Supplemental Information for:

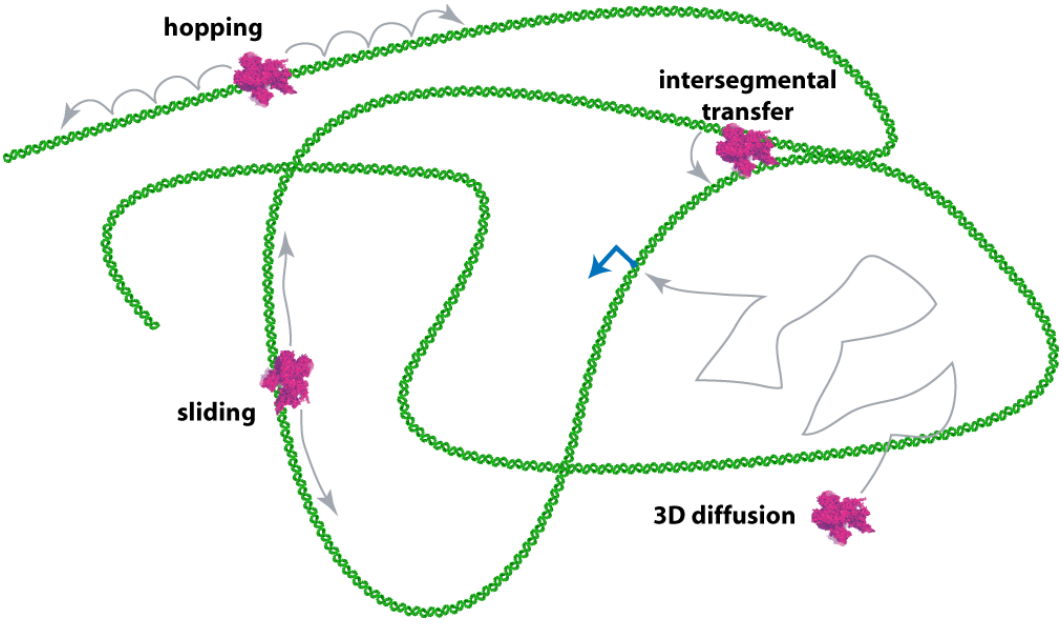
The promoter search mechanism of *E. coli* RNA polymerase is dominated by three–dimensional diffusion

Feng Wang[1,6], Sy Redding[2,6], Ilya J. Finkelstein[1,5], Jason Gorman[3,5], David R. Reichman[2], and Eric C. Greene[1,4]

[1]Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY, USA. [2]Department of Chemistry, Columbia University, New York, NY, USA. [3]Department of Biological Sciences, Columbia University, New York, NY, USA. [4]Howard Hughes Medical Institute, Columbia University, New York, NY, USA. [5]Present addresses: Department of Chemistry, and the Institute for Cellular & Molecular Biology, University of Texas at Austin, Austin, TX, USA (I.J.F.); Vaccine Research Center, National Institutes of Health, Bethesda, MD, USA (J.G.). [6]These authors contributed equally. Correspondence should be addressed to E.C.G. (ecg2108@columbia.edu)

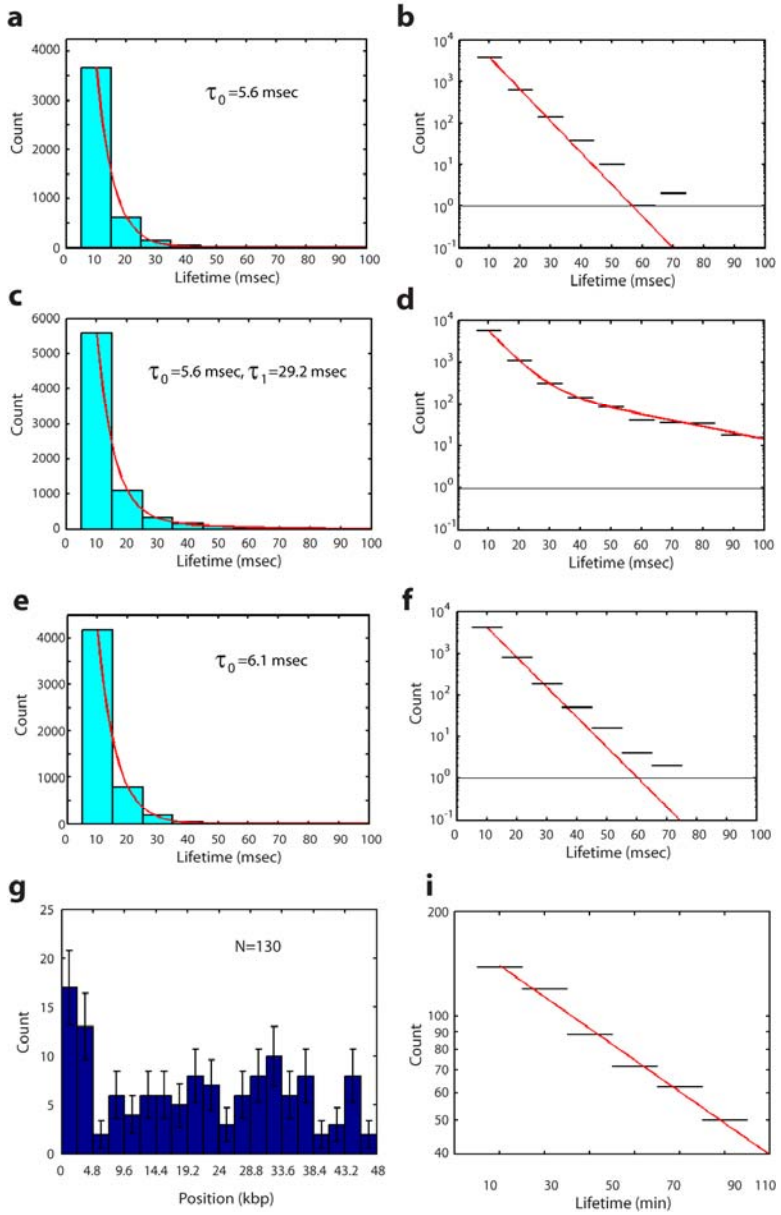
This supplementary information includes:

Nature Structural & Molecular Biology: doi:10.1038/nsmb.2472



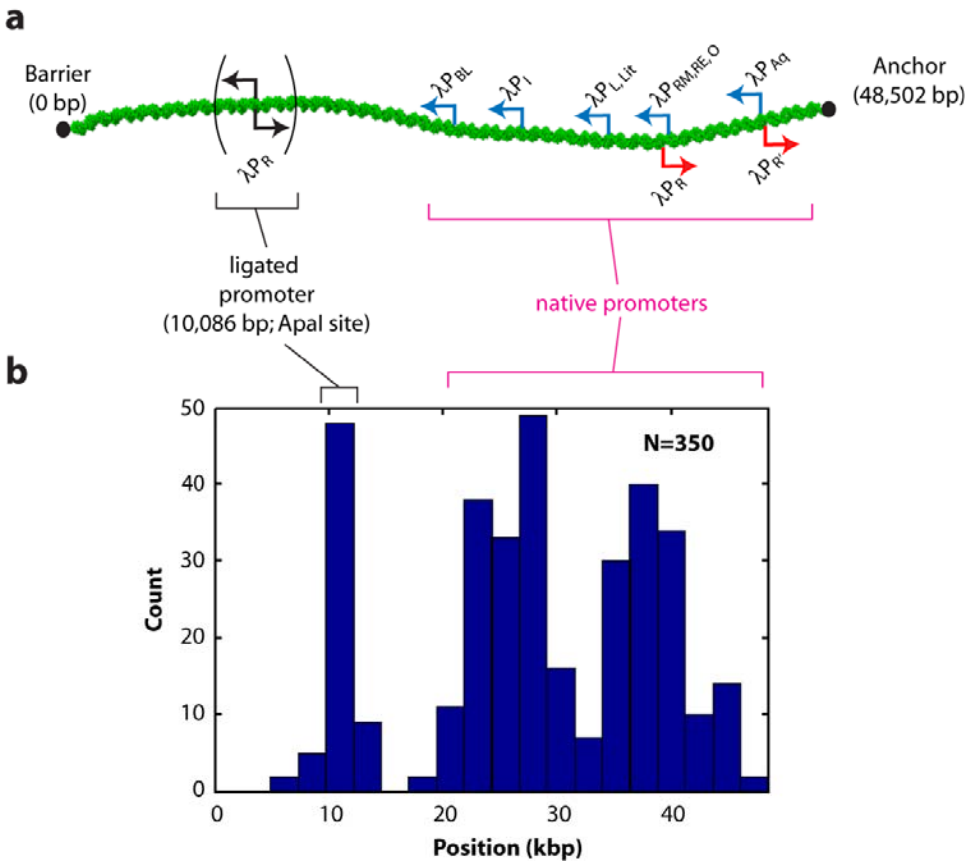
Supplementary Fig. 1. The Target Search Problem. Diffusion–based models for how proteins might search for binding targets: random collision through 3D–diffusion (*i.e.* jumping); 1D–hopping, involving a series of microscopic dissociation and rebinding events; 1D–sliding, wherein the protein moves without dissociating from the DNA; and intersegmental transfer, involving movement from one distal location to another via a looped intermediate. These mechanisms are not mutually exclusive, and the latter three are categorized as facilitated diffusion because by reducing dimensionality they allow target association rates exceeding limits imposed by 3D–diffusion. DNA is green, the target site (promoter) is blue, and RNAP is magenta.

Nature Structural & Molecular Biology: doi:10.1038/nsmb.2472



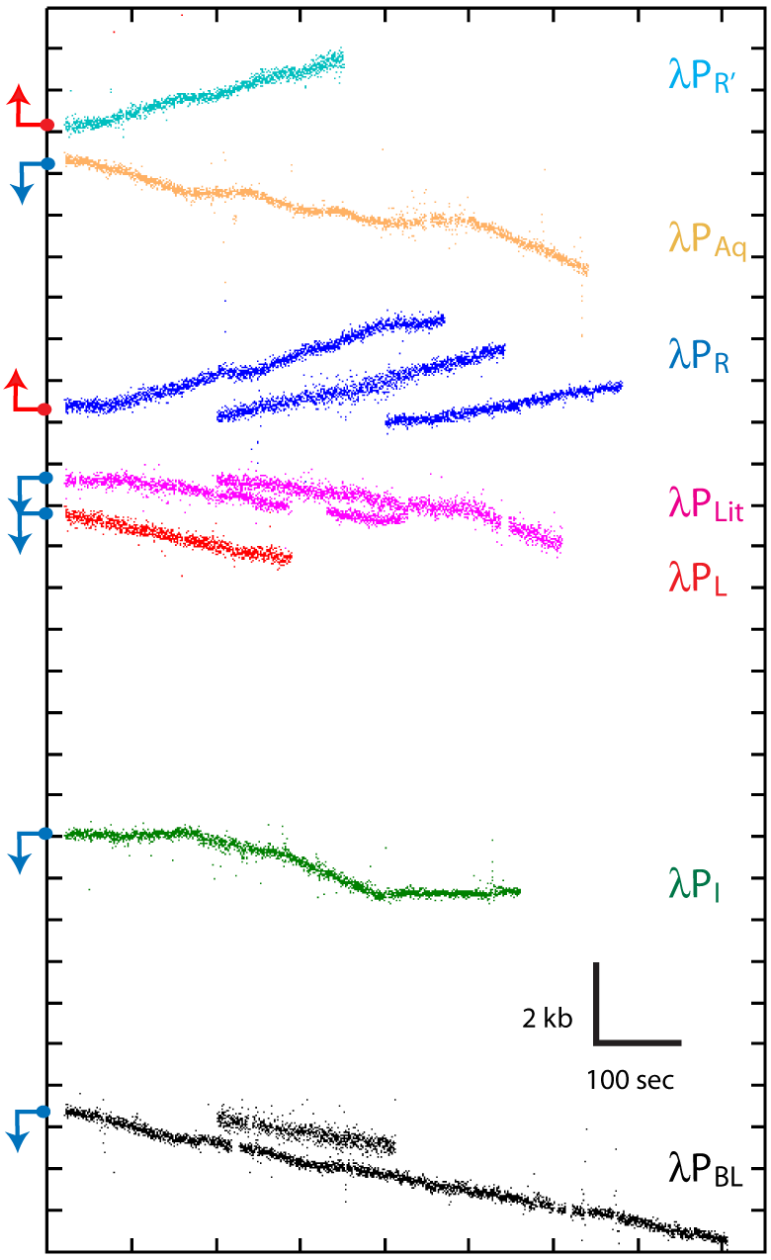
Supplementary Fig. 2. Lifetime analysis of τ_0 and τ_1 events. a, Histogram of lifetimes for QDs only in the absence of RNAP, and the red line is a single exponential fit to the histogram. **b,** Shows the same QD-only data, but the y-axis is on a logarithmic scale. **c** and **d,** Histogram of lifetimes for QD-RNAP and corresponding double exponential fit. The first time constant obtained from the double exponential fit (5.6 msec) is the same as is obtained from the single exponential fit to the QD-only data set. **e** and **f,** Histogram of lifetimes for QD-RNAP and corresponding exponential fit for data collected in the absence of DNA. **g,** This binding distribution uses the data points presented in Fig. 2c, but the data were restricted to only those events that had a lifetime of ≥40-msec. Based upon the two exponential components obtained from the lifetime measurements, this ensures that most of the events (>93%) plotted in this binding distribution histogram are τ_1 events (*i.e.* nonspecifically bound RNAP). **i,** Semi-log plot of the lifetime distributions for the τ_3 events, corresponding to the inset shown in the lower panel Fig. 2c.

