A unifying mechanistic model of bacterial transcription with three interconnected pause states and non-diffusive backtrack recovery

Richard Janissen[*1], Behrouz Eslami-Mossallam[*1], Irina Artsimovitch[#2], Martin Depken[#1], Nynke H. Dekker[#1]

[1]*Department of Bionanoscience, Kavli Institute of Nanoscience, Delft University of Technology, 2629HZ Delft, The Netherlands*
[2]*Department of Microbiology, Ohio State University, Columbus, OH43210, United States of America*

[*]Equal contribution
[#]Correspondence may be addressed to: Artsimovitch.1@OSU.edu, S.M.Depken@TUDelft.nl, and N.H.Dekker@TUDelft.nl

KEYWORDS: transcription; RNA polymerase; transcription kinetics; backtracking; RNA cleavage; transcriptional heterogeneity; state switching; single-molecule experiments

SUMMARY

Pausing by RNA polymerase (RNAp) facilitates the recruitment of regulatory factors, RNA folding, and other related processes. While backtracking and intra-structural isomerization have been proposed to trigger pausing, the mechanisms of pause formation and recovery remain debated. Using high-throughput magnetic tweezers with single-molecule sensitivity, we have examined the full temporal spectrum of *Escherichia coli* RNAp transcription dynamics. Together with probabilistic dwell-time analysis and modelling, this has identified three distinct states that compete with elongation: a short-lived elemental pause and two long-lived backtracked pause states. We demonstrate that recovery from backtracking is not governed by diffusional Brownian motion, but rather by intrinsic RNA cleavage, whereby RNAp conformational changes hinder cleavage in long-lived pause states. We further show that state switching underlies stochastic alterations in the frequency of short pauses. Based on these findings, we propose a consensus model of intrinsic pausing that unifies all key findings while resolving earlier contradictions.

INTRODUCTION

The transcription of genetic information is the first step in gene expression and is conserved throughout all domains of life.[1] The synthesis of RNA transcripts, carried out by the DNA-dependent RNA polymerase (RNAp), is a complex and tightly controlled process that defines cell identity, fitness, proliferation, and response to environmental signals.[2–4] RNAp itself is subjected to many forms of regulation, which include interactions with accessory proteins, nucleic acid signals, and small molecules, as well as conformational dynamics in the active site and global structural changes in the transcription complexes.[5–11]

Already four decades ago, pioneering work on bacterial transcription revealed that RNA chain elongation is intermittent, and frequently interrupted by pauses that reduce overall transcription velocity.[12,13] Over the past two decades, several aspects of transcriptional pausing have been systematically studied using high-resolution single-molecule techniques to monitor RNAp dynamics in real-time. These studies have shown that chain elongation by both bacterial RNAp and eukaryotic Pol II, whose structures and action mechanisms are largely conserved, is interrupted by mainly two types of pauses: sequence-induced regulatory pauses that allow timely interaction with transcription factors and the translation machinery (in

prokaryotes); and intrinsic pauses originating from RNAp backtracking on the DNA template and RNA; the latter pauses are associated with proofreading.[14–18] Previous single-molecule studies of bacterial RNAp and eukaryotic Pol II were able to identify the ubiquitous catalytically inactive elemental pause state, *cis*-acting consensus DNA sequences and backtracking-induced interruption of active RNA chain elongation.[15,19–25]

However, the temporal signals resulting from single-molecule experiments are not trivial to interpret, and not all pause signatures observed in single-molecule studies have been satisfactorily mapped to structural intermediates of the transcription elongation complex (TEC) that are thought to cause notable transcriptional pauses or arrest.[11,16,18,21,26–29] While there exists a consensus on the properties of the elemental pause (lifetime ~1 s), intrinsic transcriptional pauses with longer lifetimes have neither been fully characterized, nor has their relation to structural intermediates of the TEC been fully elucidated.[21,22,29–33] Besides hairpin-stabilized pauses, various biochemical studies have identified two distinct exponentially distributed backtracked transcriptional pause states with notably different lifetimes (~5 s vs. ~100 s).[21,22,24,28,34–37] In contrast, the majority of single-molecule studies attributed all pauses exceeding 5 s to a single state with a broad power-law distribution of pause durations, which originate from diffusive TEC dynamics during RNAp backtracking.[15,23,36,38–40] This discrepancy extends to our understanding of the recovery of backtrack-induced transcriptional pauses: while biochemical studies indicate that RNA cleavage is the dominant recovery pathway, several single-molecule and theoretical studies consider diffusional Brownian motion of the RNAp to be the main recovery pathway.[15,23,36,38–41] Due to these divergent interpretations of transcriptional pause signatures from single-molecule and bulk experiments, a consensus mechanistic model of intrinsic transcriptional pausing remains lacking.

Previous single-molecule studies may have missed some intrinsic pause states identified in biochemical assays and structural studies simply because these were rare events within limited observation times. We thus employed high-throughput magnetic tweezers to probe the full temporal spectrum of *E. coli* RNAp transcription dynamics over a time period exceeding two hours. Transcriptional pausing analysis identified three distinct pause states competing with elongation: a short-lived pause with a lifetime of ~1s, which resembles the reported consensus elemental pause, and two long-lived, serially connected pause states, with lifetimes of ~4 s and ~100 s that branch from the elemental pause.[21,22,29–33] Our results show further that backtrack-recovery is mediated by intrinsic RNA cleavage and not not by diffusional Brownian motion, indicating that a subset of paused TECs assume a conformational state in which intrinsic and GreB-assisted RNA cleavage is hindered, delaying as such the escape to productive elongation.[25,34,40,42–46] The availability of large datasets also allowed us to interrogate the sources of the poorly understood heterogeneity in transcription velocity and pause dynamics, providing support for previously postulated state-switching that we could link to stochastic alterations in the frequency of short pauses.[24,26,29,32,47–50] We present a unified mechanistic model that integrates all key findings of previous biochemical and single-molecule studies, and describes the origin and hierarchy of intrinsic pause states as framework for future studies.

RESULTS

RNAp transcription dynamics exhibit notable heterogeneity

To investigate the full spectrum of RNA synthesis dynamics and stochastic pausing, we established a single-molecule assay (**Methods M1**) based on high-throughput magnetic tweezers with instrumental drift compensation (**Methods M2**) that provides a large data volume required for statistically robust analysis.[51,52]

Similarly to previous force-spectroscopy investigations, halted RNAp transcription elongation complexes (TECs) were formed on surface-attached linear DNA templates by nucleotide deprivation. As

template, we used the *E. coli rpoB*-coding DNA sequence free of known consensus pause-inducing sequences that could overshadow the intrinsic RNAp pausing dynamics.[24] By attaching magnetic beads to singly biotinylated RNAps and varying the DNA template orientation (**Figure 1A**), we were able to apply assisting forces (AF) or opposing forces (OF) to the RNAps while simultaneously monitoring their progression along the DNA template in real time.[52,53] Resumption of transcription was achieved by the addition of all four nucleotides in equimolar concentration, and the RNAp translocation was measured ≥2 hours at constant forces.

