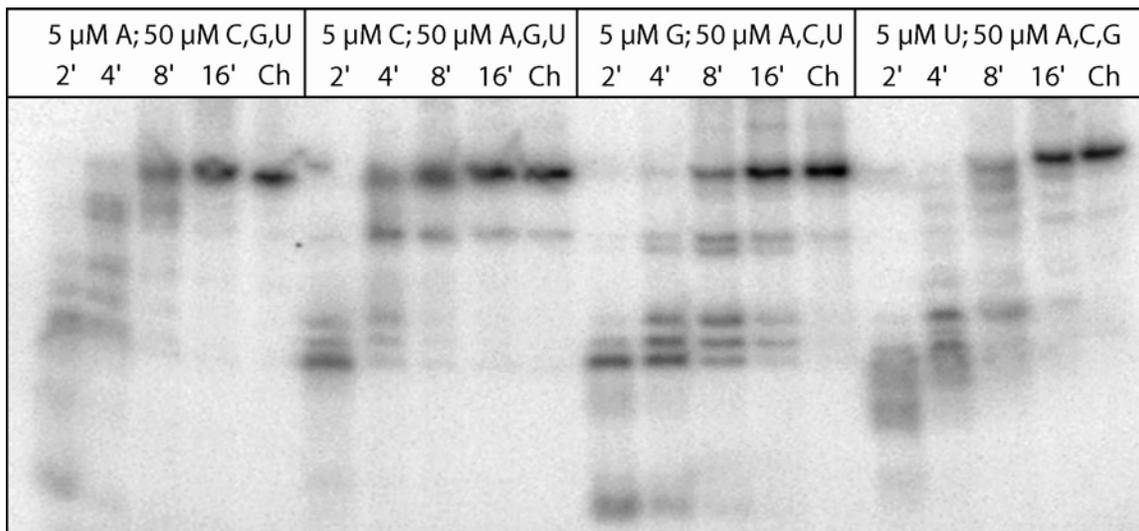
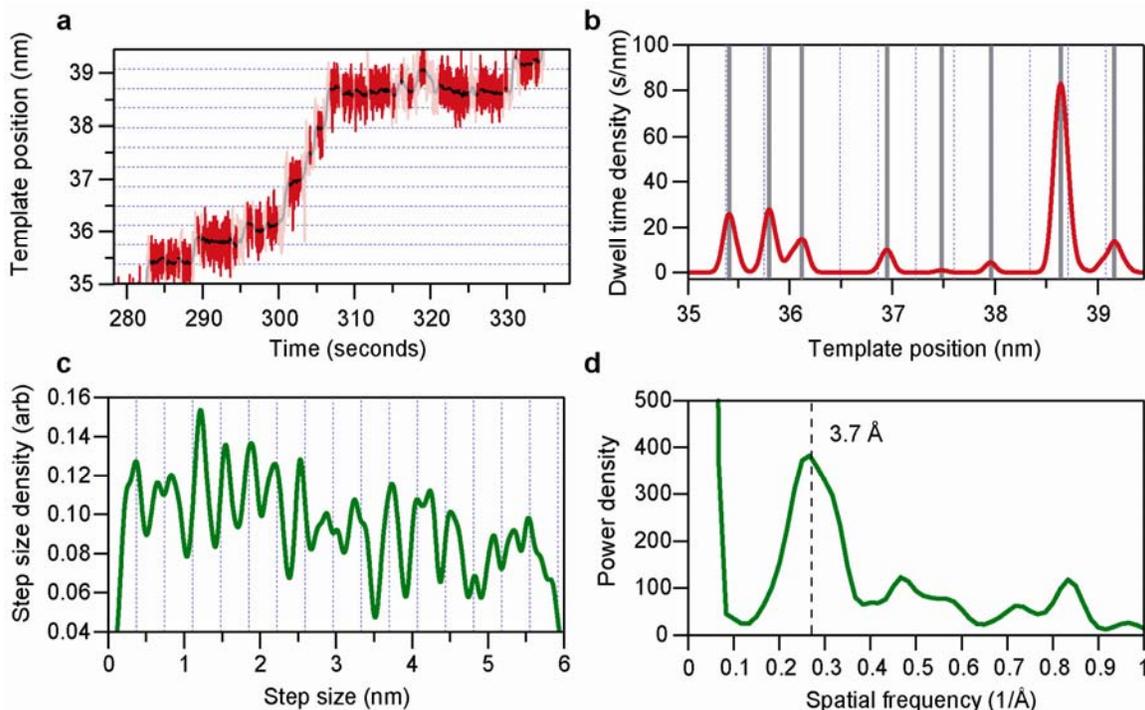