Nature Structural & Molecular Biology: doi:10.1038/nsmb.2472



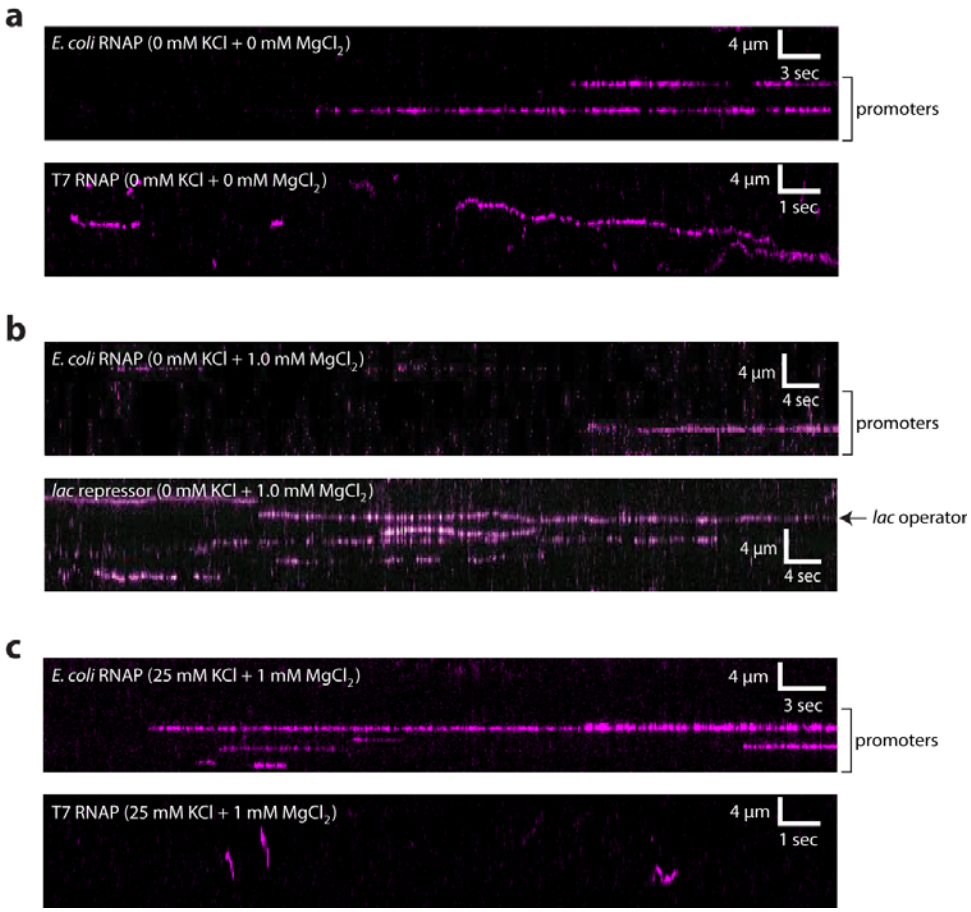
Supplementary Fig. 3. Promoter binding by QD-RNAP. a, Schematic of substrate with a ligated λP_R promoter. As a further verification that QD-RNAP was binding to promoters in the phage DNA, and not simply associating with the AT-rich right half of the molecule, a 100-bp synthetic DNA fragment (IDT) spanning positions -67 to +24 of P_R promoter was ligated into the ApaI site on the left half of the phage genome. Successful insertion of the promoter fragment destroys the ApaI site. The presence of the insert was confirmed by PCR, and products lacking the insert were then selected against by further digestion with ApaI prior to assembly of the DNA curtains (molecules that get cut with ApaI cannot be assembled into double-tethered curtains). The ligation mixtures contain a heterogeneous mixture of substrates with the promoter fragment inserted in either orientation, as depicted. **b,** Binding site distribution. The ligated DNA was used to assess QD-RNAP binding distributions in single-molecule DNA curtain assays. As shown here, the presence of the new λP_R promoter fragment resulted in a new peak of QD-RNAP in the binding distribution at the expected location (compare to Fig. 2c, bottom panel)

Nature Structural & Molecular Biology: doi:10.1038/nsmb.2472



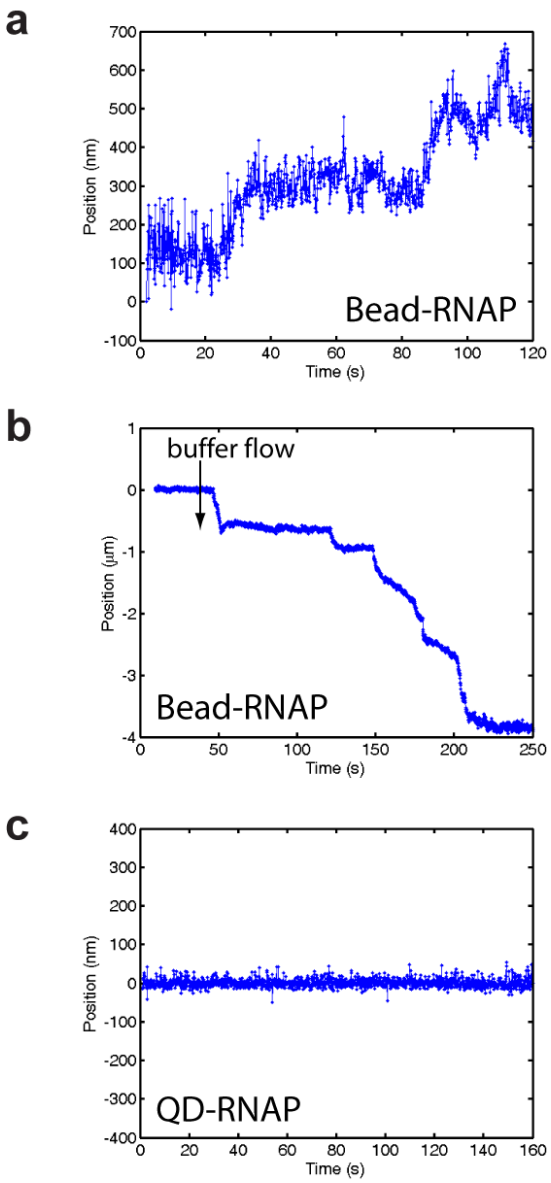
Supplementary Fig. 4. Transcription by QD-RNAP. Examples of RNAP movement along the DNA in the presence of all four rNTPs, and data were collect at room temperature. RNAP and rNTPs were pre-mixed prior to injection into the sample chamber. The trajectories are color coded for each corresponding promoter, and the relative orientation of each promoter is indicated on the left.

Nature Structural & Molecular Biology: doi:10.1038/nsmb.2472



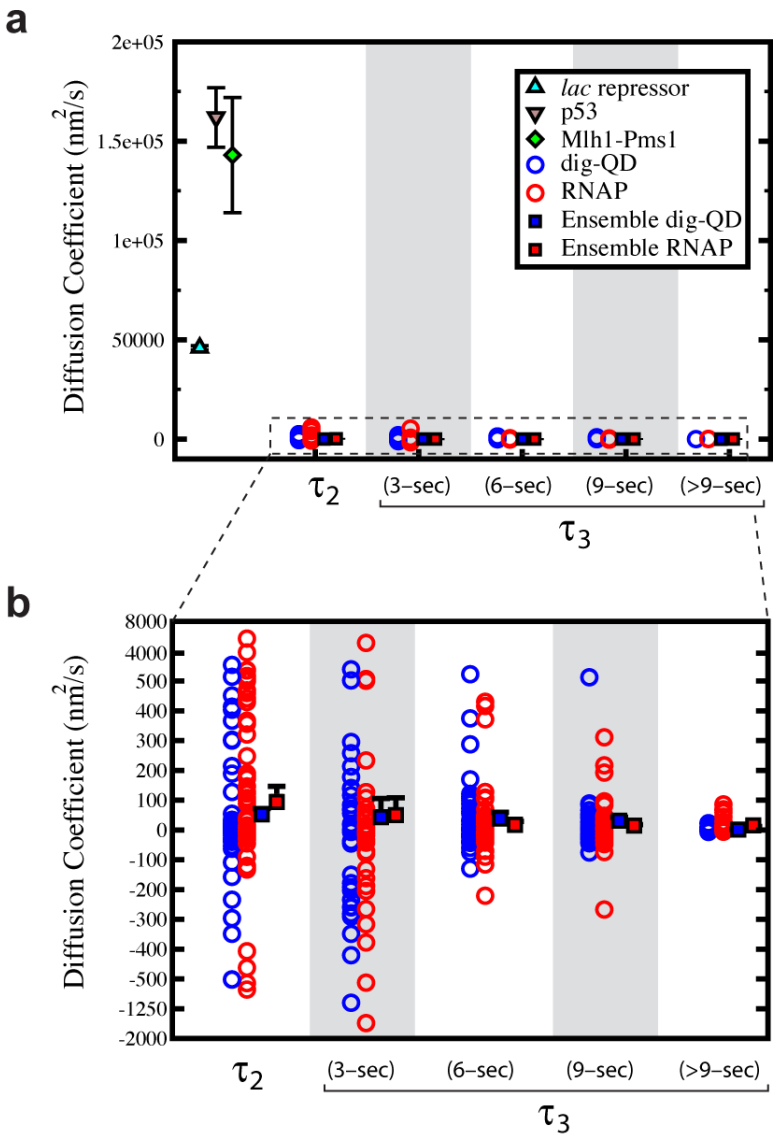
Supplementary Fig. 5. 1D diffusion of lac repressor and T7 RNAP. Kymograms comparing *E. coli* RNAP to QD-tagged T7 RNAP and lac repressor, both of which can diffuse along DNA under low ionic strength conditions.[1-3] **a,** *E. coli* RNAP compared to T7 RNAP; buffer conditions: 40 mM Tris-HCl (pH 8.0), 0.2 mg ml^{-1}, 5 mM DTT for T7 RNAP and 1 mM DTT for *E. coli* RNAP. **b,** *E. coli* RNAP compared to lac repressor; buffer conditions: 10 mM Tris HCl (pH 8.0), 1 mM $MgCl_2$, 1 mM DTT, 1 mg ml^{-1} BSA. The DNA used in the experiments with lac repressor contained a single, 21-bp symmetric lac operator, as indicated by the arrow.[4] **c,** *E. coli* compared to T7 RNAP; buffer conditions: 20 mM Tris-HCl (pH 8.0), 25 mM KCl, 1 mM $MgCl_2$, 1 mM DTT, 0.2 mg ml^{-1} BSA.

Nature Structural & Molecular Biology: doi:10.1038/nsmb.2472



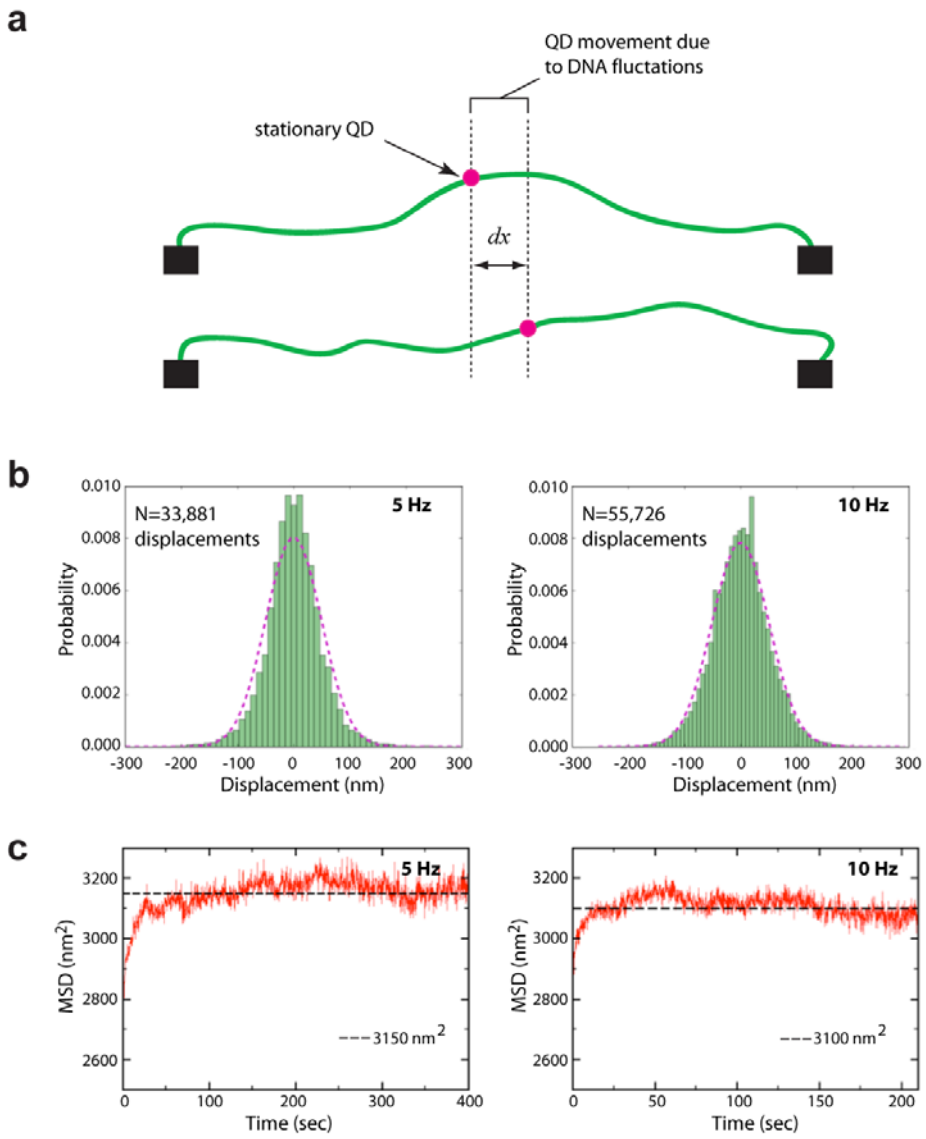
Supplementary Fig. 6. RNAP Bead-Aggregates Exhibit 1D Movement. a, Particle-tracking trajectory showing 1D diffusive motion of an RNAP-saturated bead (1.0 μm) bound to a DNA molecule in the absence of buffer flow. **b,** Trajectory of an RNAP-saturated bead (1.0 μm) when buffer flow (0.4 ml min^{-1}) was applied in the direction indicated by the arrowhead. **c,** A typical trajectory of QD-tagged RNAP bound to DNA is shown for comparison.

