the plasmids were restricted by BsaI-HF (8 units per microgram DNA for 2 h at 37  $^{\circ}$ C) and treated in parallel with shrimp alkaline phosphatase (0.4 units per microgram DNA) to generate a linear

DNA with distinct, dephosphorylated 5' overhangs. The enzymes were heat deactivated for 20 min at 65 °C, and the DNA was immediately treated with Klenow 3'-5' exo-polymerase (1 unit per microgram DNA) and 0.1 mM ddATP to generate an assisting force template. The reaction proceeded for 30 min at 37 °C, followed by heat inactivation for 20 min at 75 °C. The sample was then extracted five times with phenol-chloroform and once with chloroform, ethanol precipitated, and reconstituted in 10 mM Tris, pH 8, and 0.1 mM EDTA.

**Preparation of sigma 70.** Plasmid pIA1127 was transformed into Rosetta2 bacteria. The bacteria were grown in 2 L of 2YT medium supplemented with 1% glucose, NPS [25 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 50 mM KH<sub>2</sub>PO<sub>4</sub>, 50 mM Na<sub>2</sub>HPO<sub>4</sub>], 1 mM magnesium sulfate, 34 µg/mL chloramphenicol, and 50 µg/mL kanamycin. The culture was grown at 37 °C to an OD<sub>600</sub> of 0.5, transferred to 17 °C, and IPTG was added to 0.1 mM. Induction proceeded for 16 h.

For purification, the bacteria were dispersed in 80 mL of buffer A25 (20 mM Tris, pH 7.5, 0.5 M NaCl, 10% glycerol, 25 mM imidazole, 2 mM beta-mercaptoethanol) supplemented with 0.1 mg/mL lysozyme and protease inhibitors (Roche). The bacteria were lysed by French press, and the lysate was clarified by centrifugation and filtration and loaded on a 5 mL Ni-NTA column. The column was washed with 20 mL buffer A25 and 20 mL A50 (A25 + 50 mM imidazole), and the his-tagged sigma 70 was eluted in A300 (A25 + 300 mM imidazole). TEV protease [prepared as described (5)] was added at a molar ratio of 1:40, and the reaction proceeded overnight at 4 °C while being dialyzed against A50. The sample was then passed again through a Ni-NTA column. The flow-through, containing non-his-tagged sigma 70 was collected, concentrated twofold, and further purified by gel filtration on a sephacryl S300 column equilibrated with buffer B (20 mM Tris, pH 7.5, 0.5 M NaCl, 10% glycerol, 1 mM EDTA, 1 mM DTT). The protocol yielded ~35 mg of pure sigma 70. Aliquots were flash-frozen in liquid nitrogen and stored at -80 °C.

**Preparation of sortagged RNA polymerase holoenzyme.** Plasmid pIA1234 was transformed into Rosetta2 bacteria. Sortag-RNAP was expressed using the same protocol as sigma 70, except that ampicillin was used instead of kanamycin.

For purification, we used a modified version of a published protocol (6). The cells were dispersed in 75 mL of lysis buffer (50 mM Tris, pH 6.9, 0.5 M NaCl, 5% glycerol) supplemented by 0.1 mg/mL lysozyme and protease inhibitors and lysed by French press. The lysate was centrifuged and filtered, and imidazole was added to 20 mM. The sample was loaded on a 5-mL Ni-NTA column. The column was washed with 30 mL of lysis buffer plus 20 mM imidazole, and the his-tagged RNAP core enzyme was eluted in lysis buffer plus 250 mM imidazole.