### Determination of $[NTP]_{eq}$



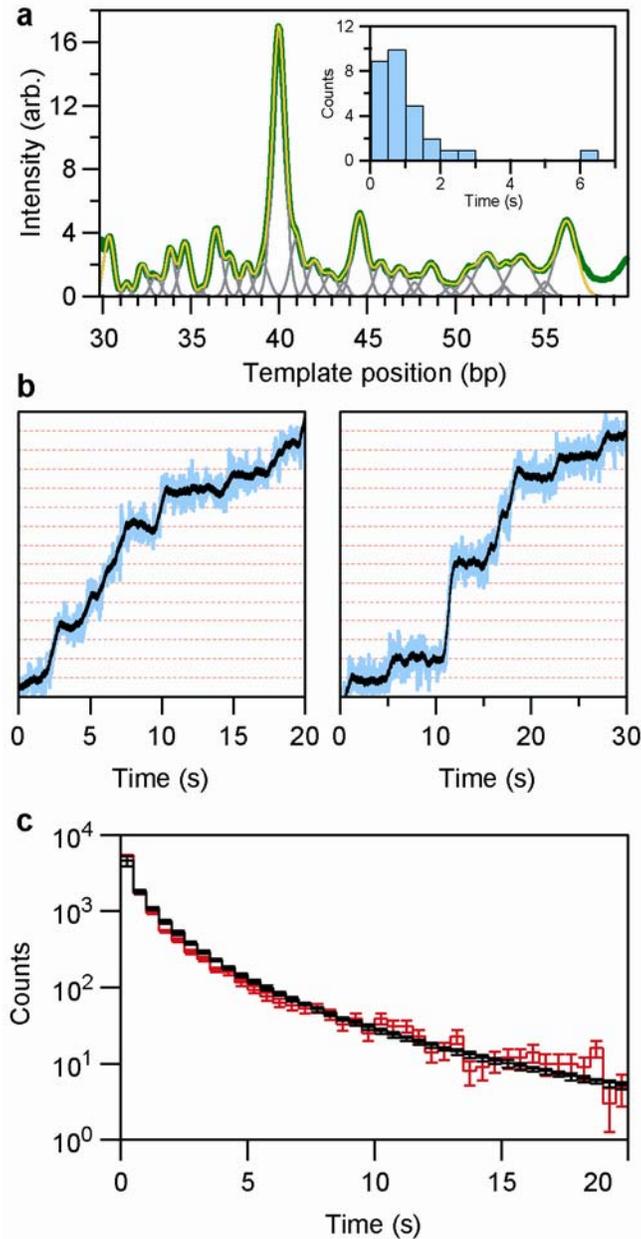
**Figure S1** | To determine the ratio of ribonucleoside triphosphate (NTP) concentrations at which each individual species becomes equally rate limiting for the overall elongation process, we proceeded as follows. First, we prepared ternary elongation complexes (TECs) that were stalled at position A29 near the start of the DNA template, as described in the text, and divided these into four aliquots. Transcription was initiated synchronously, allowing the TECs to transcribe for 2, 4, 8, or 16 min., after which the reaction was quenched and the RNA transcripts were run on the gel (lanes are indicated by the time labels). For each of the four aliquots, three of the NTPs were present at comparatively high concentration levels (50  $\mu$ M), but the concentration of the fourth NTP species (A,C, G or U) was lowered tenfold (to 5  $\mu$ M), so that its availability became rate-limiting for elongation. An additional set of chase reactions (labeled ‘Ch’ on the gel) was run with a saturating concentration of all four NTPs (1 mM), which was added to the reaction buffer after 16 minutes with transcription allowed to continue for 5 min thereafter. Rates of elongation were assayed by quantitating the amount of full-length transcript produced in each of the lanes for every time point, and fitting these data to an error function (i.e., to the integral of a gaussian, a function that models the degree of accumulation for a process involving both diffusion and drift). From the fits, we obtained the characteristic rates of addition for each of the four NTP species under these conditions: 0.67 bp/s for GTP, 0.84 bp/s for UTP; 1.29 bp/s for ATP, and 2.62 bp/s for CTP. From these four rates, we then computed the approximate NTP concentrations necessary to produce an average addition rate for any given nucleotide species of  $\sim 1$  bp/s when placed under significant assisting load (10  $\mu$ M GTP, 10  $\mu$ M UTP, 5  $\mu$ M ATP, and 2.5  $\mu$ M CTP). We defined this particular set of concentrations to be  $[NTP]_{eq}$ .