Nature Structural & Molecular Biology: doi:10.1038/nsmb.2472



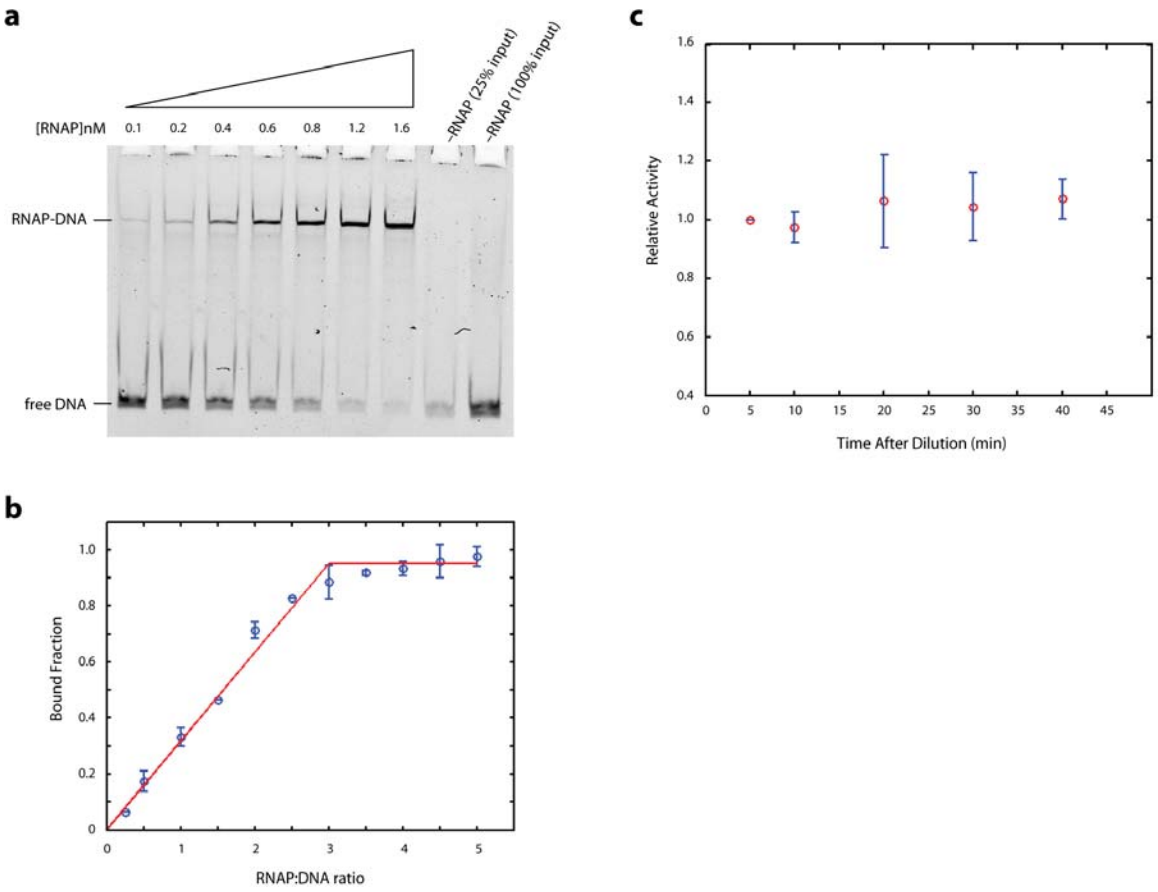
Supplementary Fig. 7. RNAP and dig-QD Diffusion Coefficient Data. a, Shows a comparison of the single molecule and ensemble diffusion coefficients obtained for QD-tagged RNAP and an immobilized dig-QD (this study), along with reported values for *lac* repressor [2], p53 [5], and Mlh1-Pms1 [6]. **b,** Magnified view of the RNAP and dig-QD data sets. Red circles represent diffusion coefficients obtained from all individual particle-tracking trajectories for RNAP, and blue circles represent diffusion coefficients from dig-QD trajectories collected and analyzed under identical conditions. Squares represent ensemble values for the diffusion coefficients obtained from the cumulative tracking data along with corresponding error bars. Diffusion coefficients are gamma distributed, therefore we report the magnitude of the square root of the variance (error bars). Also see Supplementary Notes.

Nature Structural & Molecular Biology: doi:10.1038/nsmb.2472



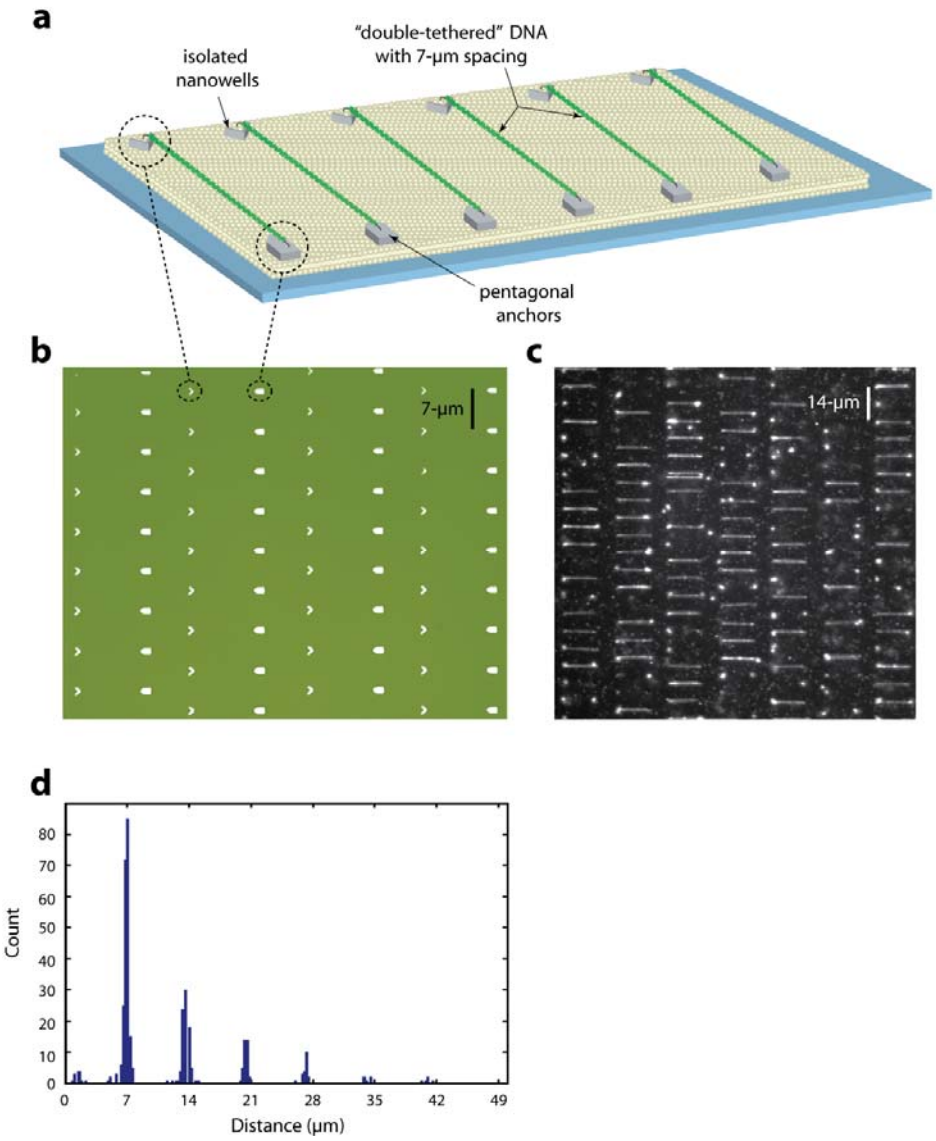
Supplementary Fig. 8. Diffusion Coefficients and DNA Fluctuations. **a,** Cartoon illustration of DNA motion giving rise to the apparent diffusion coefficients for the stationary dig-QDs The underlying fluctuations of the DNA were analyzed by linking a single QD to a fixed digoxigenen tag covalently attached to the double-tethered DNA molecules.[7] **b,** Distributions of single-frame displacements for data collected at either 5 or 10 Hz (as indicated) for the entire dig-QD data set. The distributions have been normalized, and the overlay is a Gaussian fit generated using the mean and standard error of the distribution. The number of individual displacements is indicated. **c,** Reference graphs showing the mean squared displacement analysis of the stationary dig-QD particles. See Supplementary Notes for additional details.

Nature Structural & Molecular Biology: doi:10.1038/nsmb.2472



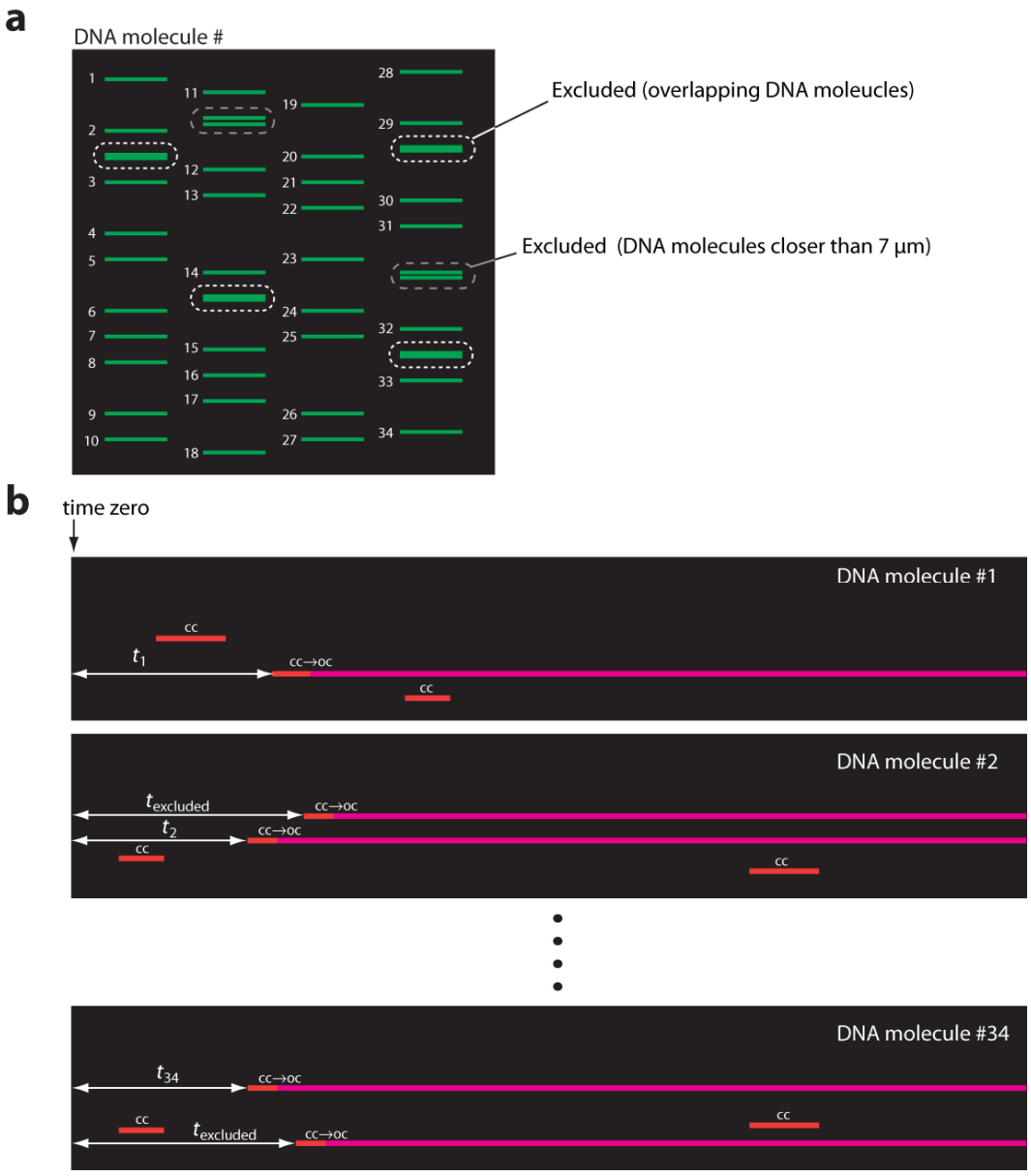
Supplementary Fig. 9. Activity of RNAP Under Dilute Conditions. a, Gel shift assay showing RNAP and promoter λP_R association. RNAP was titrated into reactions containing 0.4 nM Cy-3 labeled promoter DNA fragment and then challenged with heparin to disrupt protein-DNA complexes that had not formed open complexes. The right two lanes are examples of negative controls containing different amounts of input DNA, which was used to calibrate band intensity. **b,** Quantitation showing the fraction of bound promoter DNA fragment as a function of the RNAP to DNA ratio. **c,** Stability of RNAP under dilute conditions. RNAP was diluted to 0.6 nM and incubated at room temperature in the absence of DNA. After the indicated time intervals, the samples were assayed for DNA binding activity using the Cy3-labeled λP_R promoter DNA fragment. Bound and unbound DNA fractions were separated by native gel electrophoresis and quantitated based on the fluorescence intensity of the bands.