To form the holoenzyme, the sample was incubated with a twofold excess of purified sigma 70 overnight on ice. The sample was diluted 10-fold with buffer B0 (50 mM Tris, pH 6.9, 5% glycerol, 0.5 mM EDTA, 1 mM DTT) and loaded on a heparin 5-mL column. To avoid overloading the column, the sample was divided into three portions that were loaded separately. A gradient of 50 mM to 1 M NaCl was used to elute the protein. RNAP holoenzyme was separated clearly from excess sigma 70. The sample was dialyzed against buffer B50 (50 mM Tris, pH 6.9, 5% glycerol, 50 mM NaCl, 0.5 mM EDTA, 1 mM DTT) and then purified further on a 1-mL monoQ column using a 50 mM to 1 M NaCl gradient (again, the sample was split into three portions loaded separately). Pure fractions were pooled, dialyzed against storage buffer (20 mM Tris, pH 7.5, 200 mM KCl, 0.2 mM EDTA, 0.2 mM DTT, 5% glycerol), aliquoted, flash-frozen, and stored at -80 °C.

**Biotinilation of sortag-RNAP.** We obtained a peptide containing an N-terminal GGG tag with a biotin-modified lysine residue (Genscript): GGGDGDY{Lys(biotin)}.

We reacted 100 µL of 9.6 µM sortag-RNAP with a 200-fold excess of biotinilated peptide in 200 µL coupling buffer (50 mM

Tris, pH 7.5, 5 mM CaCl<sub>2</sub>) containing 2 µM sortase [prepared as described (7)]. The reaction proceeded for 60 min. At this point, imidazole was added to 25 mM and NaCl to 350 mM, and the sample was passed through 70 µL Ni-NTA beads to remove sortase and unreacted RNAP. The peptide was removed by dialysis into storage buffer, and the biotinilated RNAP was stored in storage buffer at -80 °C.

#### Measurements of STEPS and Transcription Traces.

**Preparation of stalled complexes and tether assembly.** Stalled complexes were prepared by incubating 2 nM DNA with 10 nM RNAP in TB20 (20 mM Tris, pH 8, 20 mM NaCl, 20 mM DTT, 10 mM MgCl<sub>2</sub>, 20 µg/mL casein) for 20 min at 37 °C.

Stalled complexes were then ligated to the beads at a ratio of 1 fmol stalled complex to 2 µg beads in TB20 in the presence of 0.1 mM ATP and 0.4 units of T4 DNA ligase, for 60 min at room temperature. For 1.5 kb DNA handles, 1 fmol handle was ligated to 3 µg beads.

Following the ligation, heparin was added to 0.4 mg/mL of the beads. To the beads ligated to the DNA handle, a 200-fold excess of neutravidin was added and incubated with the beads for 10 min before diluting with experimental buffer. For the stalled complex beads, beads were incubated for 10 min with the added heparin before dilution with experimental buffer.

**Experimental setup.** The experimental buffer was: 20 mM Tris, pH 8, 130 mM KCl, 10 mM MgCl<sub>2</sub>, 0.1 mM DTT, 0.1 mM EDTA, and 10 mM NaN<sub>3</sub>. NTP concentrations were 10 µM. To perform the experiment, we employed a three-channel setup previously described (8). The experiment was performed in a homemade flow chamber with separate entries for buffer and for NTP solution. Every experiment consisted of the following steps:

- i) Trap a DNA handle bead, followed by the trapping of a stalled complex bead.
- ii) Obtain a power spectrum of the beads, for calibration of the conversion factors between measured voltages and bead displacement (9).
- iii) Rub the beads against each other until a tether is formed (if at all).
- iv) Perform the STEPS procedure a total of six times, twice for each dwell time distribution (uniform, exponential, and gamma-2).
- v) Flow the NTPs from the shunt to restart transcription. The tether was maintained under a tension of  $15 \pm 1$  pN, by maintaining the traps positions until transcription caused the tension to drop below 14 pN, at which point the trap positions were adjusted away until the tension reached 16 pN.

**Data Analysis.** A small fraction of traces exhibited abnormally noisy behavior, including apparent two-state hopping (Fig. S1). Such traces were manually excluded after visual inspection. All other traces were included in the subsequent analysis.