## Automated analysis of RNAP step sizes



**Figure S2** | We performed a fully-automated analysis of step size on a single, long (~300 bp; 100 nm), comparatively noise-free record containing a large number of steps. Data were acquired at 2 kHz, and decimated and smoothed as described (see Methods). To identify the positions at which stepwise transitions occurred, the running standard deviation in position was computed for a set of consecutive, overlapping data windows (window width = 75 points, or 750 ms). When this value exceeded a given threshold ( $0.32 \text{ \AA}$ ), the center point of the window was discarded; otherwise, it was retained. This procedure removes segments of records corresponding to those places where transitions occurred, or where motion was too rapid for steps to be resolved, or where any backward motion took place. **a**, To illustrate, an approximately ~5-nm long section of the processed trace is displayed, with light colors indicating regions where points were discarded (*grey; pink*), and dark colors representing points retained by the analysis (*red; black*). Horizontal lines (*dotted*) are spaced at  $3.7 \text{ \AA}$  intervals. **b**, The processed trace was then used to create a dwell-time histogram (bin size =  $0.1 \text{ \AA}$ ) which was smoothed with a 50-point binomial filter. The center of each peak in this histogram was then identified (*grey lines*). **c**, All pair-wise distances between these peaks of less than 6 nm were then computed and histogrammed (bin size =  $0.01 \text{ \AA}$ ), provided that no pair of distances in question spanned a transition where the AOD displacement had been updated (see Methods). The histogram was smoothed with a 6000-point binomial filter. **d**, The power spectrum of **c**, but with the histogram smoothed with a 1000-point binomial filter to preserve higher spatial frequencies. This shows a clear periodicity at the inverse of the a fundamental step size of  $3.7 \pm 1.5 \text{ \AA}$ . Compare with the estimate of  $3.7 \pm 0.6 \text{ \AA}$  based on selected records (Fig. 2b,c, main text).

## Simulations of single-molecule data based on gel parameters



**Figure S3** | RNAP transcriptional gel data, simulated records derived from these, and the experimental dwell-time distribution for RNAP steps. **a**, A pseudo-steady-state intensity profile derived from quantitative gel analysis of radiolabelled transcripts (*green*) was fit to a series of gaussian peaks (*grey*) which were used to estimate RNAP dwell times at various transcript lengths. The sum of the gaussian peaks provides a good fit (*yellow*). Significant variation in dwell time (peak height) is readily apparent. *Inset*: A histogram of lifetimes for each of the gaussian bands, normalized to a 1 bp/s average rate. **b**, Simulated single-molecule traces based on the gel data in **a**, filtered at 50 ms (*light blue*) and 750 ms (*black*) (see Methods). Horizontal lines (*dotted*) are spaced at 3.4 Å intervals. **c**, Dwell time distributions derived from our single molecule stepping data (*red*) with associated