Nature Structural & Molecular Biology: doi:10.1038/nsmb.2472



Supplementary Fig. 10. Parallel Array of Double-tethered Isolated (PARDI) Molecules. a, Schematic diagram of the new PARDI DNA curtain design used for the promoter association rate measurements. **b,** Optical image highlighting nanofabricated PARDI pattern design. **c,** Image of a typical PARDI field-of-view, showing the double-tethered, YOYO1-stained DNA molecules. **d,** Histogram showing the measured distances between neighboring DNA molecules anchored to the PARDI patterned surface.

Nature Structural & Molecular Biology: doi:10.1038/nsmb.2472



Supplementary Fig. 11. Promoter Association Rate Analysis. a, Illustration of a PARDI curtain, where each DNA is numbered. Overlapping DNA and DNA molecules closer than 7-µm are excluded from analysis. **b,** A kmyogram is made for each DNA in the field-of-view, and the time required for detection of the first promoter bound protein on each individual DNA is extracted from the kymograms (*e.g.:* t_1 & t_2; for DNA #1 & #2, respectively). Importantly, we are not measuring the rate of either closed complex (cc; schematically represented as red lines) formation or open complex (oc; schematically represented as magenta lines) formation, but rather we are measuring the exact instant (with 100 msec resolution) at which a single molecule of RNAP is detected at a promoter for the molecules of RNAP that subsequently make a successfully transition to open complexes. Once one promoter in occupied, all subsequent binding events on that same DNA molecule are excluded from further analysis (*e.g.:* $t_{excluded}$ for DNA molecules #2 & #34); therefore in this example, we could obtain a maximum of 34 data points. These restrictions ensure we only record binding events that occur when all promoters on a given DNA are initially accessible for binding by RNAP, in accordance with calculation parameters. See Supplementary Notes for complete details.

Nature Structural & Molecular Biology: doi:10.1038/nsmb.2472

Supplementary Table 1. Summary of Prior Promoter Search Studies.

Reported $D_{1,obs}$ (nm^2 sec^{-1})	**Reported Sliding Distance**	**Nonspecific Binding Lifetime**	**Experimental Approach**	**Reference**
N.A.	No 1D Sliding.	≤30 msec	Single molecule TIRFM	This study.
$1.3x10^5$	~250 bp	N.A.	Transcription initiation kinetics	Ricchetti *et al.*, 1988. [8]
$1.5x10^5$	~2,000 bp	~3.3 sec	Rapid mixing followed by photocrosslinking	Singer and Wu, 1987. [9]
$0.62\text{-}1.5x10^5$	85-13,000 bp	N.A.	Rapid mixing followed by photocrosslinking	Singer and Wu, 1988.
$\sim 1x10^4$	~90 nm (~300 bp)	0.12 sec	Single molecule fluorescence imaging	Harada *et al.*, 1999. [10]
N.A.	>10 μm (>30 kb)	N.A.	Single molecule fluorescence imaging	Kabata *et al.*, 1993. [11]
$1.1x10^1$	~150 nm (~440 bp)	~600 seconds	AFM, did not address how promoters were located	Guthold *et al.*, 1999. [12]

Nature Structural & Molecular Biology: doi:10.1038/nsmb.2472

Supplementary Table 2. List of λ-Phage Promoters.

Promoter	Position (bp)	Orientation	Reference
λP_{BL}	23,231	left	NCBI Nucleotide data base http://www.ncbi.nlm.nih.gov/nuccore/J02459.1
λP_{I}	29,065	left	NCBI Nucleotide data base
λP_{L}	35,582	left	NCBI Nucleotide data base
λP_{Lit}	36,256	left	NCBI Nucleotide data base
λP_{RM}†	37,940	left	[13, 14]
λP_{R}†	38,023	right	NCBI Nucleotide data base
λP_{RE}	38,343	left	NCBI Nucleotide data base
λP_{O}	38,675	left	NCBI Nucleotide data base
λP_{Aq}	44,146	left	[15]
$\lambda P_{R'}$	44,587	right	NCBI Nucleotide data base

†There are ten promoters in the λ phage genome, however, binding to λP_{RM} and λP_{R} is mutually exclusive [16], therefore a maximum of nine promoters can be occupied in our assays.

Nature Structural & Molecular Biology: doi:10.1038/nsmb.2472

Supplementary Table 3. Buffer Conditions Tested in this Study.

Buffer Conditions	Reference
20 mM Tris-HCl [pH 8.0], 100 mM KCl, 10 mM $MgCl_2$, 1 mM DTT, ±250 mM rNTPs, 0.2 mg ml^{-1} BSA	This study.
20 mM Tris-HCl [pH 8.0], 25 mM KCl, 1 mM $MgCl_2$, 1 mM DTT, 0.2 mg ml^{-1} BSA	This study.
40 mM Tris-HCl [pH 8.0], 50 mM KCl, 5 mM $MgCl_2$, 1 mM DTT, 0.2 mg ml^{-1} BSA	This study.
40 mM Tris-HCl [pH 8.0], 100 mM KCl, 10 mM $MgCl_2$, 1 mM DTT, 0.2 mg ml^{-1} BSA	This study.
40 mM Tris-HCl [pH 8.0], 1 mM DTT, 0.2 mg ml^{-1} BSA	This study.
10 mM Tris-HCl [pH 8.0], 1 mM $MgCl_2$, 1 mM DTT, 1.0 mg ml^{-1} BSA	This study.
8 mM Tris-HCl [pH 7.9], 50 mM NaCl, 6 mM $MgCl_2$, 1 mM β-Me	Ricchetti *et al.*, 1998, [8] and this study.
10 mM Tris-HCl [pH 7.9], 200 mM KCl, 10 mM $MgCl_2$	Singer and Wu, 1987, [9] and this study.
10 mM Tris-HCl [pH 7.9], 50 mM KCl, 1 mM $MgCl_2$	Singer and Wu, 1988,[17] and this study.
20 mM HEPES [pH 7.8], 100 mM KCl, 1 mM $MgCl_2$, 50% sucrose, 0.5% β-Me, plus an O_2 scavenging system	Harada *et al.*, 1999, [10] and this study.
[a]20 mM Tris-HCl [pH 8.0], 100 mM KCl, 10 mM $MgCl_2$, 0.1 mM DTT, 0.4 mg ml^{-1}casein	Kabata *et al.*, 1993, [11] and this study.
20 mM Tris-HCl [pH 7.0], 5 mM KCl, 5 mM $MgCl_2$, 1 mM β-Me	Guthold *et al.*, 1999, [12] and this study.

[a]The concentration of Tris-HCl was not reported in Kabata *et al.*, and we selected 20 mM for our assay conditions.

Nature Structural & Molecular Biology: doi:10.1038/nsmb.2472

Supplementary Table 4. Diffusion coefficient values.

Protein	Diffusion coefficient ($nm^2\ sec^{-1}$)	Reference
lac repressor	$4.6(\pm 1.0) \times 10^4$	Elf *et al.*, 2007.[2]
p53	$1.62(\pm 0.15) \times 10^5$	Tafvizi *et al.*, 2011.[18]
Mlh1-Pms1	$1.43(\pm 0.29) \times 10^5$	Gorman *et al.*, 2010.[6]
[a,b]dig-QD $(3-sec)$	52.0 ± 25.25	This study.
[a,c]RNAP $(\tau_2; \sim 3-sec)$	96.66 ± 28.5	This study.
[a]RNAP $(\tau_3; 3-sec)$	49.78 ± 64.80	This study.
[a]RNAP $(\tau_3; 6-sec)$	17.44 ± 17.91	This study.
[a]RNAP $(\tau_3; 9-sec)$	14.8 ± 9.33	This study.
[a]RNAP $(\tau_3; >9-sec)$	15.76 ± 2.25	This study.
[a,b]dig-QD $(>9-sec)$	1.32 ± 0.593	This study.

[a] The diffusion coefficients we report of RNAP cannot be interpreted as diffusion of the proteins along the DNA. This is because diffusion coefficients in the small value range are dominated by DNA fluctuations rather than protein movement, and small value diffusion coefficients are also subject to large sources of error (see Supplementary Notes, Supplementary Fig. 7 & Supplementary Fig. 8).

[b] The dig-QD diffusion coefficients were obtained from particle tracking data collected for either 3-seconds or >9-seconds, as indicated. The difference between the diffusion coefficient values obtained for dig-QD $(3-sec)$ and dig-QD $(>9-sec)$ does not reflect differences in the behavior of the dig-QDs, rather it reflects the greater precision of the diffusion coefficient measurements obtained for particle tracking data collected over longer time intervals. For the same reason, the τ_3 diffusion coefficients decrease as the data collection windows increase from 3-seconds to >9-seconds.

[c] Note that the τ_2 molecules exhibited an exponentially distributed lifetime. Therefore these diffusion coefficients were obtained from the analysis of particle tracking data sets comprised of trajectories of differing time lengths, whereas all other data sets were built from tracking data spanning the indicate time intervals.

Nature Structural & Molecular Biology: doi:10.1038/nsmb.2472

Supplementary Notes

RNA Polymerase Purification and Characterization. Cells for expressing a chromosomal copy of RNA polymerase that harbors a biotinylation peptide tag on the C-terminus of the β' subunit were generously provided by Dr. Robert Landick (University of Wisconsin-Madison),[19] and RNAP holoenzyme was expressed, purified, and characterized as previously described.[4]

We sought confirmation that RNAP remained functional under the dilute conditions necessary for single molecule measurements. For this, we adapted a gel shift assay for quantifying promoter binding activity under dilute protein conditions. The active RNAP concentration under the single molecule experiment conditions was determined using a gel shift assay by titrating RNAP into reactions containing a fixed amount of promoter DNA, as previously described (Supplementary Fig. 9a).[20] To enhance detection, these assays utilized a Cy3-label 249–bp DNA fragment containing promoter λP_R, which was made by PCR using λ phage DNA as a template with the following primers: Cy3 (5'- Cy3–GGC CTT GTT GAT CGC GCT TT -3', 5'- CGT GCG TCC TCA AGC TGC TCT T -3', IDT). Varying amounts of purified RNAP (0.1 – 1.6 nM) were incubated with 0.4 nM of the Cy3-labeled λP_R DNA fragment in buffer (20 mM Tris [pH 8.0], 25 mM KCl, 1 mM $MgCl_2$, 1 mM DTT, and 0.2 mg ml^{-1} BSA) at room temperature for 40 minutes. Heparin (10 μg ml^{-1}) was then added to disrupt non-specifically bound RNAP and closed complexes, and the reactions were resolved on native 5% polyacrylamide gels to separate the free and bound DNA. The gels were then scanned for Cy3 fluorescence using Typhoon FLA 9000 (GE Healthcare), and the fractions of bound and free DNA were quantified using ImageQuantTL software (GE Healthcare). Under these reactions conditions, the formation of open complex is essentially irreversible, and fractional activity of RNAP capable of forming stable open complexes is revealed as the inverse of the saturation point in the titration curve (Supplementary Fig. 9b).[20]

To verify that RNA polymerase remained active over time after dilution, a time series of RNAP promoter-binding activity using the gel shift procedure described above. Briefly, RNAP was diluted to a final concentration of 0.6 nM in buffer (20 mM Tris [pH 8.0], 25 mM KCl, 1 mM $MgCl_2$, 1 mM DTT, and 0.2 mg ml^{-1} BSA) and the diluted samples were then incubated at room temperature for the indicated time intervals. The activity of the diluted RNAP was measured using a gel shift assay as described above. As shown in Supplementary Fig. 9c, the activity of diluted RNAP did not change significantly over a 40-minute period. Our single molecule measurements are typically completed within ≤15 minutes of diluting the RNAP stock solutions.