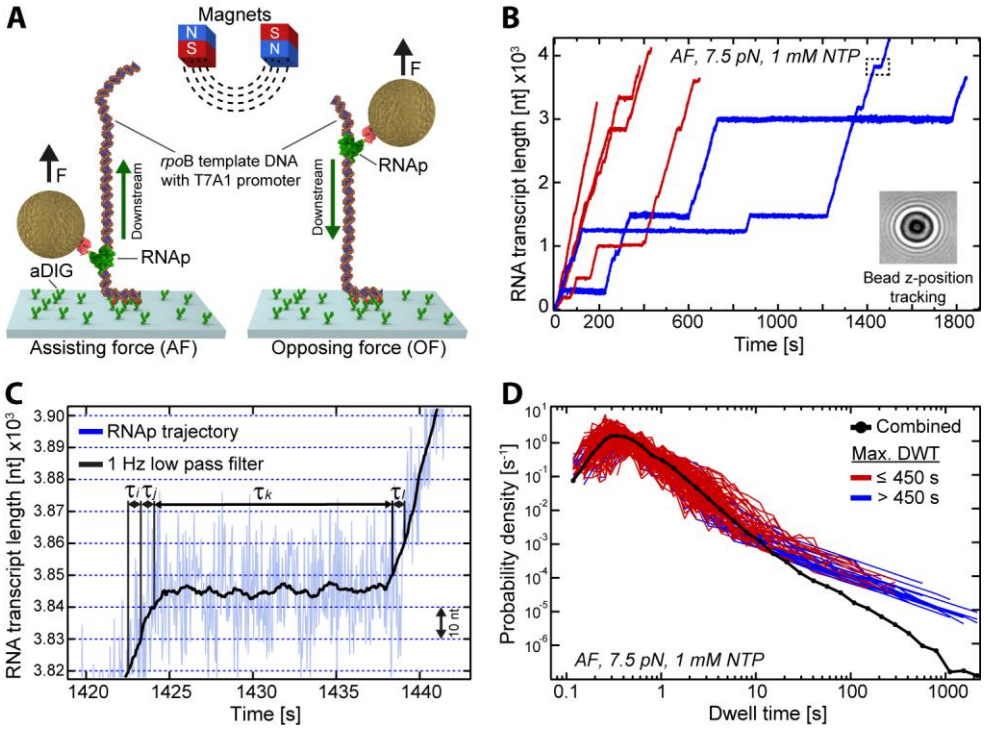

Figure 1: Multiplex measurement of single-molecule RNAp transcription dynamics. (**A**) Schematic representation of the magnetic tweezers *in vitro* transcription assay in the assisting force (AF) and opposing force (OF) configurations with respect to the RNAp downstream direction. The assay consisted of a surface-attached DNA template including a T7A1 promoter sequence, a magnetic bead bound to the singly biotin-labeled ß'-subunit of RNAp, and the magnets that exerted a constant force *F* on this bead. (**B**) Example trajectories of individual RNAps. Translocation was measured *via* the height change of the bead diffraction pattern (inset) over time. The individual trajectories are colored based on the largest pause duration found within the trajectories according to the populations identified in (D). (**C**) Magnified region of an individual RNAp trajectory corresponding to the data in the dashed rectangle in (B). Dwell times (denoted in the example with τ_i, τ_j, τ_k and τ_l) associated with RNAp advancing consecutively through 10 nt windows (blue dashed horizontal lines) were extracted from 1 Hz-filtered trajectories (black line). (**D**) Superimposition of DWT distributions from individual trajectories (blue and red) and the combined DWT distribution of all trajectories (black). The color code reflects the largest pause duration (maximum DWT) found for each individual RNAp trajectory (red: ≤ 450 s; blue: > 450 s), separated by the established selection protocol (**Methods M4**). **See also Figure S1.**

In agreement with previous studies, the RNAp trajectories (**Figure 1B**) exhibit periods of apparently constant translocation velocities, punctuated by pauses of different duration and position.[32,54] To capture the probabilistic nature of RNAp translocation and stochastic pausing, we constructed dwell-time (DWT) distributions from DWTs obtained for all trajectories.[55] These DWTs were extracted from individual trajectories (**Figure 1C**) by measuring the time needed for each individual RNAp to successively transcribe a fixed number of nucleotides (**Methods M3**). This approach allowed us to perform a probabilistic analysis on both the RNA chain elongation and on transcriptional pausing in the range of milliseconds to hours, and to extract the underlying kinetic parameters.[51,55]

Figure 1D depicts individual DWT distributions for n = 230 trajectories, measured under 7.5 pN AF at 1 mM [NTP]. All individual DWT distributions qualitatively exhibit the same features, a peak at short time scales (~ 400 ms for the provided example) and a tail of gradually decreasing probability for DWT >1 s. While the peak reflects the pause-free elongation, the tail originates from pauses.[54,55] We observe a significant variation between individual DWT distributions, particularly in the occurrence of extremely long-lived pauses. While for the majority of the analyzed RNAp trajectories the longest observed pauses lie well below 450 s (**Figure 1D,** red), a small fraction of trajectories (~ 10%) show single pauses extending to few thousand seconds (**Figure 1D,** blue). However, the combined DWT distribution (constructed by accumulating measured DWTs from all single trajectories) is not representative of all individual DWT distributions (**Figure 1D**, black). To reveal differences between these subpopulations, we used the Bayes-Schwarz information criterion (**Figure S1**) to estimate the number of single-rate pause states in the combined DWT distribution and in the DWT distributions corresponding to each subpopulation.[56,57] We found that while the pausing dynamics of each subpopulation can be captured with three single-rate pause states, the combined DWT distribution requires the incorporation of a fourth pause state, an artefact arising from combining two inherently different distributions. This result demonstrates the necessity to select a homogenous ensemble of elongation trajectories prior to the analysis. Since we found that the maximum pause duration provides a parameter that can successfully discriminate between the two subpopulations, we established a selection protocol using the Gaussian mixture model approach, which classifies individual trajectories based on the longest pause they contain (**Methods M4**).[58] We initially focus on the dominant sub-population and discuss the minority sub-population with prolonged pauses at a later stage.

Intrinsic RNAp dynamics reflects three distinct pause states

Previous studies on stochastic pausing of RNAp have identified a consensus short elemental pause (lifetime ~ 1s), which can act as an intermediate state for less frequent "stabilized" pauses with extended lifetimes.[21,22,29–33] Since the measured combined DWT distribution is incompatible with the simple power-law decay predicted when all pauses originate in diffusive returns to elongation from backtracking (red line in **Figure 2A**), and the Bayes-Schwarz information criterion analysis further supports the existence of three distinct pause states (**Figures 2A, S1B, S2A**; **Methods M5, M6**), we propose a working model (**Figure 2**) that includes a frequent short pause state (P1) and two additional rare pauses (P2, P3) of extended duration. The catalytic pathway is characterized by an effective elongation rate k_{el}, which contains the contribution of all sub-processes such as nucleotide addition, NTP hydrolysis, PP_i release, and translocation. The three pause states are connected sequentially, jointly forming a single off-pathway branch. The probability of leaving the catalytic pathway and entering the off-pathway branch through P1 is described by P_{total}. Within the off-pathway branch, the entry into the pause states P2 and P3 is governed by the rates k_{P2} and k_{P3}, respectively, and all pauses return to the catalytic pathway through the exit rates k_{e1}, k_{e2}, and k_{e3}.

We have trained this model and validated its predictive power against data collected under different empirical conditions, as described in the following sections. As a first step, we parametrized the working model by fitting it to a combined DWT distribution, constructed from a large (361.3 kb of total transcript length) reference ensemble of elongation trajectories measured under 7.5 pN AF at a saturating [NTP] of 1 mM. Under these conditions, the characteristic pause times (**Figure 2A**; determined by the green, blue, and purple "shoulders") are well-separated from the pause-free elongation peak (**Figure 2A**; yellow peak), which allows for accurate maximum-likelihood estimation of all parameters (**Methods M6**; **Figure 2B,C**; **Table S1**). [55] The parameter values estimated from the reference data set will be referred to as reference parameters in all subsequent analyses.