**LSS-HMM.** A LSS-HMM was implemented in the Cython programming language (10). The core algorithm followed the description available in refs. 11, 12, and 13. The analysis parameters (described in the previously mentioned references) were set as follows: STEPS traces and segments of transcription activity traces between two consecutive trap position adjustments were down-sampled to 200 Hz and the quantization size to 0.025 nm; a null prior was used for negative step sizes; and each HMM was run for up to 1,000 iterations, up to 60 s or until the change in likelihood between iterations dropped below  $10^{-8}$  (whichever occurred first).

The theory of the LSS-HMM algorithm (11–13) allows the molecular motor to take steps both in the forward and in the backward direction (backtracking). However, under 15 pN of assisting force, backtracking events are rare; we found it beneficial

to force the steps to be always in the forward direction. Specifically, the prior on the step size distribution  $p(d)$  was set to zero for  $d < 0$ . In the absence of such a constraint, some of the traces would be fitted to a collection of steps that quickly alternate between forward and backward motion (about every 100 ms). The origin of this motion is not known. Preventing any backward motion forces the LSS-HMM to average out these dynamics into a single state.

In certain cases, the LSS-HMM would fail to converge to a valid step distribution. Numerically, such a failure would manifest itself as a numerical underflow of the likelihood. Such failures could arise due to two reasons. First, a trace can exhibit actual backtracking, contradicting our initial assumption. Second, RNA polymerase can enter long-lived pauses (14, 15) that we also found to be, counterintuitively, detrimental to the performance of the LSS-HMM—likely due to the difficulty for the LSS-HMM algorithm to distinguish between very slow activity and residual low-frequency noise.

When the LSS-HMM failed to converge for a STEPS trace, the trace was discarded. When it failed to converge for a transcription activity segment, the segment was split into two halves that were fitted separately; the procedure was recursively repeated until the fit succeeded or the segment length dropped below 5 s. In such a manner, periods of backtracking or of long-lived pauses can be separated from the analysis, whereas segments containing processive activity would be analyzed. Note that if the abnormally noisy traces described above had not been removed manually, this algorithm would likewise reject them.

The analysis of segments containing very few steps (even if they are longer than 5 s) is difficult for any HMM-based algorithm, as it needs to learn the correct step distribution from the dataset self-consistently. As such, we also discarded segments fitted to less than an arbitrarily chosen threshold of 10 steps.

**Quantification of the accuracy of the fit for STEPS data.** To quantify the time accuracy of the fit for STEPS data, we first determined if all real steps (i.e., actual trap motions) were correctly detected or if some of them were missed (underfitting). Likewise, we evaluated if scored steps, as scored by the LSS-HMM algorithm, correspond to real steps or if some of them were spurious (overfitting).

To quantify underfitting, we paired each real step with the closest scored step. If done naively, such pairing could fail to detect a case where two temporally close real steps are fitted with a single step—thus missing a short dwell between the two real steps. Therefore, we additionally imposed the condition that the pairing between fitted steps and real steps must be one-to-one (if the number of fitted steps was smaller than real steps, some real steps were left unpaired). Specifically, we required that the number of matches be equal to the lower between the number of real steps and the number of scored steps, and we minimized the sum of the time intervals between the paired steps.

We then asked, What is the distribution of the time intervals between the real and the scored step in the pairings? In other words, how far is each real step from the closest scored step? We find that regardless of the dwell time distribution (constant, exponential, or gamma-2), more than 70% of the real steps are within 100 ms (one-fifth of the mean dwell time) of the closest scored step (Fig. 5 D–F, blue curves).

To quantify overfitting, we asked how far each scored step is from the closest real step, once again imposing one-to-one cor-

respondence. In this case too, at least 70% of the scored steps were found within 100 ms of the closest real step, with the exception of the biologically less relevant case of constant steps, where the allowance had to be raised to 165 ms (Fig. 5 D–F, orange curves).