Promoter Dissociation Kinetics. A key feature of the τ_2 and τ_3 events was that they coincided with the locations of the known phage promoters (Fig. 1 & Fig. 2c).[4] The assignment of τ_2 and τ_3 events was made by inspection of reaction trajectories, and this assignment was facilitated by the drastic difference in the observed lifetimes for these two types events. Molecules assigned as τ_2 events dissociated from the DNA during collection of real time videos, whereas those assigned as τ_3 events did not dissociate from the DNA during the typical 200-second observation windows (Fig. 2). To determine the lifetime of the τ_2 events, we measured the times over which the individual proteins remained bound to the DNA before dissociating, and resulting data was fit to a single exponential decay. RNA polymerase molecules assigned as τ_3 events (*i.e.* promoter-bound open complexes) exhibited lifetimes that greatly exceeded the typical lengths of the videos that were used to monitor DNA binding in real time. Therefore dissociation of promoter-bound open complexes (*i.e.* τ_3 events) was measured in separate experiments by injecting 1 nM QD-RNAP into the flow cell to bind the promoters in buffer (20 mM Tris [pH 8.0], 25 mM KCl, 1 mM $MgCl_2$, 1 mM DTT, and 0.2 mg ml^{-1} BSA). After an incubation of 3 minutes, free QD-RNAP was removed from solution by flushing the sample chamber with buffer lacking RNAP. The number of promoter-bound QD-RNAP molecules was then monitored versus time over a 2-hour period with data collected at 5-minute intervals. The resulting data were then corrected for broken DNA molecules, and fit to a single exponential decay (Fig. 2c).

Nature Structural & Molecular Biology: doi:10.1038/nsmb.2472

Promoter Association Kinetics. For association rate measurements, we developed a new type of sparse DNA curtain that ensures experimental measurements are made under conditions where the DNA is effectively at the infinite dilution limit, such that the binding of RNA polymerase molecules on individual DNA molecules can be regarded as independent process. The patterns were made by electron-beam lithography, as described,[21] but the pattern geometry was altered so that the surface was much more sparsely populated with double-tethered DNA molecules (Supplemental Fig. 10).

We next determined the concentration profile in the flowcell using an injection rate of 0.5 ml min^{-1}. A 700 µl sample of 1nM QDs was loaded into the sample loop and the syringe pump was then pre-run for 1 minute at 0.5 ml min^{-1}. The sample was then injected without stopping the syringe pump. Images were recorded continuously and the QD signal intensity in the sample chamber of the flowcell was measured over time. The concentration of QDs in the sample chamber relative to the concentration of the injected sample (1 nM) was then calculated based on the volume of the sample loop and the flow rate. This concentration profile was then used in the kinetic measurements of promoter association rates to calculate the actual concentration of QD-RNAP present in the sample chambers.

To begin a measurement, 700 µl of diluted QD-RNAP was loaded into the sample loop. To avoid pH changes due to the oxygen scavenger system, the GLOXY components were added to the buffer immediately prior to use; imaging buffer stored in a closed syringe will maintain a constant pH for than 1 hour after the addition of the oxygen scavenging system. To avoid any dead response time from the syringe pump, the pump was pre-run at a flow rate of 0.5 ml min^{-1} for 1 min. The QD-RNAP sample was then injected while maintaining a constant flow rate of 0.5 ml min^{-1}, and data acquisition was initiated 10 sec after sample injection. Exactly 30 sec after the injection, the buffer flow was terminated, ensuring that the concentration of QD-RNAP in the sample chamber remained constant after this time point; the selection of the 30 sec stopping point was based on the concentration profiles of our flowcells under this sample injection regime. During this procedure, imaged were continually collected at either 5- or 10-Hz for a period of up to 15 minutes.

The resulting kinetic data were analyzed to determine the association rates based upon the initial binding observed within the field-of-view after the concentration of QD-RNAP in the sample chamber has plateaued. For each individual DNA molecule, we first determine the time to initial binding (t_i), which we define as the time when the first promoter-bound QD-RNAP open complex is detected, conditioned upon its subsequently forming an open complex, and for the purpose of this analysis we defined the open complex as anything with a lifetime more than 40 sec. Binding events prior to the 30 sec time point after initial sample injection were excluded to ensure that any measured association events occurred only after the concentration of QD-RNAP in the flowcell had plateaued. Once the first promoter was occupied by an open complex, any subsequent events occurring on the same DNA molecule were excluded from analysis such that the resulting data set describes only the initial binding events that occurred under conditions where all of the λ-phage promoters were unoccupied and accessible for binding, in accordance with the theoretical calculations. For all DNA molecules we then build a data set of the promoter initial association times $\{t_1, t_2, \cdots, t_N\}$, where each of these values represents the time it took for the first promoter to be occupied on each DNA within the field-of-view. The data were then sorted into ascending order $\{t_{(1)}, t_{(2)}, \cdots t_{(N)}\}$ such that $t_{(1)} \leq t_{(2)} \leq \cdots \leq t_{(N)}$. The residual waiting times between binding events were then calculated as $\aleph_m = t_{(m+1)} - t_{(m)}$ and a corrected time for each measured event was determined as $t_C = \aleph_m(N - m)$(Supplementary Figure 11). We then determined the average association time $\tau = \langle t_C \rangle$ over the entire data set at each tested concentration of RNA polymerase (Fig. 3d). The primary advantage of this analysis is that it eliminates the need to definitively establish a zero time point prior to an initial binding event, so long as the concentration of QD-RNAP in the flowcell remains isotropic, because all calculations are based on residual waiting times between initial binding events on the different DNA molecules in the sample chamber. Therefore, in the absence of a mechanical stop-flow device, this analysis of residual waiting times is more reliable than more typical kinetic measurements of bimolecular processes that are based upon cumulative occupation probability fitting to get the reaction time.

Nature Structural & Molecular Biology: doi:10.1038/nsmb.2472

Collision frequencies and nonspecific DNA binding events. In experiments conducted at either 5 or 10 Hz, we readily detected the binding of QD-RNAP to the promoters within λ-phage DNA, but we could not detect any appreciable binding to nonspecific sites on the same DNA molecules (see Fig. 3a, upper panel). This observation indicated we were missing collisions between the QD-RNAP and the nonspecific DNA because these collisions should have occurred at equal frequency everywhere along the DNA, and also indicated the lifetimes of the nonspecific complexes must have been substantially shorter than the 100–200 msec integration time for data collected at 10 or 5 Hz, respectively. To detect nonspecific binding, we increased the data acquisition rate to 100 Hz. We also sought to measure the collision frequencies for the QDs and QD-RNAP complexes at 100 Hz, and then compare these measured rates to calculated values for the collision frequency, in order to determine whether we were capturing all potential collisions with the DNA in the single molecule experiments.

First, the intensity threshold necessary to distinguish camera noise from actual QD excursions within the evanescent field was determined for 100 Hz data sets. The EMCCD was set to frame transfer mode and an AOI (63 × 1 pixels), and one DNA molecule was imaged for 10,000 frames at an acquisition rate of 100 Hz in the absence of any QDs. A histogram of the resulting signal intensities corresponded to background noise, which dropped dramatically above an intensity of ~2000 (A.U.). From the histogram we calculate that the probability of camera background noise beyond this threshold is ~ 9.5×10^{-5}, therefore we selected a threshold value of 2040 (A.U.). Next, QDs (150 pM) were injected into the flow cell, and data were collected as described above using the same EMCCD settings. Comparison of the two histograms revealed signal intensities that exceeded the threshold values for the camera noise, and these values were scored as QD excursions into the detection volume that is defined by the penetration depth of the evanescent field (350 nm).

Next a control experiment was conducted to determine how frequently QDs alone (in the absence of RNAP) entered the detection volume, and how much time they spent within this volume before diffusing back out into bulk solution. To count the number of events (see below), a kymograph was made from data collected at 100 Hz, as described above, and a QD detection event was defined as any signal within the kymograph that exceeded the defined threshold for EMCCD noise. The lifetime of these collision events was well described by a single exponential function with τ_0 =5.68 msec (95% confidence interval [5.56, 5.78]). This lifetime corresponds to the time QDs spend within the detection volume before diffusing back into free solution.

The same experiment was then perfomed using QD-RNAP in order to measure the lifetime of interactions between RNAP and DNA. QD-RNAP (150 pM) was injected into the flow cell and data were collected at 100 Hz, exactly as described above. In 100 seconds, 1251 events exceeding the noise threshold were recorded along the length of the DNA. The resulting data were best described by the sum of two exponential functions with τ_0 =5.58 msec (5.46, 5.74) and τ_1 =29.23 msec (24.53, 36.18) (Fig. 3 & Supplementary Fig. 2b-c). The first time constant τ_0 is the same as the QD-only control (Supplementary Fig. 2a-b), and was also found for control measurements made in the absence of DNA (Supplementary Fig. 2e-f), therefore does not arise from a polymerase-specific interaction with the DNA. We conclude that τ_1 corresponds to the lifetime of a nonspecific interaction between DNA and RNAP.

The number of expected non-specific interactions between QD-RNAP and DNA can then be estimated as follows below. The ratio of τ_0 and τ_1 events is given by:

$$r = \frac{\int_0^\infty A_0 e^{-k_0 x} dx}{\int_0^\infty A_1 e^{-k_1 x}\, dx} = 13.4$$

where $A_0, A_1, k_0,$ and k_1 are the amplitude and rate constants of the two components of observed events with QD-RNAP; note that $k_0 = 1/\tau_0$ and $k_1 = 1/\tau_1$. The protein samples used for single molecule imaging are subject to dilution while passing through the microfluidics and additional loss can also occur through nonspecific adsorption to the sample tubes, injection needles, tubing components, etc. Therefore

Nature Structural & Molecular Biology: doi:10.1038/nsmb.2472

the actual concentration of protein within the sample chamber may not be the same as the concentration that was injected. To correct for the difference due to dilution in the microfluidics, the actual versus injected concentration of QD-RNAP was determined by injecting a fixed volume (50 µl) of QDs (200 pM) into the sample chamber at a defined flow rate (0.05 ml min^{-1}) while continuously monitoring the bulk fluorescence signal through the microscope objective. The resulting signal versus time curve was normalized to define the QD concentration profile as a function of these defined injection parameters. Therefore the number of non-specific interaction events between DNA and RNA polymerase is given by:

$$\frac{N}{(1+r)\times 0.15nM\times 0.35\times 55\%}=\frac{1251}{6.8\times 0.15nM\times 0.35\times 55\%}=6371$$

per 100 sec per nM, where 0.35 is the ratio of actual concentration versus injection concentration, and 33% is the percentage of active protein. This corresponds to an observed non-specific interaction between RNAP-QD and DNA at a frequency ~10.03 sec^{-1} (at 200 pM RNA polymerase).

To verify whether we were capturing all of the potential collision events from the data collected at 100 Hz, we next calculated the expected collision frequency for comparison to the experimentally measured values. For this purpose we define two volumes, $V_{R\prime}$ and V_R, where V_R is defined by the region $R \geq r \geq 0$, and $V_{R\prime}$ is defined by the region $R' \geq r \geq 0$, where R is the experimentally detectable volume surrounding the DNA and R' is chosen such that the total volume contains a single particle, *i.e.* it satisfies the following equation, where l is the length of DNA and C is the concentration of protein:

$$R' = (\pi C l)^{-\frac{1}{2}}$$

We then have an equation that relates the rate of QD-RNAP diffusing out of each volume to the radial extent of each region through the law of mass action:

$$k_{RR\prime} = \frac{V_{R\prime}}{V_R} k_{R\prime R} = k_{R'R}\left(\frac{1}{\pi C l R^2} - 1\right)$$

Given an experimentally observed lifetime of QDs within the volume defined by V_R, we then calculated an expected value for the collision frequency. For an experimentally measured lifetime of $\tau_0 = 5.68$ msec (see above), at a QD-RNAP concentration of 150 pM, with an excitation radius of 178-nm, we find an expected collision frequency of 13.8 sec^{-1}, which is in excellent agreement with the experimentally observed value of ~10.03 sec^{-1}. The good agreement between the experimentally observed collision frequency and the calculated collision frequency indicates that we are capturing most, if not all, potential collision events at a data acquisition rate of 100 Hz.