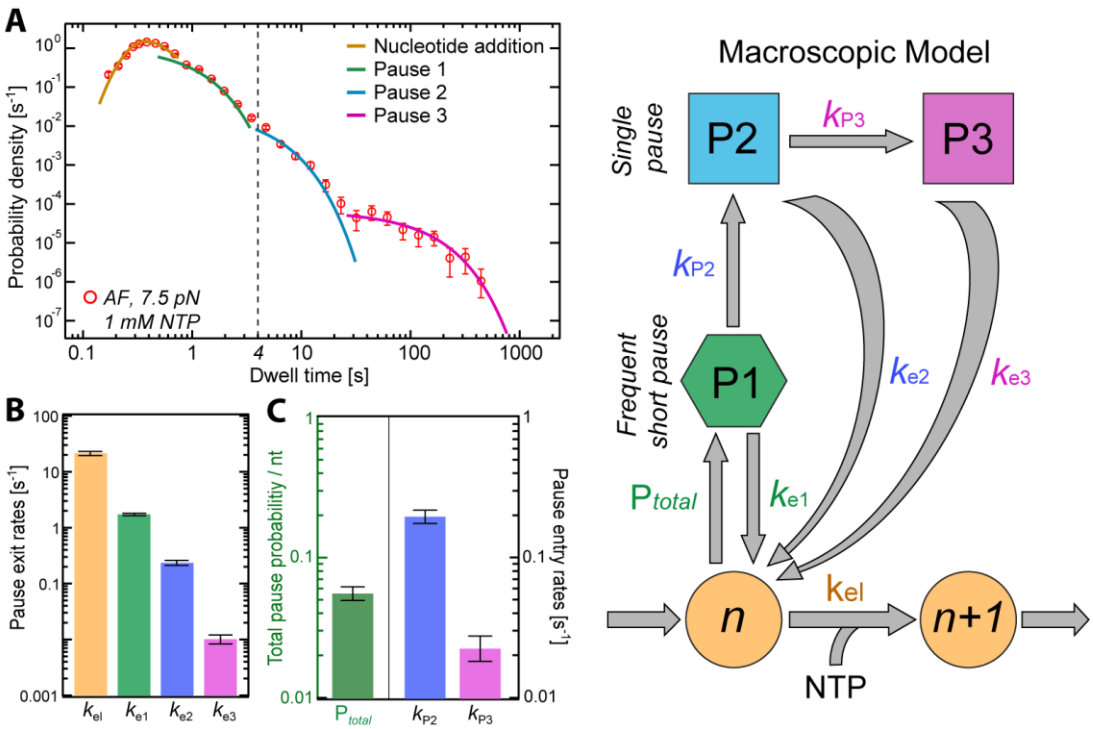

Figure 2: Statistical analysis and extraction of kinetic values of RNAp transcription dynamics. (**A**) Probability distribution (red circles) of extracted DWTs for n = 230 individual RNAp elongation trajectories measured under 7.5 pN AF at a saturating [NTP] of 1 mM. Solid curves show the best fit of the model (rightmost panel) to the data, where the contributions of individual components of the model are separated: the orange curve at short timescales captures the effective pause-free elongation, while the three curves covering the range ~1 s to 1,000 s correspond to three identified exponential pauses (P1: green, P2: cyan, and P3: magenta), determined by the Bayesian information criteria. Dashed black line depicts the pause lifetime threshold of 4 s to calculate the average transcription velocity. (**B**) and (**C**) show the values of the model parameters, resulting from fitting the hypothesized model (rightmost panel) to the distribution shown in (A). The coloring corresponds to the colors used for the different kinetic states in the model (rightmost panel). One-sigma confidence intervals are determined *via* bootstrapping with 1,000 iterations.[55,57] **See also Figures S1, S2, and S3.**

We confirmed that the fit remains robust against the use of different DWT-window sizes (e.g., 4 and 20 nt instead of 10 nt; **Figure S2B, C**); we thus chose a window of 4 nt for all subsequent DWT analyses. Furthermore, we found that the model fits well to the pause-free velocity distribution (**Figure S2D, E**; **Methods M7**), which has been employed in previous studies to estimate the effective elongation rate k_{el}.[23,39,47,59] The estimated k_{el} (**Figure S2F**) from the pause-free velocity distribution (k_{el}: 21.6 ±0.2 nt/s) is compatible with both the value we obtain from DWT-distribution analysis (k_{el}: 21.7 ±1.8 nt/s), and with previously reported pause-free velocity estimations.[39,47] Similarly, the lifetime of ~1 s for the frequent pause state P1 is compatible with the duration of the previously described elemental pause, and pause lifetimes compatible with our estimates of ~4 s (P2) and ~100 s (P3) have also been previously reported.[21,22,24,30,32–36] Our large dataset (**Table S2**) and accompanying analysis procedure thus allow us to unify these three independently observed pause states in a single model.

Estimating next the values of k_{el} and P_{total} for individual trajectories in the reference data set (**Figure S3A**; **Methods M7**), we found that both quantities possess wide distributions (**Figure S3A, B, C**). While the distribution for k_{el} is largely symmetric, P_{total} shows a skewed, bimodal distribution with peaks at 0.05 and 0.08. This heterogeneity in transcription dynamics is also clearly evident in the velocity of individually measured trajectories (**Figure S3D**), as previously reported.[19] To measure the transcription velocity for each trajectory, we removed all segments corresponding to DWTs longer than the P2 lifetime (4 s; dashed line in **Figure 2A**) to exlude the fraction of pauses associated with P2 and P3 state. We then calculated the displacement of RNAp in consecutive temporal windows of 1 s. The velocity is therefore

dominated by the pause-free elongation and frequent pauses associated with the P1 state. To characterize the heterogeneity in velocity, we analyzed two groups of trajectories separated based on their estimated pause probabilities (i.e. $P_{\text{total}} < 0.08$ and $P_{\text{total}} > 0.08$, respectively). While the majority of RNAps progress with an average velocity of ~10 nt/s (**Figure S3D,E**; purple), a small fraction (~30%) of pause-prone RNAp advance more than twice as slow (~5 nt/s; **Figure S3D,E**, green). Interestingly, we detected a small fraction of RNAp (< 5%) in the reference data set in which the velocity alternates between these two mean values (**Figures S3D,F**), a phenomenon that may reflect state-switching.[48,60] Together with the lack of clear correlation between k_{el} and P_{total} for individual trajectories (Pearson correlation coefficient r = 0.16854; **Figure S3A**), these results imply that variations in the frequency of short pauses, and not changes in k_{el}, underlie the observed heterogeneity in velocities. We therefore propose that the state switching reflects a stochastic transition into a more pause-prone mode.

Given the variation in model parameters k_{el} and P_{total} between individual trajectories, we next asked how representative the combined DWT distribution is for the dynamic behavior of individual RNAps. We found that although k_{el} and P_{total} exhibit rather wide distributions, the combined DWT distribution that includes all trajectories (**Figure S3F**) can be described impressively well by their average values. This establishes that, despite the inherent heterogeneity found in model parameters, they exhibit a self-averaging property in the sense that fitting the combined DWT distribution results in a population-averaged estimation of the parameters. Nonetheless, in all subsequent analyses we have estimated k_{el} and P_{total} for individual RNAp trajectories and report the mean values with the corresponding standard errors (**Table S1**). The overall low pausing probability precluded applying the same method to the remaining model parameters due to low statistics. Therefore, a population-averaged estimation for these parameters were obtained by directly fitting the combined DWT distributions (**Methods M8**).

The performed validation assessments ensure the correctness and applicability of our DWT-distribution analysis method with which we ultimately dissect the stochastic dynamics of transcribing RNAp. We next probe the changes in transcription kinetics under different experimental conditions to validate the model structure in a step-wise and bottom-up fashion, and provide insights into the nature and hierarchy of the observed transcriptional pause states.

All transcriptional pauses compete with the active catalytic pathway

In the next step we conducted a systematic experimental approach to discover whether all three identified pause states compete with the active catalytic pathway. To do so, we measured the response to [NTP] ranging from 1 µM to 500 µM at a constant force of 7.5 pN AF, and compared the results with our designated reference data. Overall, in agreement with previous results obtained from bulk and comparable single-molecule experiments, we found that the deficiency of NTPs reduced the average transcription velocity significantly (**Figure 3A**), resulting in a proportional decrease in the processivity of single TECs (**Figure 3B**) due to the statistically unchanged TEC lifetime (**Figure S4A**).[19,39,59]

The combined DWT distributions (**Figure 3C**) present a more detailed view of the effect of NTP availability on the transcription dynamics. As the [NTP] decreases, the elongation peak shifts to larger DWTs, indicating a slower k_{el}. At the same time, the probability of all pauses gradually increases, suggesting that all three pauses are off-pathway states in competition with active elongation. To further validate this hypothesis, we fitted our model with three off-pathway pause states (**Figure 3**, rightmost panel) to the empirical DWT distributions measured at decreasing [NTP] (1 µM - 500 µM). For the fits (**Methods M8**) to the different DWT distributions, we allow only the effective elongation rate k_{el} and the pause probability P_{total} to vary while fixing all other parameters to the previously determined reference values (**Table S1**). The resulting two-parameter fits (**Figure 3C**, solid lines) are able to fully capture the observed

NTP-dependent variations in the combined DWT distributions, demonstrating that all three pause states compete with the active catalytic pathway.