counting errors, compared with the predicted dwell time distribution based on quantitative gel analysis (*black*), along with bootstrapped errors. The close correspondence between these curves suggests a tight coupling between the lengths of RNA transcripts (as measured in gels) and stepping of the RNAP enzyme along DNA (as measured through the experimental attachment point to the enzyme, located on the  $\beta'$  subunit).

**Methods for single-molecule simulations.** To obtain gel data used in stochastic simulations, stalled ternary transcription complexes were quenched at  $t = 0, 15, 30, 45, 60, 90, 120, 180, 240, 360, 600,$  and  $900$  seconds after the addition of NTPs (at  $[NTP]_{eq}$ ) at room temperature and run on a denaturing 8% polyacrylamide gel. Intensities of this gel were quantified, and positions were scaled using an NTP sequencing ladder for calibration and assuming that position on the gel varied inversely with transcript length. Band intensity profiles derived from each lane (time point) were fit to gaussians, which model the stochastic decoherence over time for an initially synchronized population of molecules. The fit gaussians were then used to normalize each intensity profile to obtain a steady-state approximation to the time spent at each gel position. Normalized intensity profiles from three time points (30, 45, and 60 s) were averaged to obtain a pseudo-steady-state intensity profile. This profile was fit to a series of narrow gaussians, one for each sequential base of transcript length ( $N = 29$ ). The areas under these gaussians supplied the relative dwell times at the corresponding transcript lengths (Fig. S3a). Areas were histogrammed and scaled such that the mean dwell time matched the overall observed elongation rate ( $\sim 1$  bp/s; Fig. S3a, inset). Single-molecule traces were simulated by choosing an average dwell time from this measured distribution for each simulated step, then drawing a specific dwell time from an exponential distribution, the mean of which corresponded to the average dwell time. Noise, derived by statistical sampling from representative records of stalled RNAP dumbbells under 18 pN of tension was added to these dwells, which were spaced  $3.4 \pm 0.5 \text{ \AA}$  apart.

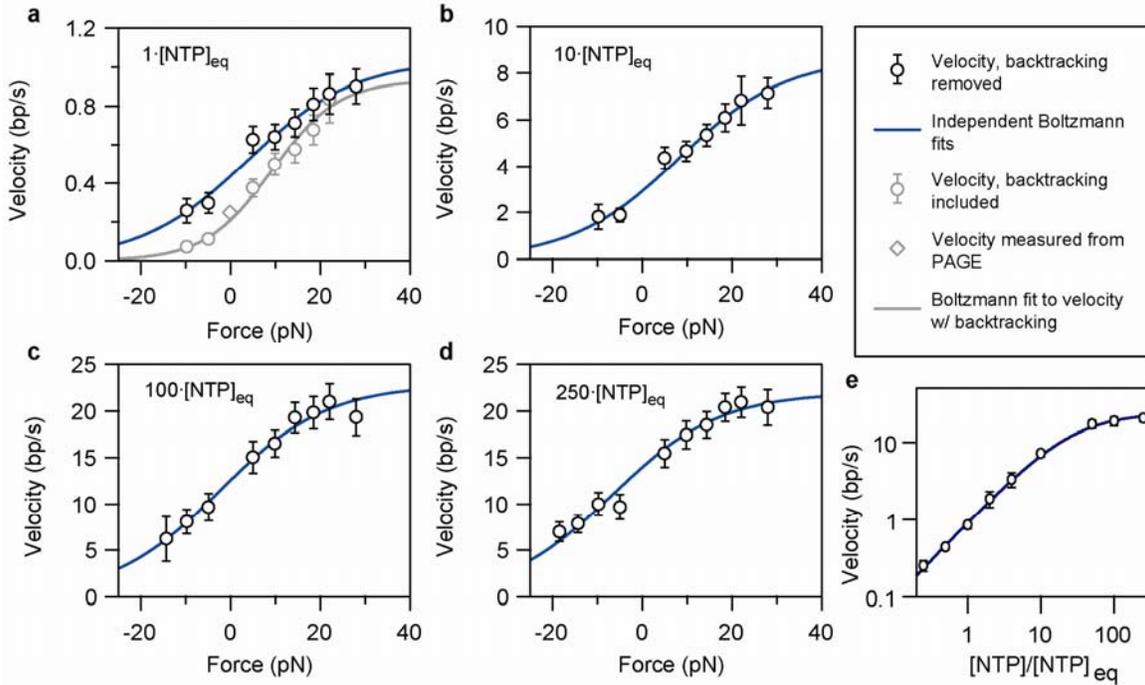
**Methods for comparison of dwell-time distributions.**