Diffusion Coefficients and DNA Fluctuations. In single molecule experiments, the diffusion coefficient is commonly determined by performing a linear regression of the mean square displacement (Ψ), against time. In Michalet,[22] it was noted that for a given length of trajectory, diffusion coefficient (D), frame rate Δt, and level of measurement noise, σ, there exists an optimal point to perform the regression, which minimizes the error in the estimation of $D_{1,obs}$. We applied this method to our 1D system and calculated 1D-diffusion coefficients for τ_2 and τ_3 events. A full derivation can be found for the 2-dimensional case in Michalet,[22] and here we present only the necessary equations to extend the calculations of the variance and covariance of Ψ to a 1-dimensional system:

$$\langle \Psi_n^2 \rangle - \langle \Psi_n \rangle^2 =$$

Nature Structural & Molecular Biology: doi:10.1038/nsmb.2472

$$\begin{cases} -\frac{4}{3(N-n)^2}\left(D^2 n\Delta t^2(n+5n^3-2N-4n^2N)+12Dn\Delta t(n-N)\sigma^2+3\sigma^4(-1+4n-3N)\right) & \text{for } 2n<N \\ \frac{4}{3(n-N)}\left(D^2\Delta t^2\left(11n^3+N-17n^2N-N^3+n(-5+7N^2)\right)-12Dn\Delta t\sigma^2-6\sigma^4\right) & \text{for } 2n\geq N \end{cases}$$

and

$$\langle \Psi_n \Psi_m \rangle - \langle \Psi_n \rangle \langle \Psi_m \rangle =$$

$$\begin{cases} \frac{1}{3(n-N)(N-m)}\left(4D^2n\Delta t^2\left(6m^2n+\left(n^2-1\right)(n+2N)-2m\left(3nN+n^2-1\right)\right)+48Dn\Delta t\sigma^2(m-N)+6(-1+4m+2n-4N)\sigma^4\right) & \text{for m}+n<N \text{ and} m>n \\ \frac{4D^2\Delta t^2}{3(N-n)}\left(m-m^3+4n-4m^2n-6n^3+N\left(-1+3m^2+8mn+6n^2\right)-N^2(3m+4n)+N^3\right)-4\left(\frac{4Dn\Delta t\sigma^2-\sigma^4}{n-N}\right) & \text{for m}+n\geq N \text{ and} m>n \end{cases}$$

For convenience we also present the equations for the mean and dispersion of a least squares regression fit to *P* points, respectively.

$$\langle B(\Psi)\rangle = \frac{N\sum_{n=1}^{P}\langle n\cdot\Psi_n\rangle - \sum_{n=1}^{P} n \sum_{n=1}^{P}\langle\Psi_n\rangle}{N\sum_{n=1}^{P} n^2 - (\sum_{n=1}^{P} n)^2}$$

and

$$\langle B(\Psi)^2\rangle - \langle B(\Psi)\rangle^2 = \left(N\sum_{n=1}^{P} n^2 - \left(\sum_{n=1}^{P} n\right)^2\right)^{-2}$$

$$\begin{pmatrix} N^2\sum_{n,m=1}^{P} nm\langle\Psi_n\Psi_m\rangle - \langle\Psi_n\rangle\langle\Psi_m\rangle - N\sum_{n=1}^{P} n\sum_{n,m=1}^{P} m\langle\Psi_n\Psi_m\rangle - \langle\Psi_n\rangle\langle\Psi_m\rangle \\ -N\sum_{m=1}^{P} m\sum_{n,m=1}^{P} n\langle\Psi_n\Psi_m\rangle - \langle\Psi_n\rangle\langle\Psi_m\rangle - (\sum_{n=1}^{P} n)^2\sum_{n,m=1}^{P}\langle\Psi_n\Psi_m\rangle - \langle\Psi_n\rangle\langle\Psi_m\rangle \end{pmatrix}$$

Using these equations, we performed a check of predicted values versus values calculated from simulation experiments for both $\langle\Psi_n^2\rangle - \langle\Psi_n\rangle^2$ and $\langle B(\Psi)^2\rangle - \langle B(\Psi)\rangle^2$, and for each set of parameters, 1000 simulation experiments were performed and the average values obtained. The appearance of a minimum in $\langle B(\Psi)^2\rangle - \langle B(\Psi)\rangle^2$ leads to the method by which we should calculate $D_{1,obs}$, as outlined:

Algorithm to calculate best estimate of $D_{1,obs}$:

1. Perform a LSF (least squares fit) to n% of Ψ, and calculate the diffusion coefficient.
2. Using the diffusion coefficient from step one, determine the optimal point at which to perform a regression, and recalculate the diffusion coefficient.
3. Repeat step two until convergence.
4. Repeat entire process again from additional starting points to ensure global convergence.

Applying the above analysis to the tracking data for individual RNAP molecules resulted in three populations. The first group, ~84 percent of trajectories in this study, converged to a single optimum fit. The second group, ~11 percent, exhibited two or three 'optimums' in close proximity to one another. In this case, choosing between the values was arbitrary. In the final group multiple 'optimums' were measured which were sizably different. For these cases we based our decisions on the output of the convergence loop, and a holistic view of the particular data set.

Nature Structural & Molecular Biology: doi:10.1038/nsmb.2472

The ensemble mean squared displacement $\tilde{\Psi}$ is calculated from a number of individual trajectories, and for notation purposes we introduce two variables: Q as an index over trajectories, and T_n, which is defined below:

$$T_n = \sum_{j=1}^{Q} (N_j - n)$$

Then $\tilde{\Psi}$ may be calculated as:

$$\tilde{\Psi}_n = \frac{1}{T_n}\left(\sum_{j=1}^{Q}\sum_{i=1}^{N_j-n}\left(x_{n+i}^{(j)} - x_i^{(j)}\right)^2\right)$$

The variance and covariance of $\tilde{\Psi}$ are calculated as in the single trajectory case, just appropriately weighted by the number of terms that originate from each set:

$$\langle\tilde{\Psi}_n^2\rangle - \langle\tilde{\Psi}_n\rangle^2 = \frac{1}{T_n^2}\sum_{j=1}^{Q}(N_j - n)^2\left(\langle\Psi_n^2\rangle_{(j)} - \langle\Psi_n\rangle_{(j)}^2\right)$$

and

$$\langle\tilde{\Psi}_n\tilde{\Psi}_m\rangle - \langle\tilde{\Psi}_n\rangle\langle\tilde{\Psi}_m\rangle = \frac{1}{T_nT_m}\sum_{j=1}^{Q}(N_j - n)(N_j - m)\left(\langle\Psi_n\Psi_m\rangle_{(j)} - \langle\Psi_n\rangle_{(j)}\langle\Psi_m\rangle_{(j)}\right)$$

We can then determine the dispersion of slopes resulting from a LSF analysis of $\tilde{\Psi}$, which is identical to the single trajectory case aside from the use of $\tilde{\Psi}_n$ in place of Ψ_n, and these relationships were also verified with synthetic data.

The above analysis presents a rigorous means of obtaining diffusion coefficients for experimental particle tracking data, however, one must still recognize that when diffusion coefficients are very small (on the order of the noise and error), or when the trajectories are very short (N<10 frames) negative diffusion coefficients can begin to emerge (see Supplementary Fig. 7). Diffusion coefficients are obtained from correlation functions, so that any error in calculation of the positions will propagate through to the final value. The first is there is a localization error, which encompasses DNA fluctuations (see below), as well as camera noise (including frame averaging), florescence fluctuations, and Gaussian fitting errors. This error is systematic at short time separation due to the DNA fluctuations. Secondly, the mean squared displacement is known to carry significant statistical error from two sources: (*i*) The mean squared displacement time average from a single trajectory only equals the *ensemble* average in the limit that the trajectory is infinite (see below), so only when a trajectory is infinitely long, is the apparent diffusion coefficient obtained from that trajectory precisely correct; and (*ii*) there is also a correlated error, which comes about from overlaps in calculation of displacements. Readers should be directed to Qian *et. al.*, and X. Michalet for in depth discussion.[22,23]

The sources of error described above can be considered negligible when the diffusion coefficients are relatively large, as is the case for proteins that can diffuse long distances on DNA. However, when the diffusion coefficients are in the small-valued range the DNA fluctuations contribute substantially to the obtained values. Detailed analysis of such small-value diffusion coefficients would require the development of equations that implicitly include the DNA fluctuations in the density functions used to calculate the diffusion coefficients, which is beyond the scope of our current work. Instead, we analyzed the motion of quantum dots that were covalently linked to the DNA via a directed antibody interaction,[7] and use these results to qualitatively assess the contribution of DNA fluctuations to the $D_{1,obs}$ obtained from promoter-bound closed and open complexes (dig-QD; Fig. 3e, Supplementary Fig. 8,

Nature Structural & Molecular Biology: doi:10.1038/nsmb.2472

Supplementary Table 4).

We begin by recognizing that the probability density function for 1-D Brownian motion in the presence of Gaussian localization error can be shown to be the following:[6]

$$P(x, t+\Delta t|x', t) = \frac{1}{\sqrt{2\pi}\sqrt{2D_{1,obs}\Delta t + 2\sigma^2}} \exp\left[-\frac{(x-x')^2}{4D_{1,obs}\Delta t + 4\sigma^2}\right]$$

Multiplication by $(x-x')^2$ and integration reveals that, for a stationary particle, $D_{1,obs} = 0$, the MSD should be a straight line at $2\sigma^2$. Furthermore, detailed calculations reveal that measured MSD values are gamma distributed about their true means.[6] That is to say, when $D_{1,obs}$ is small, it is common to attain MSD values in the range $[0,2\sigma^2)$. Since the diffusion coefficient is determined through a linear fit to MSD values, it is not irregular to obtain a negative measured value for $D_{1,obs}$ such as when early time points produce values in the range $[2\sigma^2, \infty]$, and later time points yield values on $[0,2\sigma^2)$. This, of course is also the source of positive values of $D_{1,obs}$ for stationary particles (see Supplementary Fig. 7).

For a completely stationary particle, the MSD plots should be independent of time, and while the resultant curves for the QD-labeled DNA do exhibit time independence at long time separation, at short times there is clearly some motion, as revealed in the rise of the MSD plots at early time points (Supplementary Fig. S8b). Errors resulting from the camera and fitting functions are not correlated in time, and won't induce this kind of time dependent behavior in the MSD. However, the motion of the DNA will produce time dependent effects if the fluctuations occur on a time scale comparable to the QD motion (Fig. 8c). The MSD plots yield a straight line corresponding to $2\sigma^2$, yielding values for σ of ~39 nm^2 sec^{-1}, which we ascribe to the underlying DNA fluctuations (Fig. 8c). Importantly, this dig-QD data set reflects an ensemble of approximately 34,000 – 58,000 individual diffusive steps (as indicated, Supplementary Fig. 8b), and thus reflects what can be considered the average noise of the system arising from the movement of the DNA collected under ideal conditions. As indicated above, in the case of small value diffusion coefficients obtained from smaller data sets (such as those arising from shorter time trajectories for QD-RNAP), one can expect a large variation in $D_{1,obs}$ even for stationary particles. Based on these results for stationary dig-QDs, we conclude that our results are upwardly biased by the DNA chain motions, and that the diffusion coefficients obtained for the promoter-bound RNAP molecules assigned as closed or open complexes do not likely reflect motion of the proteins along the DNA, but rather reflect motion of the DNA itself.

Lac repressor experiments. FLAG-tagged lac repressor protein was expressed, purified, and labeled with anti-FLAG quantum dots, as previously described.[4] Double-tethered DNA curtains were made using a version of λ-DNA containing either a single 21-bp symmetrical operator sequence (Fig. 4),[4,24] or a 5x tandem repeat of the same operator sequence (Supplemental Fig. 5). We have previously shown that QD-tagged lac repressor co-localizes with the operator site, and dissociated rapidly in the presence of IPTG, as expected.[4] Target search experiments (Fig. 4 & Supplemental Fig. 5) were conducted in low ionic strength buffer containing 10 mM Tris-HCl (pH 8.0), 1 mM $MgCl_2$, 1 mM DTT, 1 mg ml^{-1} BSA, using the indication amounts of QD-tagged lac repressor (0.1–1 mM). Proteins were injected into the sample chamber and data were collected at 10 Hz. For each target-binding event, two position measurements were collected. First, the position of the protein was recorded in the first frame in which it appeared. Subsequently, the position of the particle was measured after it had bound stably at the target for a time greater than or equal to 1 min. The difference between these two points, Δx, was then determined. For $\Delta x \leq 100nm$ (three standard deviations of noise) events are scored as direct collisions. Furthermore, $\Delta x > 100nm$ corresponds to facilitated transport to the operator. The initial binding location of non-specific events that did not result in target capture (failed searches) were also collected, provided the binding event occurred prior to target engagement by the successful protein (*i.e.* the target was still unoccupied). For the failed searches, Δx was calculated by measuring the initial binding location of the

Nature Structural & Molecular Biology: doi:10.1038/nsmb.2472

failed searcher relative to the location of the operator.

T7 RNAP experiments. The gene for T7 RNAP was fused to a C-terminal AviTag (GGGLNDIFEAQKIEWHE) and cloned into the vector pTBX3 (NEB). The vector was transformed into BL21 cells, which were then grown in LB (1L) containing carbenicillin, and induced at an OD_{600}~0.8 with 0.8 mM IPTG. After 4 hours of induction, cells were harvested by centrifugation and resuspended in buffer containing 20 mM Tris-HCl, pH 8.0, 1 M NaCl, 1 mM EDTA, plus a Halt protease inhibitor cocktail (Pierce). The cell paste was then flash frozen on liquid nitrogen and stored at -80˚C until lysis. For lysis, the cells were thawed at room temperature, lysed by sonicatation, and the lysate was clarified by centrifugation. The clarified lysate was loaded onto a 10-ml Chitin bead column (NEB), and washed extensively with buffer containing 20 mM Tris-HCl, pH 8.0, 1 M NaCl, 1 mM EDTA, following the manufactures protocol. The column bed was then quickly flushed with 20 mM Tris-HCl, pH 8.0, 1 M NaCl, 1 mM EDTA, plus 50 mM DTT, and incubated at 4°C for ~20 hours. The protein was then eluted and dialyzed into T7 RNAP storage buffer (50 mM Tris-HCl [pH 7.9], 100 mM NaCl, 20 mM β-mercaptoethanol, 1 mM EDTA, 50% glycerol, 0.1% triton X-100) at 4°C overnight. Protein activity was tested by *in vitro* run off transcription assays. Single-molecule experiments using double-tethered DNA curtains were conducted exactly as described for *E. coli* RNAP, under the indicated buffer conditions (see Supplemental Fig. 5).