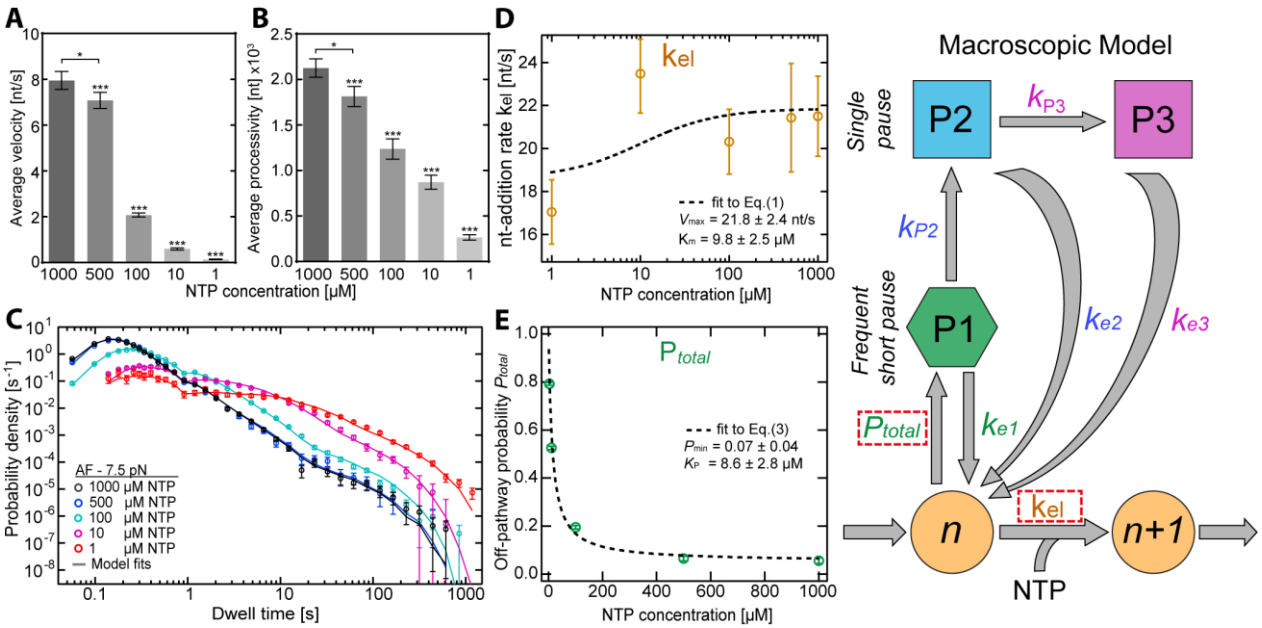

Figure 3: NTP deficiency increases pausing probability and reduces effective elongation rate. (**A**) Measured average (± SEM) end-to-end transcription velocity and (**B**) processivity (± SEM) at different [NTP]. (**C**) Superimposed DWT distributions (circles) of combined elongation trajectories at different [NTP]. The lines represent the best ML fit of the macroscopic model (rightmost panel) to each data set by only setting the nucleotide addition rate k_{el} and the total off-pathway pause probability P_{total} as free fit parameters (rightmost panel, dashed rectangles), while constraining all other kinetic parameters to their reference values (**Figure 2B,C**). (**D**) Extracted effective elongation rates k_{el} from (C), fitted with **Eq. 1** (dashed line). (**E**) Off-pathway entry probabilities P_{total} resulting from the ML fits to data sets in (C), fitted with **Eq. 3** (dashed line). Data shown in (A, B) were subject to statistical analysis using one-way ANOVA, with comparative Tukey post-hoc test (significance level α: *** = *0.001* and $*$ = *0.05*). Error bars for (C, D, E) were determined *via* bootstrapping. **See also Figure S4.**

We next investigated the NTP dependency of both the effective elongation rate and the pause probability. We here note that the mere existence of a competing pause branch affects the lifetime associated with the catalytic pathway - a fact largely ignored in previous studies. Taking this into account (**Methods M9**), the effective elongation rate k_{el} is given by

$$k_{\text{el}} = \left(1 + \frac{K_{\text{p}}}{[\text{NTP}]}\right) k_{\text{el}}^{0}, \qquad \text{(Eq.1)}$$

where the term in the parentheses represents the effect of the pause branch, and k_{el}^{0} follows the expected Michaelis-Menten kinetics [32]

$$k_{\text{el}}^{0} = \frac{V_{\text{max}}}{1+\frac{K_{\text{m}}}{[\text{NTP}]}}\,. \qquad \text{(Eq.2)}$$

Equations **Eq.1** and **Eq.2** imply that k_{el} saturates at a finite value at very low [NTP]. While this may at first seem counterintuitive, it can be understood considering that $\frac{1}{k_{\text{el}}}$ represents the lifetime of the on-pathway elongation, and that even at vanishing [NTP] this lifetime will remain finite due to the exit into off-pathway pausing. The pause probability, however, approaches unity at the lowest [NTP], in accordance with:

$$P_{\text{total}} = 1 - \frac{(1-P_{\text{min}})}{1+\frac{K_{\text{p}}}{[\text{NTP}]}}. \quad \text{(Eq.3)}$$

In the above equations, the quantities V_{max}, K_{m}, P_{min}, and K_{p} depend only on the rates that constitute the active pathway and on the entry rate into the pause branch (**Figure S4B**; **Methods M9**). Their precise values are expected to vary for different nucleotides.[19] We found that the values of k_{el} and P_{total} estimated by the fits are well captured by the **Eqs. 1** and **3** (**Figures 3D**, **E**), leading to averaged (over 4 nt, based on the applied DWT-window) estimations of $V_{\text{max}} = 21.8 \pm 2.4$ nt/s, $K_{\text{m}} = 9.8 \pm 2.5\ \mu$M, $K_{\text{p}} = 8.6 \pm 2.8\ \mu$M, and $P_{\text{min}} = 0.07 \pm 0.04$. Here, the estimated value of K_{m} should be considered as a lower bound for the averaged Michaelis-Menten constant. With respect to the upper bound, we have confirmed that the model remains consistent with the empirical DWT distributions even if K_{m} is increased to 100 µM. The only notable deviations occur in the left tail of the DWT distribution (**Figure S4C**; i.e. at DWTs shorter than 0.2 s), which we found is most sensitive to, and thus can be biased by, the experimental noise.

Our estimations for k_{el} and P_{total} follow the theoretical predictions (**Figure 3D**, **E**), supporting our proposed model with three off-pathway pauses branching from the same active catalytic state. Our results are thus in agreement with the current view in which the elemental pause represents an off-pathway intermediate state that can isomerize into stabilized, long-lived pauses.[21,29,61]

Solely the entry to backtrack-associated pauses exhibits force dependency

With the knowledge that the three identified pause states are off-pathway pauses that branch from the same active catalytic state, we next investigated their origin and interconnection. Since previous single-molecule and bulk studies associated longer pause lifetimes with diffusive return from backtracking, which is expected to be susceptible to force, we probed the force dependency of the RNAp dynamics to identify the pause states which are coupled to RNAp translocation.[15,23,36,38–41]

We systematically measured the transcription dynamics of RNAp at different applied forces ranging from 5 to 12.5 pN, in both the AF and OF directions (**Figure 1A**). Globally, with the gradual increase in the magnitude of the OF, we observe a significant decrease in average velocity (**Figure S5A**), indicating a notable increase of either the pausing frequency or duration, or both, since the average TEC lifetimes remain comparable (**Figure S4**). **Figure 4A** depicts examples of combined DWT distributions measured at 12.5 pN in AF and OF orientations (resulting DWT distributions for all applied forces are shown in **Figure S5**). We found that switching the force direction from AF to OF produces a significant effect similar to that observed at low [NTP]: enhanced pausing probability accompanied with a potential decrease of the effective elongation rate.

To explain the observed force dependency, we first fitted our model (**Figure 4**, rightmost panel) to the 12.5 pN OF data by only allowing k_{el} and P_{total} to change while fixing all other parameters to the reference values (**Figure 4A**, purple dashed line). While the model fits the data at short timescales <10 s, it slightly deviates at longer timescales. The remaining discrepancy between the model and the data vanishes when we introduce force dependence in the entry rate of the P2 state:

$$k_{\text{P2}} = k_{\text{P2}}^{\text{ref}} \exp\left(-\frac{(F-F^{ref})(1-\delta)a}{k_{\text{B}}T}\right), \quad \text{(Eq.4)}$$

where $a \approx 0.37$ nm is the average step size of RNAp, $k_{\text{P2}}^{\text{ref}}$ is the value corresponding to the reference force $F^{\text{ref}} = 7.5$ pN, δ identifies the position (in nucleotides) of the energy barrier against RNAp backward

translocation with respect to the backtracked state, and F is the applied external force.[41,62] Sweeping the range for δ between 0 and 1 with increments of 0.1 (**Methods M8**) in our fits, we found that the entry to P2 exhibits force-dependency with an optimal value of δ =0.7±0.1 (**Figure 4B,** blue solid line). As shown in **Figure 4A**, increasing the force dependency of k_{P2} to its maximal extent by setting δ = 0 results in an overestimation of the probability of the long-lived pauses P2 and P3 (**Figure 4A**, black dashed line).