The times taken by individual RNAP enzymes to traverse successive  $3.4 \text{ \AA}$ -wide windows were plotted in histogram form to derive a global dwell-time distribution from the single molecule records, shown in red in Fig. S3c. To calculate the corresponding dwell-time distribution from gels, the set of mean dwell times derived from each of the different transcript lengths ( $N = 29$ ) was normalized to give the same average elongation rate as each of the single molecule traces ( $N = 27$ ). After normalization, 29 exponential distributions with time constants equal to these normalized dwell times were summed together to supply an estimate of the overall dwell-time distribution for each simulated trace. These 27 simulated distributions, one with the same average elongation rate as each of the single molecule traces, were then added together to give an overall estimate of the expected single molecule dwell time histogram, shown in black in Fig. S3c.

## **RNAP backsteps under assisting load and backtracks under opposing load.**

We observed infrequent rearward displacements through distances of single or multiple bases during transcription at  $[NTP]_{eq}$  against moderate (18 pN) assisting loads (Fig. 4a of main text). At the extremely low concentrations of  $PP_i$  present in these experiments ( $< 10$  nM), transcriptionally-upstream motion is more likely to be associated with a backstepping event or hypertranslocation<sup>1</sup> than with pyrophosphorolysis. In addition to single backsteps, we observed occasional backtracking by RNAP (defined as more than 3 bp of transcriptionally-upstream movement) under hindering loads of 9 pN (Fig. 4b of main text). After backtracking, polymerase could recover and continue elongation against load. Once backtracked, RNAP appeared to shuttle backwards and forwards along the template, dwelling at preferred locations, consistent with observations made previously on stalled complexes<sup>2-4</sup>: this quasi-stability suggests that the enzyme diffuses along a rough energy landscape, spending additional time in local minima while periodically exploring other regions. Such diffusional motion demonstrates the energetic contribution to translocation associated purely with local sequence (*i.e.*, with the formation and breakage of hydrogen bonds, changes in base stacking, and interactions between polymerase and specific bases) from the energetic contributions associated with NTP-binding and hydrolysis. As such, it supports the notion that local sequence information significantly affects the propensity of RNAP to translocate.