Mechanisms of Diffusion-Controlled Reactions. To address the rate at which diffusion-controlled processes occur, we begins with the descriptions of colloidal aggregation developed by Smoluchowski,[25] where the rate, k_{smol}, of the reaction $A + B \xrightarrow{k_{smol}} AB$ can be shown to be proportional to the sum of the Stokes-Einstein diffusion coefficients, D_3, of the two reactants.

$$k_{smol} = 4\pi\rho\left(D_{3,A} + D_{3,B}\right)\left(1 + \mathcal{O}\left(t^{-1/2}\right)\right) \qquad \text{(Eq. 1)}$$

Here, ρ is the reaction radius, which will be defined below. This relation is commonly cited as the upper limit on the speed at which a diffusion-controlled reaction may occur.[26,27] However, early measurements of the rate at which the lac repressor associates to its operon sequence exceeded this limit by two orders of magnitude.[28] This paradox of faster-than-diffusion association was resolved for site-specific association of proteins by including mechanisms of lower dimensionality: hopping, sliding, and intersegmental transfer, in what has been termed the facilitated diffusion model (Supplementary Fig. 1).[26] With facilitated diffusion, the association rate has the form $k = k_{smol}(1 + \zeta)$ where ζ depends on the dissociation rate of the protein from non-specific DNA, the diffusion coefficient of the protein on the surface of DNA, and the protein concentration.

In the simplest terms, a facilitated association process of a protein to its cognate sequence consists of three states: (*i*) a free state, (*ii*) a non-specifically bound state, wherein the protein is bound to non-target DNA, and (*iii*) a specifically bound state, where the protein has located and bound target DNA. In general, the search process that a protein undergoes consists of cycling through the non-specifically bound and free states until eventually locating the target. When the concentration of available non-specific states outnumbers specific states, this process is slow. The "facilitation" occurs due to two factors. First, the affinity of proteins for non-target stretches of DNA localizes the protein to the DNA for extended periods of time, allowing for many successive rebinding events. Second, when the protein is able to translocate along the DNA during its time in the bound state, it may interrogate multiple sites during a single binding event.

Reaction Radius & Target Size. The theoretical framework of for facilitated diffusion separates the bound and free states at a distance ρ, which is termed the reaction radius. The motion of the protein beyond this distance is expected to be free thermal diffusion in three dimensions. While, within ρ the protein's motion is constrained to only allow movement along the dimension of the DNA (Fig. 3a-d). This motion is

Nature Structural & Molecular Biology: doi:10.1038/nsmb.2472

expected to be Brownian as well, however the diffusion coefficient must now include the average effect of the potential along the DNA as well as the viscous forces from solution. The reaction radius is then dependent on the size of the protein and the DNA, as well as the ionic strength of the solution, as it describes the point at which one can disregard the gradient of the radial portion of the electrostatic potential of DNA. Here, ρ is chosen to be the sum of the radii of the searching protein and the DNA plus the Debye screening length, r_{db}, under our reaction conditions

Investigations of the sequence-specific binding of proteins commonly consist of foot printing and mutation methods, revealing both the extent of the protein-DNA interface and minimal consensus sequences. However, a fundamental question remains: over what range can the protein be out of register with the target sequence and still recognize it (Fig. 3b)? If we move the protein 1-bp to the left or right of a perfectly centered target, is the protein heavily biased toward registered binding or does the protein act as if it is has been placed on a random sequence for which it has no preference? If the answer is the former, then this begs the question, exactly how far out of register is it necessary to move the protein until the latter is true? This is the concept of the linear target size, which we will term a. If the position z_0 defines when the protein is perfectly in register with the target sequence, then whenever the proteins position, z, satisfies $z_0 - \frac{a}{2} \leq z \leq z_0 + \frac{a}{2}$, it recognizes the target sequence and rapidly moves to z_0.

Above we considered the effect of lateral displacement on target recognition, we must also consider the effect of the protein's angular orientation with respect to the target (Fig. 3a). The importance of orientation arises due to the fact that the entire surface of the protein does not carry out the function of sequence recognition. For example, consider a protein as a Janus particle, where half of the surface recognizes DNA sequences, and the other half does not. If we consider the search process to consist of only proteins colliding with the DNA from solution, it should be clear that half of the particles which collide with the target sequence will recognize it, as the other half would have encountered the DNA in an unproductive orientation. While, it is difficult to consider the motion of a protein about its own axis in calculation, the effect of orientation can easily be accounted for by altering the target size. For the example above, an effective target size (ψ) equal to half of the size of the usual size, $\psi = \frac{a}{2}$, would account for the non-binding surface, while allowing every encounter to be productive. In the calculations that follow, we will treat the protein-quantum dot complex as a sphere. The portion of this sphere corresponding to the reactive portion of the protein is then defined by the half-angle, θ, subtended by the area of an equivalent circle of the sphere. Then, the above condition for recognition becomes $z_0 - \frac{\psi(a,\theta)}{2} \leq z \leq z_0 + \frac{\psi(a,\theta)}{2}$.

While the mapping of the orientation of the protein onto the target size accounts for the probability of successful collisions between the protein and target DNA when the protein originates in bulk solution, it underestimates the probability of locating the target via sliding mechanisms. If $\widetilde{D}_1$ is the one-dimensional diffusion coefficient in the reduced system, it is then straightforward to recover the usual one-dimensional diffusion coefficient, $D_{1,\rho} = \widetilde{D}_1 \frac{a^2}{\psi^2}$.

Association Rate to the DNA from Solution. We consider the DNA to be initially void of bound protein and immersed in an isotropic distribution of RNAP molecules at concentration C_0. Then, the initial (first encounter) association rate of proteins to the DNA is identical to the flux of proteins across an absorbing cylinder of radius ρ and length L, where $L = 48{,}502$-bp. This flux can be found from the solution to the radial diffusion equation, subject to the following boundary conditions.

$$C(r,0) = C_o \qquad radius > \rho$$
$$C(\rho,t) = 0$$
$$C(\infty,t) = C_0$$

The Laplace transformed solution satisfying these boundary conditions is given by:

Nature Structural & Molecular Biology: doi:10.1038/nsmb.2472

$$C(r,s) = \frac{C_0}{s}\left(1 - \frac{K_0\left(r\sqrt{s/D_3}\right)}{K_0\left(\rho\sqrt{s/D_3}\right)}\right) \qquad \text{(Eq. 2)}$$

Where the Laplace transform is defined as $f(s) = \int_0^\infty e^{-st} f(t)dt$, and $K_0(z)$ is the modified Bessel function of the second kind. The solution to the above in the time domain can be written as:[29]

$$C(r,t) = \frac{2C_0}{\pi}\int_0^\infty e^{-D_3 u^2 t}\frac{J_0(\rho u)Y_0(ru) - J_0(ru)Y_0(\rho u)}{u\left(J_0^2(\rho u) - Y_0^2(\rho u)\right)}du \qquad \text{(Eq. 3)}$$

To determine the rate of association per unit length, we find the flux (ϕ) of proteins across the boundary at ρ, and then integrate this flux over the entire surface of the DNA:

$$k_\alpha(t) = 2\pi\rho D_3 \frac{d}{dr}C(r,t)|_{r=\rho} = \frac{8}{\pi}D_3 C_0 \int_0^\infty e^{-D_3 u^2 t}\left[u\left(J_0^2(u\rho) + Y_0^2(u\rho)\right)\right]^{-1}du \qquad \text{(Eq. 4)}$$

From the above we also find $k_\alpha^\psi(t) = \psi k_\alpha(t)$. In general, we will make use of the Asymptotic solutions of $C(r,s)$ by considering the small argument (long time) and the large argument (short time) expansions of $K_0(z)$. Note, here we have scaled time as $\tau = \frac{D_3 t}{\rho^2}$, which leads to:[29]

$$k_\alpha(t)_{t\to 0} = 2\pi D_3 C_0\left[\frac{1}{\sqrt{\pi\tau}} + \frac{1}{2} - \frac{1}{4}\sqrt{\frac{\tau}{\pi}} + \frac{1}{8}\tau - \ldots\right] \qquad \text{(Eq. 5)}$$

$$k_\alpha(t)_{t\to\infty} = 4\pi D_3 C_0\left[\frac{1}{\ln(4\tau) - 2\gamma} - \frac{\gamma}{(ln(4\tau) - 2\gamma)^2} - \frac{\frac{\pi^2}{6} - \gamma^2}{(ln(4\tau) - 2\gamma)^3} + \ldots\right] \qquad \text{(Eq. 6)}$$

Effect of Protein Concentration on Association Rates & Calculation of Effective Target Size. The hallmark of facilitated diffusion, is that the overall association rate can be greatly accelerated by the mechanisms of sliding and hopping we have described. Notably, the magnitude of this effect is proportional to the concentration of reactants. That is, the acceleration, which may be present at lower protein concentrations, vanishes as the concentration increases. To see this consider the flux to the operator to be comprised of three terms: the first, k_α, is described above, the second k_h, the hopping rate into the promoter, and the third, k_s, is the sliding rate into the promoter. Notably, while k_α is proportional to the initial concentration, k_h and k_s are also both proportional to the concentration but they are additionally scaled by the nonspecific lifetime, and in the case of k_s, by the 1D diffusion coefficient. Conceptually, the domination of k_α over the association rate is then easily inferred from the limiting cases of the non-specific lifetime and 1D diffusion coefficient. For example, when the dissociation rate is zero, (*i.e.* the DNA is infinitely sticky), hopping cannot exist. Then the concentration at which hopping is effectively eliminated from the association rate corresponds to establishing $k_\alpha \gg nk_h$, where n is the average number of hops to the target. To estimate the limiting rate of association, recall that the average number of initial binding events per unit length up to a time t is given by.

$$\langle n(t)\rangle = \frac{8}{\pi}C_0\int_0^\infty \frac{1 - e^{-D_3 u^2 t}}{J_0^2(u\rho) + Y_0^2(u\rho)}\frac{du}{u^3} \qquad \text{(Eq. 7)}$$

Nature Structural & Molecular Biology: doi:10.1038/nsmb.2472

Now, if there are N promoter sites, each of an effective length ψ on the DNA, then the probability of randomly choosing one of these sites is $N\psi/L$. That is to say, on average, it takes $L/N\psi$ random collision events until a promoter is found. We then ask for the time $\bar{t}$ such that $\langle n(\bar{t})\rangle = L/N\psi$. If we again let $\tau = tD_3/\rho^2$, then for each value of C_0, D_3, ρ, N, and ψ, there is a $\bar{\tau}$ such the following equality is true.

$$\frac{1}{\psi} = \frac{8}{\pi} N\rho^2 C_0 \int_0^\infty \frac{1-e^{u^2\bar{\tau}}}{J_0^2(u)+Y_0^2(u)} \frac{du}{u^3} \qquad \text{(Eq. 8)}$$

When the concentration reaches a value necessary for $k_\alpha(t)$ to be the predominant contribution to the association rate, the above calculation yields the effective target size. Furthermore, at any concentration higher than this value, the above continues to give the same result for ψ. However, at lower concentrations, this calculation will over estimate ψ, due to the combined influence of hopping and sliding. Traditionally, this would be referred to as the "antenna" effect.[8,27]

Comparison to Previous Single-molecule Promoter Search Studies. The work of Kabata *et al.* is often cited as evidence for long-distance 1D-diffusion of *E. coli* RNA polymerase along DNA.[11] However inspection of the data presented in Kabata *et al.* shows that the reported single molecule trajectories were not diffraction limited fluorescent spots, as would be expected for single molecules of RNAP. We surmised that these data might have reflected behavior of large aggregates of RNA polymerase. An RNAP aggregate would have numerous DNA binding sites, and the collective effect of these binding sites could cause an aggregate to appear to slide on DNA. To test this hypothesis, we saturated large (1.0–μm dia.) streptavidin-coated beads (Chemicell GmbH, Cat. No. 2205-1) with biotinylated RNAP, and asked whether these artificial mimics of an RNAP aggregate could slide on DNA. As shown in Supplementary Fig. 6, the RNAP-coated 1.0–μm beads were observed moving along the DNA by 1D-diffusion and could also be pushed in 1D along the DNA when flow was applied, confirming that an aggregate of RNA polymerase might display apparent sliding behavior. Alternatively, Kabata *et al.* could not define the number of DNA molecules that gave rise to each observed sliding event, and the DNA belts they were using were 2-3 μm thick (10-300 mg ml^{-1}), so the apparent sliding they observed may have arisen as the cumulative outcome of multiple nonspecific binding events involving numerous DNA molecules.