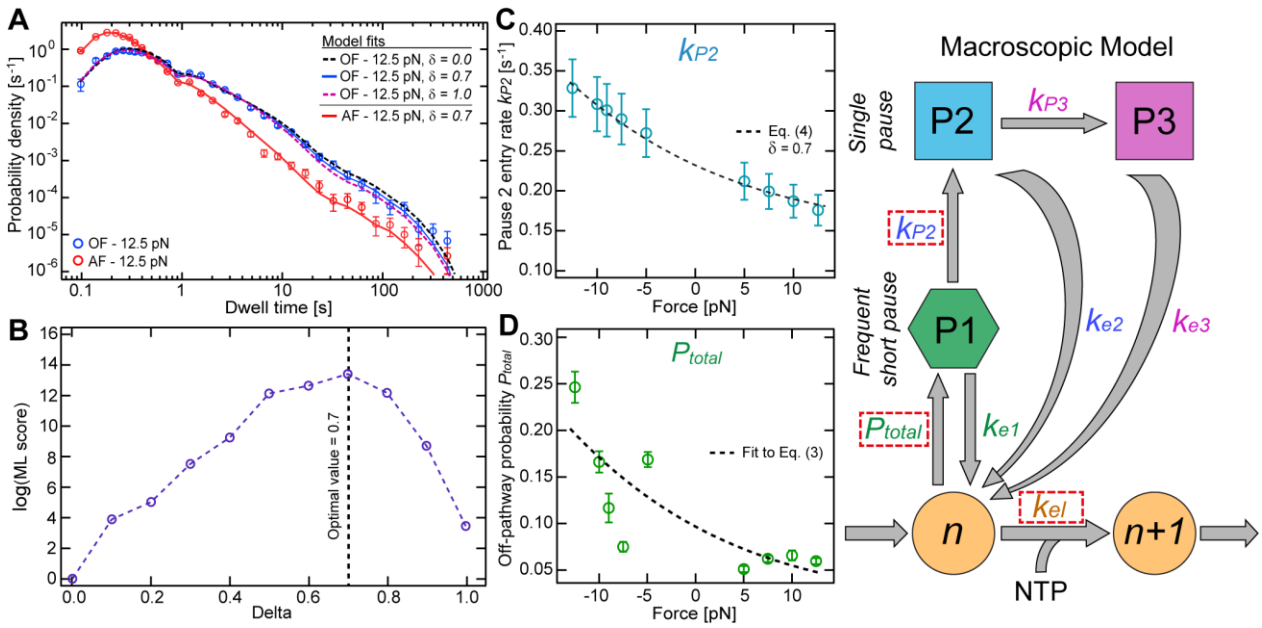

Figure 4: RNAp backtracking exhibits force-dependency in the entry rate, but not diffusion-governed recovery. (**A**) Combined DWT distributions (circles) measured at 12.5 pN AF and OF. The (dashed) lines represent the best ML fits of the macroscopic model (rightmost panel) to each data set. Three fit parameters, the nucleotide addition rate k_{el}, total off-pathway pause probability P_{total} , and P2 entry rate k_{P2} assumed to be force dependent via the parameter δ (δ: 0 = maximal force dependency, 1 = no force dependency; see **Eq. 4**), were used. All other kinetic parameters were constrained to their reference values (**Figure 2B,C**). (**B**) ML fit quality as a function of the force-dependency of P2 entry δ, from fitting the OF 12.5 pN data in (A). The maximum at 0.7 marks the optimal value for δ. (**C**) k_{P2} as a function of force, calculated from **Eq. 4** with δ= 0.7±0.1. The dashed line shows the interpolation between the measured forces. (**D**) P_{total} probabilities from the ML fits to data sets at different applied forces (**Figure S5B,C**). The dashed black line depicts the theoretical prediction (**Eq. 3**) with δ= 0.7. **See also Figures S4, S5,** and **S6.**

To validate proposed force-dependency for the P2 entry rate k_{P2}, we performed two-parameter fits to the data obtained at different forces by using k_{el} and P_{total} as fit parameters, fixing k_{P2} at its expected value for each force assuming δ= 0.7 (**Eq. 3**; **Figure 4C**) and fixing all other model parameters to their reference values (**Table S1**). The resulting fits to the data at 12.5 pN AF (**Figure 4A,** red solid line), as well as to all other measured forces (**Figure S5B,C**), demonstrate that our model remains robust at different forces.

We next explored the force dependency of the catalytic pathway. Each NMP addition cycle involves transition of RNAp from pre- to the post-translocated state, which is expected to be affected by an externally applied force.[19,23,36,39,41,59,63] We found that the effective elongation rate k_{el}, as estimated by our fits, remains rather constant with force (**Figure S5D**), implying that at saturating 1 mM NTPs, RNAp translocation does not become rate limiting even when the OF is as large as 12.5 pN. In contrast, the off-pathway pause probability P_{total} shows an increasing trend when the exerting force switches from AF to OF orientation (**Figure 4D**). Using **Eq. S36** to describe the force dependency of P_{min} and K_p in **Eq. 3** (**Methods M9**), with the reference values P_{min}^{ref} = 0.07 and K_p^{ref}= 8.6 μM corresponding to the force F^{ref} = 7.5 pN (**Figure 3E**), we fitted **Eq. 3** (**Figure 4D**, dashed line) to the estimated values of pause probability (**Figure 4D**, circles).

The best fit corresponds to $\Delta \approx 0.5 \pm 0.16$, which identifies the position of the energy barrier against RNAp transition from the pre- to post-translocated state (**Eq. S35**). The evident discrepancy between the model prediction and the estimations might be due to an inhomogeneity in the magnitude of force applied to the magnetic beads since the amount of embedded iron oxide particles is not stringently constant.[64] This could explain the non-monotonic variation of the pause probability around 5 pN OF. Alternatively, the sharp increase in the off-pathway pause probability P_{total} at large OFs might indicate an additional force dependency, e.g., in the entry rate to the elemental pause; this is compatible with the recent observation that the elemental pause may be in fact a half-translocated state.[21]

The fact that the entry rate k_{P2} into the first longer-lived pause state P2 increases with OF (**Figure 4C**) implies that this entry is accompanied by the RNAp backward translocation. We therefore propose that P2, and in consequence, P3 are backtracked pauses. It has been shown that RNAp can stochastically switch to a catalytically-inactive backtracking mode, where it can diffuse back and forth along the template DNA.[25,36,44,65] In a backtracked TEC, the nascent RNA is threaded through the active site, blocking nucleotide addition. The RNAp can recover from backtracking either by diffusing forward, placing the 3'end of the transcript in the active site, or by cleaving the backtracked RNA, thereby freeing the active site.[40,43,45,66–68] Previous single-molecule studies have suggested that diffusion provides the dominant recovery mechanism.[15,23,36,38–41] Our data contradicts this view: we observe only a limited degree of force dependency for k_{P2}, and the lifetimes of P2 and P3 states do not seem to be affected by the external force,. To examine this further, we parametrized the previously suggested diffusive backtracking model (**Figure S6**, rightmost panel) by fitting it to the reference data (**Methods M10**).[41] We found that the model captures the combined DWT distribution of the reference data sufficiently well, provided that the maximum backtrack depth (BT) exceeds 5 nt (**Figure S6A)**. This is compatible with previous studies that assumed unrestricted diffusive backtracking (reflected by the $t^{-\frac{3}{2}}$ power-law relationship) to describe long-lived pauses.[15,23,36,38,39,41] However, the agreement between the data and the diffusive backtracking model fails to persist at high OFs. This is illustrated in **Figure S6B**, where the DWT distribution was simulated at 12.5 pN OF (**Methods M10**) with different values of δ and the maximum BT (**Figure S6**, rightmost panel). None of the simulated DWT distributions, including the one with the highest likelihood score (**Figure S6B,** red line), satisfactorily capture the data. Moreover, the simulated trajectories corresponding to the best fit typically show a significantly shorter processivity than the measured data, an inherent property of the diffusive backtracking model, where at sufficiently high OFs essentially all RNAps will be trapped in an irreversible backtrack (**Figure S6C**).[41] We found that this occurs in the diffusive backtracking model provided that the maximum BT exceeds 5 nt, leading to simulated trajectories which show, on average, at least one order of magnitude lower processivity compared to our measured RNAp trajectories (**Figure S6D**).