## SI Discussion

**Effect of Tether Compliance on the Step Size in STEPS Data.** In classical optical tweezers measurements, the distances of the beads to the centers of their respective traps,  $\Delta x_{\text{bead1-trap1}}$  and  $\Delta x_{\text{bead2-trap2}}$ , are subtracted from the distance between the traps,  $\Delta x_{\text{trap1-trap2}}$  (the latter being set by the experimentalist). Such an approach is suitable to track the time-varying extension of the tether between the traps. In the case of STEPS data, however, because the tether is  $\sim 17\times$  stiffer than the traps taken together (1.7 pN/nm at 15 pN vs. two traps of 0.2 pN/nm each), the tether extension only changes by 1/18 of the trap motion, whereas the total displacement of the beads away from the traps changes by 17/18 of this motion (i.e., 0.32 nm as the traps were moved by 0.34 nm). Thus, we chose to analyze the total displacement data.

**Width of the Recovered Step Size Distributions.** Surprisingly, we observed that even though dwell times are well recovered in our STEPS analysis, the corresponding step size distributions are very wide. We rationalized this observation on the basis that LSS-HMM can assign variation in the measured bead position to two sources—actual spread of the step size distribution and additional Gaussian noise; thus, LSS-HMM can choose to report a wider step size distribution to narrow the reported Gaussian noise distribution.

**Fraction of the Data Analyzed.** The fitted segments of the transcription traces were at least 10 bp long and reached as much as 70 bp (Fig. 6B). Note that the maximum length of the fitted segment is due to the need to maintain the force within a range of 2 pN; with a trap stiffness of 0.2 pN/nm per trap, the traps must be displaced—and thus a new fit region must be started—every 20 nm (10 nm on each trap)—that is, approximately every 70 bp. In the absence of PPI, the fitted segments amounted to 75% of the total distance transcribed but only 43% of the total duration of the traces, due to the selective removal of long-lived pauses from the analyzed datasets (in the presence of PPI, the respective percentages are 51% of the distance and 24% of the duration). As such, the dwell time distribution obtained from the analysis faithfully represents the true distribution of pause-free translocation in short time scales; however, the distribution is underestimated at longer time scales due to the rejection of slow segments. Due to this underestimation, we only considered events shorter than 1 s when comparing the dwell time histograms in the absence and the presence of PPI.

Note that the segments rejected are well-defined and could be subjected to further analysis by pooling short segments together before fitting with LSS-HMM to increase the statistics. Our results should be contrasted with earlier reports of the observation of single base-pair stepping in optical tweezers, which was limited to short segments ( $\sim 15$  bp) corresponding to  $\sim 10\%$  of the distance transcribed in  $\sim 10\%$  of the collected traces (16).

- Moffitt JR, Chemla YR, Izhaky D, Bustamante C (2006) Differential detection of dual traps improves the spatial resolution of optical tweezers. *Proc Natl Acad Sci USA* 103:9006–9011.
- Bustamante C, Chemla YR, Moffitt JR (2009) High-resolution dual-trap optical tweezers with differential detection: Managing environmental noise. *Cold Spring Harb Protoc* 2009:pb0172.
- Comstock MJ, Ha T, Chemla YR (2011) Ultrahigh-resolution optical trap with single-fluorophore sensitivity. *Nat Methods* 8:335–340.
- Herbert KM, et al. (2006) Sequence-resolved detection of pausing by single RNA polymerase molecules. *Cell* 125:1083–1094.
- Tropea JE, Cherry S, Waugh DS (2009) Expression and purification of soluble His 6-tagged TEV protease. *High Throughput Protein Expression and Purification: Methods and Protocols*, ed Doyle SA (Humana, New York), pp 297–307.
- Svetlov V, Artsimovitch I (2015) Purification of bacterial RNA polymerase: Tools and protocols. *Bacterial Transcriptional Control: Methods and Protocols*, eds Artsimovitch I, Santangelo TJ (Humana, New York), pp 13–29.
- Chen I, Dorr BM, Liu DR (2011) A general strategy for the evolution of bond-forming enzymes using yeast display. *Proc Natl Acad Sci USA* 108:11399–11404.
- Galbur EA, Grill SW, Bustamante C (2009) Single molecule transcription elongation. *Methods* 48:323–332.