## Unconstrained Boltzmann fits to force-velocity data

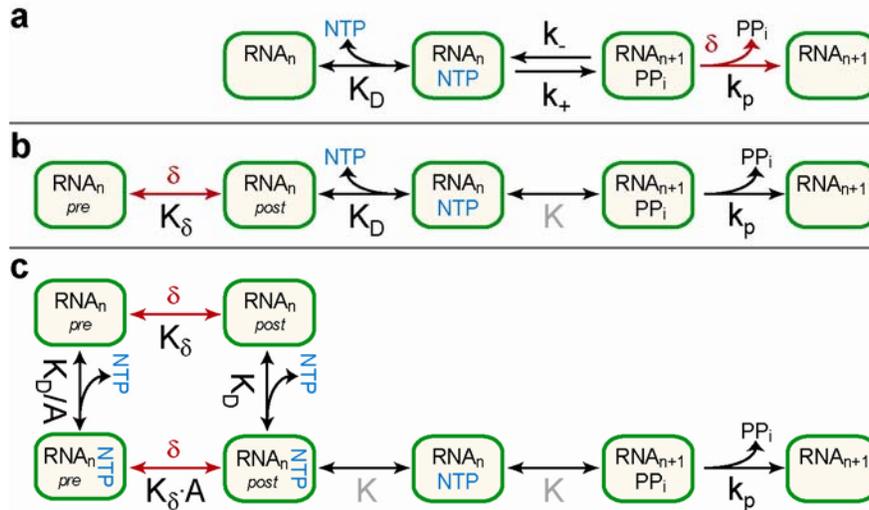


**Figure S4** | Force-velocity and Michaelis-Menten relationships with unconstrained fits to Eq. 1. **a-d**, Force-velocity relationships for RNAP at  $[NTP]_{eq}$  (**a**),  $10 \cdot [NTP]_{eq}$  (**b**),  $100 \cdot [NTP]_{eq}$  (**c**) and  $250 \cdot [NTP]_{eq}$  (**d**). Negative forces correspond to hindering loads; positive forces to assisting loads. Experimental data and Boltzmann fit at zero load, prior to the removal of backtracking events (*grey circles and line*), agree well with the velocity of elongation determined independently in a polyacrylamide gel electrophoresis (PAGE) experiment (*grey diamond*). After removal of backtracking events, as described in the main text (*open circles*), data were fit to Eq. 1 (*blue lines*), requiring the distance parameter,  $\delta$ , to be the same for all individual fits, but with no other parameter constraints. These fits returned the unconstrained values of  $v_{max}$ ,  $F_{1/2}$ , with the errors (s.e.) shown in Table 1 below. The global estimate for the distance parameter was  $\delta = 3.4 \pm 0.5 \text{ \AA}$ . **e**, Transcription velocity as a function of  $[NTP]$  for transcription under 27 pN assisting load. An independent fit to the Michealis-Menten relation for these data give  $k_{cat} = 24 \pm 2 \text{ s}^{-1}$  and  $K_M/[NTP]_{eq} = 25.5 \pm 2.9$ .

**Table 1: Unconstrained fit parameters.**

[NTP]	$v_{max}$ (bp/s)	$F_{1/2}$ (pN)
1×	$1.0 \pm 0.1$	$4 \pm 4$
10×	$9 \pm 1$	$8 \pm 4$
100×	$23 \pm 2$	$-2 \pm 3$
250×	$23 \pm 2$	$-7 \pm 2$

### Analysis of the kinetic models considered



**Figure S5** | The diagram shows the kinetic schemes and associated rate or equilibrium constants for each of the three models described in the main text. The equilibrium constants shown in grey represent parameters that are not probed by variations in either force or nucleotide concentration, and are therefore not relevant to comparisons made here. At fixed [NTP], all three of the models predict a force-velocity relationship described by the general Boltzmann relation:

$$v(F) = \frac{v_{\max}}{1 + \exp\left[-\frac{(F - F_{1/2}) \cdot \delta}{k_B T}\right]},$$

which represents Eq. 1 in the main text.

The position of the translocation step within the biochemical cycle has a dramatic effect on the force sensitivity for the rate of nucleotide addition. To understand this, it is useful to consider the behaviour of  $F_{1/2}$ , the force above which translocation ceases to exponentially affect the overall velocity. In the figure above, we show the RNAP biochemical cycle with nucleotide binding at the beginning and pyrophosphate release at the end. Pyrophosphate release is irreversible under the experimental conditions of our experiments. If a power stroke tightly coupled to PP<sub>i</sub> release were responsible for translocation (Fig. S5a), then the sensitivity to load should be greatest (i.e.,  $F_{1/2}$  maximal) at saturating nucleotide concentrations: these conditions tend to drive RNAP into the pyrophosphate-bound state and towards the load-dependent transition at the end of the cycle. Conversely, in a brownian ratchet mechanism (Fig. S5b,c) where the force-sensitive step is coupled to nucleotide binding, the force sensitivity of velocity will be