The work of Guthhold *et al.* also presented evidence for 1D diffusion of RNA polymerase along nonspecific DNA. In this study, the authors used AFM to image RNAP bound to nonspecific DNA adsorbed onto a mica surface, and reported a value for $D_{1,obs}$ of $1.1\text{x}10^1$ $nm^2\,s^{-1}$ and a nonspecific lifetime of 600 seconds.[12] The authors of this study concluded that the extraordinarily long 600-second lifetime of RNA polymerase bound to nonspecific sites was likely a consequence of both the DNA and the protein being absorbed to the mica surface, as was necessary for the AFM measurements. In this scenario, the exceedingly small diffusion constant that Guthold *et al.* report for nonspecifically bound RNA polymerase is in full agreement with our data, and we infer that they detected 1D-diffusion because of the extraordinarily long lifetime of the nonspecifically bound complexes adsorbed to the mica surface.

Harada *et al.* studied DNA binding using Cy3-tagged *E. coli* RNAP and λ-phage DNA held suspended above a surface by a dual optical trap, and concluded that 1D diffusion may contribute to the promoter search.[10] In this study the authors reported different lifetimes for RNAP bound to either the AT-rich or GC-rich halves of the λ-DNA, which correspond to the side of λ containing all of the promoters and the side lacking promoters, respectively. For the AT-rich half, they reported lifetimes of 330–msec and 1.5–sec, and for the GC-rich half they reported lifetimes of 120–msec. However, they did not detect a population of proteins consistent with open complexes, and they suggested that inability to detect open complexes was due to the relatively high 5 pN of tension on the DNA. They also demonstrated a drastic increase in binding at lower DNA tensions, therefore it seems plausible that the application of 5 pN of tension may have also altered the lifetimes of the other binding intermediates, thus yielding different values than reported in our study. Moreover, the lifetime of 120–msec reported by Harada *et al.* for the nonspecifically bound intermediate was 4-fold higher than the 30–msec upper bound we have placed on

Nature Structural & Molecular Biology: doi:10.1038/nsmb.2472

this lifetime. Harada *et al.* also reported that 10 out of 381 RNAP molecules (2.6%) underwent 1D diffusive motion detectable above instrument resolution (0.2 μm). Notably, the experiments of Harada *et al*. were conducted in 50% sucrose,[10] and the high viscosity of this buffer ($\eta_{50\%suc}/\eta_w = 15.4$) may have artificially prolonged the lifetime of the nonspecifically bound intermediates by reducing the 3D diffusion coefficient of RNAP and restricting its ability to diffuse away from the DNA upon dissociation, or through the increase osmotic stress, which is commonly reflected as an increase in the affinity for nonspecific DNA relative to specific DNA sites[30-35]; either effect would have led to an overestimate of 1D-sliding. More importantly, our results show that even if the protein were able to slide, 3D-diffusion will still dominate the promoter search mechanism at physiologically relevant protein concentrations regimes.

Influence of DNA tension. DNA wrapping or bending by RNA polymerase should be antagonized under tension, and as such might be expected to perturb open complex formation. However, this effect should occur at much higher DNA tensions than are used in our study. The DNA in our experiments is typically stretched to ~75% its contour length, and the tension on the DNA can be estimated based upon relative mean extension using the Worm-Like Chain model:

$$F = \frac{k_B T}{L_p}\left[\frac{1}{4\left(1 - {\langle x\rangle}/{L}\right)^2} - \frac{1}{4} + \frac{\langle x\rangle}{L}\right]$$

Where L_p is the DNA persistence length (53 nm), $k_B T$ is termal energy, ⟨x⟩ is the mean observed extension, L is the full contour length of the DNA (16.49 μm for λ-DNA), and F is the calculated tension force.[36] Based on this calculation the DNA in our experiments experiences a tension force of ~0.35 pN. The total free energy change required for site-specific binding under tension for a protein that bends DNA can then be estimated as:

$$\Delta\Delta G(F) = \frac{20 \cdot a}{K} F^2 + 20 \cdot a\left(1 - \cos\left(\frac{\theta}{2}\right)\right) F + \frac{\theta^2}{4}\sqrt{L_p k_B T}\sqrt{F}$$

Where θ is the bend angle of the DNA molecule induced by the bound protein, K is the DNA stretch modulus (~1,200 pN), and $20 \cdot a$ is the length of the bound site times the length of an unperturbed base pair ($20\ bp \cdot 0.34\ nm/bp$).[37] This equation predicts a simple, linear relationship between the energetic cost of bending the DNA as a function of applied tension: for a protein with bend angle of 60˚ and binding site of 20-bp, on DNA under 0.35 pN of tension, $\Delta\Delta G(F) \approx 0.078\ k_B T$. Based on these rough calculations, the low tension experienced by the stretched DNA in our experiments should have little or no impact on promoter binding by RNAP, and much higher tension forces than those that are accessible by simple flow-stretched DNA experiments would be required to substantially perturb the binding of proteins that bend the DNA. Note that the value for $\Delta\Delta G(F)$ changes from ≈ 0.004 to $\approx 0.4\ k_B T$ for binding site sizes ranging from 1-bp to 100-bp, and bend angles of 45˚ and 90˚ (for a 20-bp binding site) yield values of $\approx 0.05\ k_B T$ and $\approx 0.77\ k_B T$, respectively; so our conclusion that DNA bending will not be impacted at low tension holds true for a range of site sizes and bend angles. The conclusion that low DNA tensions experienced in our curtain assays should not drastically impact binding is reflected by our data in that RNAP recognizes and binds to the promoters on the extended DNA substrates, the lifetime we obtained for open complex formation closely match literature values, and RNAP moves along the DNA when provided with all four rNTPs. In addition, RNA polymerase can transcribe against applied forces of

Nature Structural & Molecular Biology: doi:10.1038/nsmb.2472

up to ~14-25 pN,[38,39] which again suggests that the relatively low tension used in our assays should have little or no impact upon the proteins ability to bind promoters.

Supplementary References

1. Wang, Y., Austin, R. & Cox, E. Single molecule measurements of repressor protein 1D diffusion on DNA. *Phys Rev Lett* **97**, 048302 (2006).
2. Elf, J., Li, G. & Xie, X. Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* **316**, 1191 - 4 (2007).
3. Kim, J.H. & Larson, R.G. Single-molecule analysis of 1D diffusion and transcription elongation of T7 RNA polymerase along individual stretched DNA molecules. *Nucleic Acids Res* **35**, 3848-58 (2007).
4. Finkelstein, I., Visnapuu, M. & Greene, E. Single-molecule imaging reveals mechanisms of protein disruption by a DNA translocase. *Nature* **468**, 983-7 (2010).
5. Tafvizi, A., Huang, F., Fersht, A., Mirny, L. & van Oijen, A. A single-molecule characterization of p53 search on DNA. *Proc Natl Acad Sci U S A* **108**, 563 - 8 (2011).
6. Gorman, J., Plys, A., Visnapuu, M., Alani, E. & Greene, E. Visualizing one-dimensional diffusion of eukaryotic DNA repair factors along a chromatin lattice. *Nat Struct Mol Biol* **17**, 932 - 8 (2010).
7. Visnapuu, M.-L. & Greene, E. Single-molecule imaging of DNA curtains reveals intrinsic energy landscapes for nucleosome deposition. *Nat Struct Mol Biol* **16**, 1056-1062 (2009).
8. Ricchetti, M., Metzger, W. & Heumann, H. One-dimensional diffusion of Escherichia coli DNA-dependent RNA polymerase: a mechanism to facilitate promoter location. *Proc Natl Acad Sci U S A* **85**, 4610 - 4 (1988).
9. Singer, P. & Wu, C. Promoter search by Escherichia coli RNA polymerase on a circular DNA template. *J Biol Chem* **262**, 14178 - 89 (1987).
10. Harada, Y. et al. Single-molecule imaging of RNA polymerase-DNA interactions in real time. *Biophys J* **76**, 709 - 15 (1999).
11. Kabata, H. et al. Visualization of single molecules of RNA polymerase sliding along DNA. *Science* **262**, 1561 - 3 (1993).
12. Guthold, M. et al. Direct observation of one-dimensional diffusion and transcription by Escherichia coli RNA polymerase. *Biophys J* **77**, 2284 - 94 (1999).
13. Fong, R., Woody, S. & Gussin, G. Direct and indirect effects of mutations in lambda PRM on open complex formation at the divergent PR promoter. *J Mol Biol* **240**, 119 - 26 (1994).
14. Mita, B., Tang, Y. & deHaseth, P. Interference of PR-bound RNA polymerase with open complex formation at PRM is relieved by a 10-base pair deletion between the two promoters. *J Biol Chem* **270**, 30428 - 33 (1995).
15. Hoopes, B. & McClure, W. A cII-dependent promoter is located within the Q gene of bacteriophage lambda. *Proc Natl Acad Sci U S A* **82**, 3134 - 8 (1985).
16. Hershberger, P. & deHaseth, P. RNA polymerase bound to the PR promoter of bacteriophage lambda inhibits open complex formation at the divergently transcribed PRM promoter. Implications for an indirect mechanism of transcriptional activation by lambda repressor. *J Mol Biol* **222**, 479 - 94 (1991).
17. Singer, P. & Wu, C. Kinetics of promoter search by Escherichia coli RNA polymerase. Effects of monovalent and divalent cations and temperature. *J Biol Chem* **263**, 4208 - 14 (1988).
18. Tafvizi, A. et al. Tumor suppressor p53 slides on DNA with low friction and high stability. *Biophys J* **95**, L01 - 3 (2008).
19. Shaevitz, J., Abbondanzieri, E., Landick, R. & Block, S. Backtracking by single RNA polymerase molecules observed at near-base-pair resolution. *Nature* **426**, 684 - 7 (2003).
20. Roe, J.H., Burgess, R.R. & Record, M.T., Jr. Kinetics and mechanism of the interaction of Escherichia coli RNA polymerase with the lambda PR promoter. *J Mol Biol* **176**, 495-522 (1984).

Nature Structural & Molecular Biology: doi:10.1038/nsmb.2472

21. Greene, E., Wind, S., Fazio, T., Gorman, J. & Visnapuu, M.-L. DNA curtains for high-throughput single-molecule optical imaging. *Methods Enzymol* **472**, 293-315 (2010).
22. Michalet, X. Mean square displacement analysis of single-particle trajectories with localization error: Brownian motion in an isotropic medium. *Phys Rev E Stat Nonlin Soft Matter Phys* **82**, 041914 (2010).
23. Qian, H., Sheetz, M. & Elson, E. Single particle tracking. Analysis of diffusion and flow in two-dimensional systems. *Biophys J* **60**, 910 - 21 (1991).
24. Sadler, J., Sasmor, H. & Betz, J. A perfectly symmetric lac operator binds the lac repressor very tighly. *Proc Natl Acad Sci U S A* **80**, 6785-6789 (1983).
25. Smoluchowski, M. Three lectures on diffusion, Brownian motion and the coagulation of colloids. *Phys Z* **17**, 557-571, 585-599 (1916).
26. von Hippel, P. & Berg, O. Facilitated target location in biological systems. *J Biol Chem* **264**, 675 - 8 (1989).
27. Mirny, L. et al. How a protein searches for its site on DNA: the mechanism of facilitated diffusion. *J Physics A* **42**, 434013 (2009).
28. Riggs, A., Bourgeois, S. & Cohn, M. The lac repressor-operator interaction. 3. Kinetic studies. *J Mol Biol* **53**, 401 - 17 (1970).
29. Carslaw, H. & Jaeger, J. *Conduction of heat in solids*, (Oxford University Press, 1959).
30. Lynch, T. & Sligar, S. Macromolecular hydration changes associated with BamHI binding and catalysis. *J Biol Chem* **275**, 30561 - 5 (2000).
31. Garner, M. & Rau, D. Water release associated with specific binding of gal repressor. *EMBO J* **14**, 1257 - 63 (1995).
32. Parsegian, V., Rand, R. & Rau, D. Macromolecules and water: probing with osmotic stress. *Methods Enzymol* **259**, 43 - 94 (1995).
33. Sidorova, N. & Rau, D. Differences in water release for the binding of EcoRI to specific and nonspecific DNA sequences. *Proc Natl Acad Sci U S A* **93**, 12272 - 7 (1996).
34. Robinson, C. & Sligar, S. Molecular recognition mediated by bound water. A mechanism for star activity of the restriction endonuclease EcoRI. *J Mol Biol* **234**, 302 - 6 (1993).
35. Robinson, C. & Sligar, S. Changes in solvation during DNA binding and cleavage are critical to altered specificity of the EcoRI endonuclease. *Proc Natl Acad Sci U S A* **95**, 2186 - 91 (1998).
36. Bustamante, C., Marko, J., Siggia, E. & Smith, S. Entropic elasticity of lambda-phage DNA. *Science* **265**, 1599-1600 (1994).
37. van den Broek, B., Noom, M. & Wuite, G. DNA-tension dependence of restriction enzyme activity reveals mechanochemical properties of the reaction pathway. *Nucleic Acids Res* **33**, 2676-2684 (2005).
38. Wang, M. et al. Force and velocity measured for single molecules of RNA polymerase. *Science* **282**, 902-907 (1998).
39. Yin, H. et al. Transcription against an applied force. *Science* **270**, 1653-1657 (1995).

Nature Structural & Molecular Biology: doi:10.1038/nsmb.2472