We conclude that while the diffusive backtracking model can be fitted to any individual DWT distribution at a constant force, it does not provide a consistent description of our data over the entire force range and significantly underestimates the RNAp processivity. This indicates that the diffusive RNAp dynamics alone cannot account for the observed long-lived pauses in our empirical DWT distributions. Instead, the intrinsic cleavage capability of RNAp may be responsible for the recovery from backtracking, which we experimentally verify next.

Recovery from backtracked states is primarily mediated by intrinsic cleavage

To test our hypothesis that the long-lived pauses arise from backtracked TECs, with RNA cleavage being the predominant backtrack-recovery process, we experimentally probed the contribution of transcription elongation factors GreA and GreB to the backtracking dynamics. The prokaryotic Gre factors augment the intrinsic cleavage rate of RNAp, promoting cleavage-mediated recovery of backtracked TEC.[25,34,40,42,43,45,66]

To increase the measurable effect of Gre factors, we applied an OF that favors the long-lived pauses (**Figure 4**; **Figure S5**). **Figure 5A** shows the effect of 2 µM GreA and 2 µM GreB have on the DWT distribution measured at 9 pN OF, 1 mM [NTP]. It is evident from the DWT distributions that solely pauses exceeding 4 s, corresponding to the P2 and P3 pause states (**Figure 2B**; **Table S1**), are affected by both Gre factors, and that the effect is considerably more pronounced in case of GreB. To evaluate which of the two pause states is affected, we fitted our model (**Figure 5**, rightmost panel) to the data (**Methods M8**) by only allowing the exit rate k_{e2} from P2, the entry rate k_{P3} to P3, and the exit rate k_{e3} from P3 to vary; all other model parameters were fixed according to the best fit to the corresponding control data acquired in the absence of Gre factors. We found that the three-parameter fits were able to fully capture the effects of both Gre factors (**Figure 5A**); the resulting fit values for k_{e2}, k_{P3}, and k_{e3} are shown in **Figure 5B, 5C**, and **5D**, respectively. Compared to the control data, the effect of GreA was statistically insignificant, while GreB significantly increased k_{e2} while keeping k_{P3} and k_{e3} unchanged within errors.

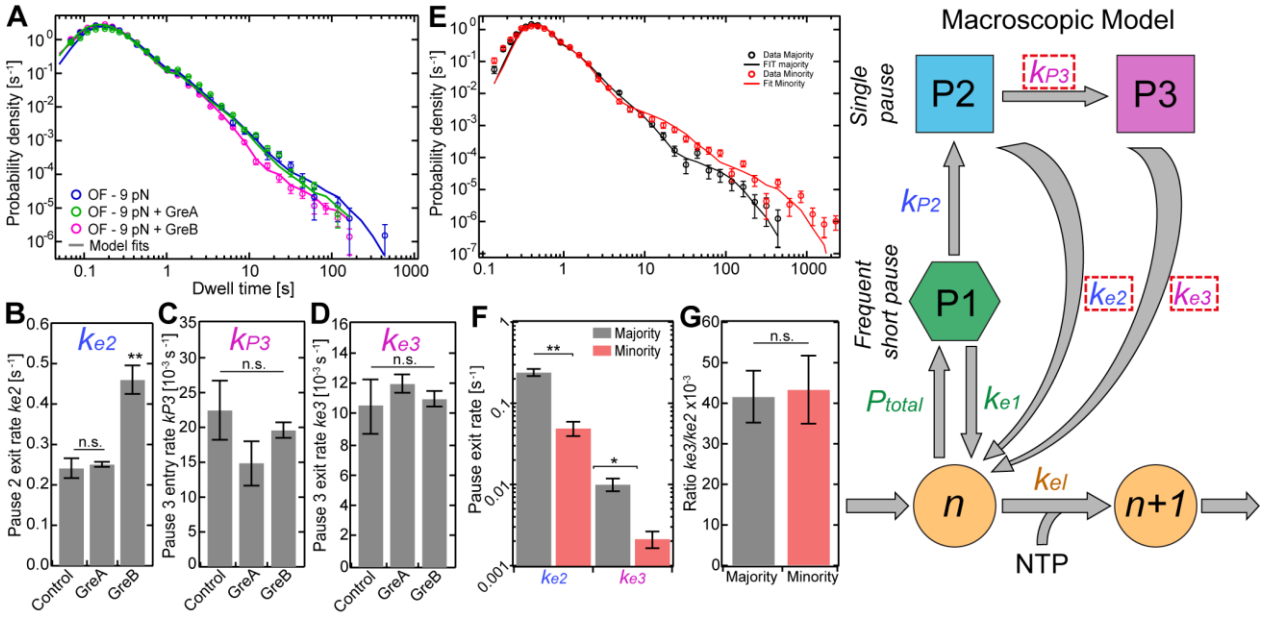

Figure 5: GreA/B-mediated cleavage and heterogeneity in intrinsic cleavage. (**A**) Superimposed combined DWT distributions (circles) of transcription trajectories measured at OF 9 pN in presence of GreA (green) or GreB (purple), and absence of either (blue). The lines in (A) represent the ML fits of our model (rightmost panel) to each data set. The combined DWT distributions in presence of Gre factors were fitted by using the P2 exit rate (k_{e2}), the P3 entry rate (k_{P3}), and the P3 exit rate (k_{e3}) as the only free fit parameters (dashed rectangles, rightmost panel), while constraining all other kinetic parameters according to the best fit to the control data (i.e. in absence of Gre factors) measured at 9 pN OF, 1 mM [NTP] (**Table S1**). (**B**) P2 exit rate k_{e2}, (**C**) P3 entry rate k_{P3}, and (**D**) P3 exit rate k_{e3}, estimated from the ML fits shown in (A). (**E**) Superimposed DWT distributions (circles) and ML fits (lines) for the minority and dominant sub-populations, extracted from the reference data and separated *via* Gaussian mixture clustering (**Methods M4**) with respect to their maximum pause durations. (**F**) P2 exit rates k_{e2} and P3 exit rates k_{e3} for the majority (grey) and minority (red) sub-populations, estimated from the ML fits shown in (E). (**G**) The ratio between the exit rates of P2 and P3 for the dominant (grey) and minority (red) sub-populations. Statistical analyses consisted of unpaired, two-tailed t-tests (significance level p: ** ≤ 0.01, * = ≤ 0.05; n.s. = non-significant). **See also Figures S5, S7,** and **S8.**

Since GreB is known to facilitate recovery from deep backtracks, the observed acceleration of the exit rate k_{e2} is a clear indication that the P2 pause originates from backtracked TECs.[25,34,40,42–44,66] The concomitant observed reduction in the occurrence of the P3 pause state, reflected by a decrease in the appearance of pauses exceeding 10 s in **Figure 5A,** supports our assumption that the entry to P3 competes with recovery from P2, implying that P3 may also originate from a backtrack-induced pause. Backtrack depth analysis reveals that P2 and P3 are associated with backtrack depths >4 nt (**Figure S7**). Surprisingly, and in contrast to P2, the lifetime of P3 (**Figure 5D**) is largely unaffected by GreB while showing similar backtrack depths (**Figure S7B**). This suggests that P3-paused RNAp may undergo major conformational

changes that render it resistant to GreB-assisted cleavage, which in turn leads to a 20-fold slower recovery rate as compared to the recovery from P2.

Our observation that GreA does not affect the apparent RNAp pause dynamics can be understood by the fact that GreA mainly affects short backtracks of ≤3 nucleotides.[34,40,42,45,69,70] This result indicates that short backtracks are short-lived and either recover quickly (≤1 s) *via* cleavage or diffusion, or lead to extended backtracking. This scenario could explain why short backtracks remain undetected in our analysis, as their contribution to the DWT distribution might be overshadowed by the elemental pause (lifetime ~ 1 s), as proposed by Saba *et al.*[21] Furthermore, if short backtracks are predominantly recovered by diffusion, which we can neither confirm nor exclude from our data, the GreA-mediated cleavage might not be able to substantially suppress the occurrence of deeper backtracks.