greatest at subsaturating nucleotide concentrations, where RNAP is driven into states at the beginning of the cycle, waiting for NTP incorporation. To quantify these effects, we derive below explicit expressions for velocity as a function of force and nucleotide concentration,  $v(F, [\text{NTP}])$ , in terms of the specific rate and equilibrium constants for each of the reaction cycles above.

In the irreversible power stroke model (Fig. S5a), the rate constant for phosphate release is directly associated with a displacement that carries the system over an energy barrier from its previous state (the irreversible transition marked  $k_p$  in the diagram above), and therefore carries a load dependence, as follows:

$$k_p(F) = k_p(0) \cdot e^{F \cdot \delta / k_B T} .$$

Here, the parameter  $\delta$  represents the physical distance to the transition state, which is necessarily less than the total distance translocated by the enzyme during one reaction cycle. The velocity with which RNAP moves through the complete nucleotide addition cycle at steady state can then be written as:

$$v(F, [\text{NTP}]) = \frac{k_p \cdot k_+}{k_+ e^{-F \cdot \delta / k_B T} + (k_- \cdot e^{-F \cdot \delta / k_B T} + k_p)(1 + K_D / [\text{NTP}])} .$$

In the reaction cycle for the brownian ratchet models (described in Fig. S5b,c), a rapid equilibrium exists between the pre- and post-translocated states associated with NTP binding. This process is load-dependent, and the equilibrium constant describing it is affected by force, according to:

$$K_\delta(F) = K_\delta(0) \cdot e^{-F \cdot \delta / k_B T} .$$

Here, the parameter  $\delta$  represents the physical distance between the pre- and post-translocated steps, *i.e.*, a full base pair. In Fig S5b we present a brownian ratchet model in which translocation must precede nucleotide binding to the enzyme active site. The force-velocity relationship predicted by this model can therefore be written:

$$v(F, [\text{NTP}]) = \frac{k_p}{1 + \frac{K_D}{[\text{NTP}]} (1 + K_\delta \cdot e^{-F \cdot \delta / k_B T})} .$$

In Fig. S5c, we present a brownian ratchet model in which a secondary NTP binding site can capture nucleotides either before or after translocation occurs. The force-velocity relationship predicted by this model is:

$$v(F, [\text{NTP}]) = \frac{k_p}{1 + \frac{K_D}{[\text{NTP}]} (1 + K_\delta \cdot e^{-F \cdot \delta / k_B T}) + A \cdot K_\delta \cdot e^{-F \cdot \delta / k_B T}} .$$

Here, parameter  $A$  represents the natural log of the energetic penalty associated with having an NTP in the secondary binding site while the enzyme remains in its pre-translocated position, at  $[NTP] = K_D$  and  $F = F_{1/2}$ . A value of  $A = 0$  would preclude any binding at all in this position, whereas a value of  $A = 1$  would supply no energetic penalty.

We note that the kinetic mechanism outlined in Fig. 5Sc can be used to describe any one of several structurally distinct secondary binding sites. In particular, we note that two competing models for secondary binding sites have been proposed in multi-subunit RNA polymerases. The so-called ‘E’ site, located at the base of the secondary channel, could non-specifically bind an NTP in a flipped orientation relative to the enzyme active site<sup>5</sup>. Alternatively, other kinetic studies have postulated the existence of a templated binding site which may be accessed via the main channel of the polymerase<sup>6-8</sup>. Both of these secondary binding site mechanisms introduce a new kinetic state that supplements binding to the active site, and therefore would be expected to affect velocity in a manner consistent with Fig. 5Sc.

