www.sciencemag.org/cgi/content/full/313/5788/801/DC1

Supporting Online Material for

Single-Molecule, Motion-Based DNA Sequencing Using RNA Polymerase

William J. Greenleaf and Steven M. Block*

*To whom correspondence should be addressed. E-mail: sblock@stanford.edu

Published 11 August 2006, *Science* **313**, 801 (2006)
DOI: 10.1126/science.1130105

This PDF file includes:

Materials and Methods
Fig. S1
References

Supporting Online Materials

Single-Molecule, Motion-based DNA Sequencing using RNA Polymerase

William J. Greenleaf and Steven M. Block, Stanford University

Materials and methods

Experimental setup

The dumbbell assay geometry used for single molecule studies of transcription by RNA polymerase (RNAP) is shown in Fig. S1; this system is described in greater detail elsewhere (5, S1). The DNA template used for these experiments was modified slightly from that used in Ref. 5 by the inclusion of an additional *his* terminator sequence at the end of the *rpoB* gene (S2, 6). Records were digitally acquired at 2 kHz (6), decimated with a boxcar filter to a 100 Hz sampling rate, then median filtered at 110 ms.

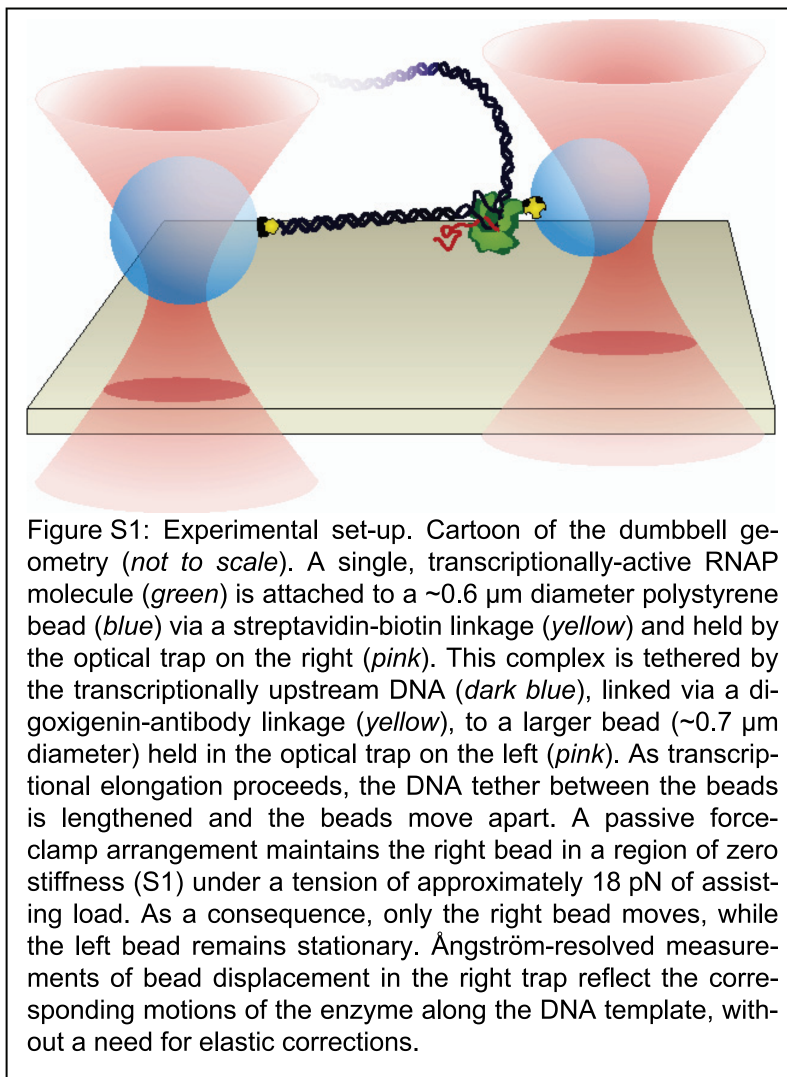

Figure S1: Experimental set-up. Cartoon of the dumbbell geometry (*not to scale*). A single, transcriptionally-active RNAP molecule (*green*) is attached to a ~0.6 µm diameter polystyrene bead (*blue*) via a streptavidin-biotin linkage (*yellow*) and held by the optical trap on the right (*pink*). This complex is tethered by the transcriptionally upstream DNA (*dark blue*), linked via a digoxigenin-antibody linkage (*yellow*), to a larger bead (~0.7 µm diameter) held in the optical trap on the left (*pink*). As transcriptional elongation proceeds, the DNA tether between the beads is lengthened and the beads move apart. A passive force-clamp arrangement maintains the right bead in a region of zero stiffness (S1) under a tension of approximately 18 pN of assisting load. As a consequence, only the right bead moves, while the left bead remains stationary. Ångström-resolved measurements of bead displacement in the right trap reflect the corresponding motions of the enzyme along the DNA template, without a need for elastic corrections.

Record alignment procedure

First, single-molecule records of transcriptional elongation vs. time were roughly aligned, as follows. If the end of a transcription record occurred near a location corresponding to a terminator sequence, then the absolute position was assigned to be that of the terminator. If the end of the record did not occur near a terminator, a contiguous portion of the record devoid of long pauses was used instead to anchor the position: such segments correspond to sequence regions where one of the four NTPs is absent. Because these data were acquired using an optical force clamp, one of the two traps was moved episodically in increments of 20 nm to maintain the position of the detected bead inside the force-clamped region (5). Log dwell-time histograms of these 20-nm-long segments of continuous transcription were smoothed, then individually aligned on an expected histogram "mask." Outside the alignment region, this mask consisted of unit positive peaks at the positions of the limiting nucleotide and quarter-unit negative peaks were at the positions of non-limiting nucleotides. Inside the alignment region, this mask consisted of quarter-unit positive peaks for every base-window (*i.e.* the average of all the nucleotide-specific masks), and therefore contained no specific sequence information. The log dwell-time histograms for each segment were autocorrelated against this mask (5). In this process, the segments were allowed to shift by ± 1 nm with respect to the previous segment, and allowed to stretch by ± 8%. A stretching parameter allows compensation for the variation in sensitivity in position detection due to size heterogeneity of the polystyrene beads, as well as the sequence-dependent heterogeneity in the linear rise per base for double stranded DNA (6). A shift pa-

rameter allows for compensation of the positional uncertainty generated due to periodic updates of the force clamp. Because of a broad distribution of dwell times at the limiting base, peaks in the dwell-time histogram were used as the primary signal of base identity, using the base-assignment heuristic described in the text. Segments of records used for sequencing analysis were selected from continuous single-molecule records of transcription along templates (~1,000 bp) carrying the known sequence of the *rpo*B gene from *E. coli*, which codes for the beta subunit of RNA polymerase.

Data analysis

The histograms in Fig. 1 were generated with a 0.1 bp bin width. These histograms were then normalized to have the same integrated area within 44 bp centered on the region to be sequenced. The histograms were then smoothed with a 5-pt boxcar filter followed by a 3-pt binomial filter. Analysis was carried out in Igor Pro 5.01 (Wavemetrics).

Supplemental References

S1. W. J. Greenleaf, M. T. Woodside, E. A. Abbondanzieri, S. M. Block. *Phys. Rev. Lett.* **95**, 208102 (2005).

S2. We thank to Matt H. Larson of the Block lab for the gift of this template and Robert Landick for providing the RNAP enzyme.