A minority population reveals the nature of long-lived backtrack-associated pauses

As described above, variations in the duration of long pauses between individual RNAp trajectories (**Figure 1D**) forced us to separate the data sets (**Methods M4**) into a dominant sub-population (~90%), to which we fitted our working model, and a minority sub-population (~10%) that contained trajectories with unusually long pauses exceeding thousands of seconds; this population has thus far been excluded from our analyses. Here, we address the question of whether the transcription dynamics of RNAp in the minority sub-population can be understood in the context of our model.

We constructed a combined DWT distribution using only the minority sub-population from the reference data set (**Figure 5E**, red circles), and performed a maximum likelihood fit (**Figure 5E**, red line) to obtain a new set of model parameters. We found that only the exit rates from P2 and P3 vary significantly between the two sub-populations (**Figure 5F**; **Figure S8A**), while all other parameters remain statistically unchanged (**Figure S8A**). Strikingly, the ratio between these exit rates is comparable in both sub-populations (**Figure 5G**), suggesting that rate-limiting sub-processes that contribute to the P2 recovery, also contribute to the P3 recovery. Alteration of the rates associated with these common sub-processes would change the lifetimes of both P2 and P3 proportionally, indicating that the recovery from P3 also involves RNA cleavage. While P2 and P3 do not appear to differ in backtracking depth, recovery from P3 occurs ~20 times slower and cannot be accelerated by GreB. These observations imply that P3 results from an infrequently occurring, long-lived conformational change in backtracked TEC that considerably suppresses RNA cleavage. As a result, recovery from P3 requires reversion to the cleavage-competent conformation.

To verify that the observed difference between the two sub-populations is not simply a result of insufficient statistics originating from short trajectories, we simulated trajectories of the same length distribution we have in the reference data set with the model parameters obtained from both sub-populations (referred to as dominant and minority parameter sets). We then compared the distributions of the longest pause per trajectory constructed from the simulated data with the empirical distribution (**Figure S8B**). We found that for the dominant parameter set the simulated distribution closely resembles the empirical data, while the minority parameter set results in a significant overestimation in the frequency of long pauses. This confirms that the observed heterogeneity in pause duration cannot be solely explained by the limited statistics, and justifies our separate treatment of the minority sub-population as a fundamentally distinct population of TECs.

DISCUSSION

A unifying and extended kinetic model of intrinsic transcription dynamics

Our empirical data, complemented by *in silico* modeling, provide a complete view of the RNAp pausing dynamics that unifies and extends previously proposed models. Based on our findings, we propose the full

kinetic model depicted in **Figure 6**, in which frequent short-lived pauses, including the elemental pause (EP) and short backtracks (BT) of lifetimes <1 s, lead to two long-lived pauses (P2 and P3) accompanied by backtracks deeper than 4 nt.

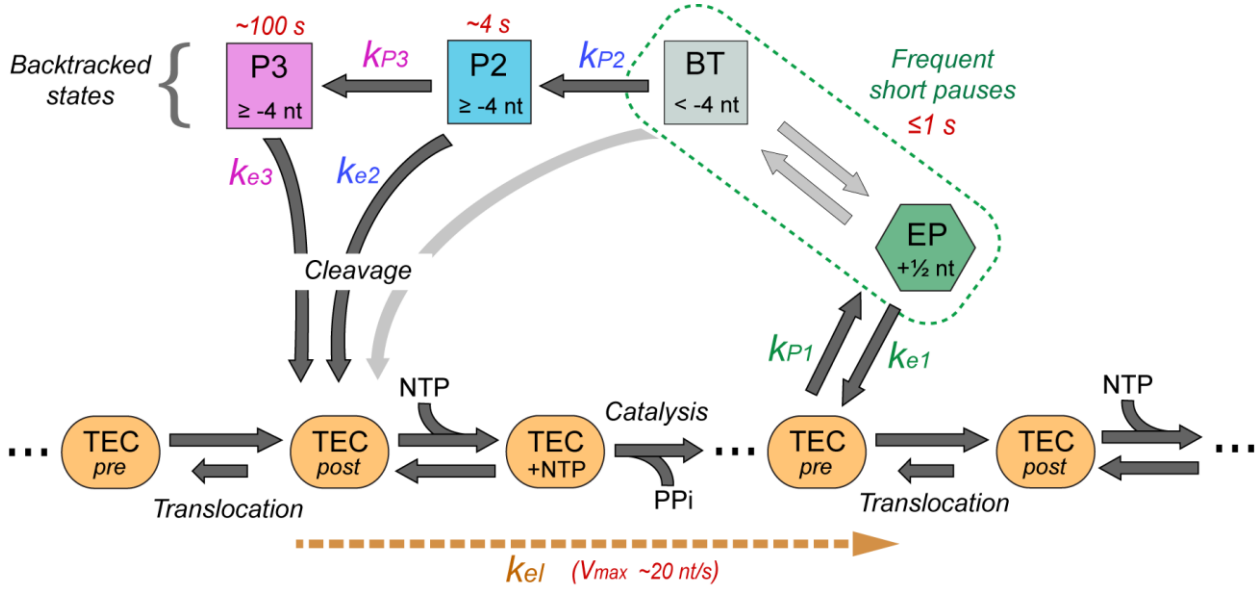

Figure 6: Unified mechanistic model of RNAp transcription. Kinetic model depicting the hierarchy of kinetic states involved in RNA elongation. The active pathway includes the pre-translocated ("*pre*"), post-translocated ("*post*"), NTP-bound ("*+NTP*"), and catalysis (with PP_i release) states of the TEC. The sub-processes involved in transcript elongation, namely translocation, NTP-addition, and catalysis, collectively contribute to the effective elongation rate k_{el}. The pause pathway includes the - half translocated - elemental pause ("EP"), the -undetected- fast pause (BT) associated with backtracks shorter than 4 nt and two long-lived pauses (P2 and P3) associated with deeper backtracks with cleavage-mediated recovery mechanism.[21] For each pause state, the assumed position of the RNAp's active site with respect to the pre-translocated state is identified. The estimated lifetimes are also denoted in red. The estimated maximum transcription velocity is denoted in red as $V_{\max}$ (Eq. 1).

In agreement with published models, our data show that pauses are serially.[18,21,30,37,40,61,71,72] The shortest pause state, with a lifetime <1 s, resembles the elemental pause identified in previous biochemical and single-molecule studies.[21,22,29–33] We have shown that the two long-lived pauses are promoted by OF and are associated with backtracking, based on our direct observation of backtracking events in transcription trajectories. Given that diffusive backtracking over several nucleotides is incompatible with our data, and that the recovery from P2 state is accelerated by GreB, we conclude that this state corresponds to backtracks deeper than four nucleotides, where GreB is known to be effective.[25,34,40,43,44,66,69] Our data and simulations provide strong indications that intrinsic RNA cleavage is the dominant recovery pathway for this state. The fact that shorter backtracks remain undetected in our analysis implies that the recovery from short backtracks, whether through diffusion or intrinsic cleavage, is faster than the elemental pause.

Our observation that the backtrack pauses show only a limited dependency on the external applied force has been reported previously.[19,36,38,39,41,59,63,73] However, such a lack of force sensitivity has been commonly attributed to a force-independent power-law tail of the DWT distribution that originates from diffusive backtracking.[15,23,36,38–41] Our longer observation times allowed us to better discriminate power-law behavior from exponentially distributed states. We find that the diffusive backtracking model does not provide a consistent description for our data across the entire force range. Instead, our findings support biochemical and structural studies that suggest the existence of two distinct backtracking modes, distinguished by different conformational states of the TEC with unequal effective cleavage rates.[21,22,34,35,66]

The nature of long-lived backtrack-stabilized kinetic states

Our analysis revealed that the long-lived pauses originate from two distinct backtracked states, which differ primarily in the susceptibility to assisted cleavage by GreB and the apparent intrinsic RNA cleavage rates. These two different pause states, with lifetimes comparable with our estimations, were previously inferred from biochemical and structural studies, where it has been proposed that both the GreB-competent (pause P2) and the GreB-incompetent (pause P3) backtracked states involve structural rearrangements that render intrinsic RNA cleavage temporarily inactive.[21,22,35,47,66] All catalytic reactions of RNAp are associated with conformational transitions of the key mobile elements in the largest RNAp subunit, the bridge helix (BH) and the trigger loop (TL). The inability of the BH and TL to fold correctly to the cleavage-competent conformations is thought to sterically hinder the backtracked RNA transcript, delaying cleavage and stabilizing the backtracked TEC in a ratcheted confomation.[30,34,61,66,74] The resistance to GreB-assisted cleavage and the significantly decreased intrinsic RNA cleavage rate in the P3 state as compared to P2 suggest that major structural changes accompany the P3 state formation. Indeed, several previous studies reported comparable cleavage rates for backtracked TECs, attributed to conformational rearrangements of several structural components in addition to TH and BH. For example, RNAp swiveling involves structural rearrangements of the clamp, shelf, SI3, and jaw domains, leading to a formation of an inert TEC.[21,30,34,75] The observed extended recovery time for the GreB-incompetent pause state (P3) thus reflects the lifetime of the altered configurations, i.e. the time necessary for RNAp to revert to a RNA cleavage-competent state. Importantly, both types of backtrack-stabilized TEC states are eventually able to resume transcription.