Global fit parameters from the dataset of Fig. 5 (main text) are supplied in Table 2 below:

**Table 2: Model parameters for the fits displayed in Fig. 5 (main text)**

Model	$\delta$ (Å)	$K_\delta$	$K_D/[NTP]_{eq}$	$k_p$ (s <sup>-1</sup> )	$A$	$k_+$ (s <sup>-1</sup> )	$k_-$ (s <sup>-1</sup> )
<b>A</b>	$2.7 \pm 0.6$	—	$21 \pm 5$	$59 \pm 15$	—	$37 \pm 10$	$0.0 \pm 0.3$
<b>B</b>	3.4 (fixed)	$4.3 \pm 0.4$	$14 \pm 1$	$20.8 \pm 0.7$	—	—	—
<b>C</b>	3.4 (fixed)	$1.9 \pm 0.5$	$22 \pm 2$	$25.5 \pm 1.1$	$0.29 \pm 0.08$	—	—

## Supplemental Methods

**Transcription assays.** Construction of bead:RNAP:DNA:bead dumbbells and buffer conditions for single-molecule assays have been described previously<sup>9</sup>. For gel assays, transcription was initiated in the presence of  $\alpha$ -<sup>32</sup>P-labelled GTP using a T7 A1 promoter and stalled at A29 on a 262 bp-long template derived from the *rpoB* gene of *E. coli*.

**Selection of transcription records.** Single-molecule elongation records were characterized by levels of associated noise that varied considerably from dumbbell to dumbbell. The power spectrum of this residual noise had non-Lorentzian properties, and therefore could not be fully attributed to the thermal noise of a bead trapped in a harmonic potential. We conjecture that the additional noise may arise from any or all of several factors. Candidates include molecular motions of the DNA linkages (biotin:avidin and digoxigenin:antidigoxigenin), the possibility that additional DNA strands tethered by one end to a bead may interfere with the dumbbell, or small amounts of asphericity in the shapes of trapped beads<sup>9</sup>. Despite the heterogeneous noise, we were able to identify a subpopulation of dumbbells (roughly 10% of the total) that were particularly quiet, and these allowed for the detection of sub-nanometre motions.

### SUPPLEMENTAL MATERIAL REFERENCES

1. Artsimovitch, I. & Landick, R. Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proc Natl Acad Sci U S A* 97, 7090-5 (2000).
2. Komissarova, N. & Kashlev, M. RNA polymerase switches between inactivated and activated states By translocating back and forth along the DNA and the RNA. *J Biol Chem* 272, 15329-38 (1997).
3. Nudler, E., Mustaev, A., Lukhtanov, E. & Goldfarb, A. The RNA-DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase. *Cell* 89, 33-41 (1997).
4. Reeder, T. C. & Hawley, D. K. Promoter proximal sequences modulate RNA polymerase II elongation by a novel mechanism. *Cell* 87, 767-77 (1996).
5. Westover, K. D., Bushnell, D. A. & Kornberg, R. D. Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. *Cell* 119, 481-9 (2004).
6. Gong, X. Q., Zhang, C., Feig, M. & Burton, Z. F. Dynamic error correction and regulation of downstream bubble opening by human RNA polymerase II. *Mol Cell* 18, 461-70 (2005).
7. Holmes, S. F. & Erie, D. A. Downstream DNA sequence effects on transcription elongation. Allosteric binding of nucleoside triphosphates facilitates translocation via a ratchet motion. *J Biol Chem* 278, 35597-608 (2003).
8. Zhang, C. & Burton, Z. F. Transcription factors IIF and IIS and nucleoside triphosphate substrates as dynamic probes of the human RNA polymerase II mechanism. *J Mol Biol* 342, 1085-99 (2004).
9. Shaevitz, J. W., Abbondanzieri, E. A., Landick, R. & Block, S. M. Backtracking by single RNA polymerase molecules observed at near-base-pair resolution. *Nature* 426, 684-7 (2003)