What triggers the formation of the inert P3 state? While some studies suggested the backtrack depth as a probable cause, the majority of studies found no dependency on the backtrack depth.[34,36,62,76] Our results confirm this observation since we found no significant difference in the backtrack depth between the GreB-competent (P2) and the GreB-incompetent (P3) states.We therefore suggest that, independent of the backtrack depth, TECs residing in the GreB-competent backtrack state (P2) can stochastically either revert to the catalytic pathway through RNA cleavage, or undergo additional conformational changes of several mobile components that lead to a GreB-incompetent state (P3). Based on our estimations, the majority of TECs that remain in the backtracked state longer than $k_{\mathrm{p3}}^{-1}\sim 1$ minute will be found in the GreB-incompetent state (P3). This hypothesis is supported by several biochemical studies which have investigated backtracked TECs. The majority of these studies 'walked' TECs into backtracked states of various depths and stalled the RNAp for ≥10 min before measuring.[28,34,43,44,66,69,77] In these studies a long cleavage time between 100 s to 200 s, comparable to our estimation of the P3 lifetime, is reported. Future structural studies, focusing on TEC stalling and pause durations, can provide further insights into the structural nature of the different backtrack states.

The origin of heterogeneity in transcription velocity and pausing dynamics

Since long transcriptional pauses are rare, accurate characterization of their lifetimes and frequencies requires a large ensemble of sufficiently long elongation trajectories. To achieve such statistics, the majority of previous studies aggregated all measured RNAp transcription trajectories based on the assumption that all RNAp behave similarly.[31,47,48,78] Nonetheless, heterogeneities between individual RNAp velocities have been reported, and two studies presented empirical and theoretical evidence in support of dynamic state switching, which involves alternation between different transcription velocities of a single RNAp. [26,29,32,48–50,60,79] However, most studies failed to detect state-switching, concluding that the individual RNAps transcribe at a constant velocity.[24,32] Apparent differences in transcription velocities were attributed either to variations in short pause frequencies or to possible chemical variations between individual RNAps, such as post-translational modifications.[24,47,80]

With our large ensemble of RNAp trajectories that captures the full extent of temporal transcription dynamics, we were able to estimate the elongation rates and the pause frequencies of individual RNAp, revealing considerable heterogeneity in both quantities. While the majority of RNAp transcribed at similar average velocity (~10 nt/s), a small but notable fraction of RNAp (~10%) transcribed the template substantially more slowly (~5 nt/s). Most interestingly, a small faction (~5%) of RNAp trajectories showed switching between these two velocities. These observations are not only quantitatively compatible with the previously described state switching, but also demonstrate that the heterogeneity of the transcription velocity is intimately linked to state switching.[48] Since we excluded long backtrack pauses from the estimation of transcription velocities, potential contributing factors to the observed heterogeneity are the pause-free elongation rate or short frequent pauses, which include the elemental pause and shallow backtracks. We provide evidence that the variations in the frequency of short pauses, and not in the pause-free elongation rate, underlie the heterogeneity in transcription velocity and the state-switching phenomena.

The mechanistic origin of state switching remains unknown. We speculate that the observed variation reflects changes in several mobile RNAP modules whose dynamic structural arrangements have shown to be important for substrate binding, catalysis, and translocation. Point mutations in these structural elements are known to significantly affect the transcription rate.[29,39,75,78,81,82] We propose that the observed slow transcription velocity of ~5 nt/s might be a result of a potentially unfavorable or hampered reorganization of these elements, leading to a dynamic switching between fast and slow transcription velocities or, in extreme cases, to a permanently slow transcription. Since the described transcription heterogeneity consistently persists throughout our experiments, we hypothesize that chemical modifications of conserved amino acid residues may underlie this phenomenon.

We have further identified a separate source of heterogeneity that is reflected by significant differences in long-lived pause lifetimes associated with backtracked TEC. A minority sub-population of RNAp (~10%) showed notably slower backtrack-recovery rates from both the GreB-competent (pause P2) and the GreB-incompetent (pause P3) backtracked states, indicating a slower intrinsic cleavage rate. We note that the longest pause durations we encountered for this sub-population exceed thousands of seconds, and most probably have been previously categorized as arrested states unable to recover from backtrack.[34,36,38,43,48,59] Since several aforementioned structural elements involved in RNA synthesis also participate in the RNA cleavage, we assume that the suggested possible chemical modification of different structure motifs during RNAp synthesis may also be responsible for the observed RNA cleavage-deficient sub-population. Elucidating the microscopic origin and cause of the observed different dynamic RNAp behavior remains an important subject for future biochemical and structural studies.

SUPPLEMENTAL INFORMATION

Supplemental Information includes 8 figures, 3 tables, and a detailed methodological description, and can be found with this manuscript online.

ACKNOWLEDGEMENTS

Funding to N.H.D. was provided by the Netherlands Organization for Scientific Research (NWO) via its TOP-GO program and by the European Union via an ERC Consolidator Grant (DynGenome, no 312221). Funding to I.A. was provided by the National Institutes of Health grant GM67153. B.E.M. forms part of the research program ‘‘Crowd management: the physics of genome processing in complex environments”, supported by the Netherlands Organisation for Scientific Research (NWO). We are grateful to Theo van Laar for DNA construct synthesis, Nina Turk for assistance in single-molecule measurements, Mariana Köber for backtrack depth analysis, and Georgiy Belogurov for discussions.

AUTHOR CONTRIBUTIONS

R.J., B.E.M., I.A., M.D., and N.H.D. conceived and designed the experiments. R.J. performed single-molecule transcription experiments and data processing. B.E.M and M.D. conceived of the analysis pipeline, and B.E.M. performed the dwell-time analysis and *in silico* modeling. All authors contributed to writing the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Lee, J. & Borukhov, S. Bacterial RNA Polymerase-DNA Interaction -The Driving Force of Gene Expression and the Target for Drug Action. *Front. Mol. Biosci.* **3**, (2016).
2. Atkinson, T. J. & Halfon, M. S. REGULATION OF GENE EXPRESSION IN THE GENOMIC CONTEXT. *Comput. Struct. Biotechnol. J.* **9**, e201401001 (2014).
3. Rockman, M. V. & Kruglyak, L. Genetics of global gene expression. *Nat. Rev. Genet.* **7**, 862–872 (2006).
4. Gehring, A. M., Walker, J. E. & Santangelo, T. J. Transcription Regulation in Archaea. *J. Bacteriol.* **198**, 1906–1917 (2016).
5. Yakhnin, A. V & Babitzke, P. NusG/Spt5: are there common functions of this ubiquitous transcription elongation factor? *Curr. Opin. Microbiol.* **18**, 68–71 (2014).
6. Svetlov, V., Belogurov, G. A., Shabrova, E., Vassylyev, D. G. & Artsimovitch, I. Allosteric control of the RNA polymerase by the elongation factor RfaH. *Nucleic Acids Res.* **35**, 5694–5705 (2007).
7. Charoensawan, V., Wilson, D. & Teichmann, S. A. Genomic repertoires of DNA-binding transcription factors across the tree of life. *Nucleic Acids Res.* **38**, 7364–7377 (2010).
8. Ponting, C. P. Novel domains and orthologues of eukaryotic transcription elongation factors. *Nucleic Acids Res.* **30**, 3643–52 (2002).
9. Lisica, A. *et al.* Mechanisms of backtrack recovery by RNA polymerases I and II. *Proc. Natl. Acad. Sci.* **113**, 2946–2951 (2016).
10. Nudler, E. RNA Polymerase Backtracking in Gene Regulation and Genome Instability. *Cell* **149**, 1438–1445 (2012).
11. Hein, P. P. *et al.* RNA polymerase pausing and nascent-RNA structure formation are linked through clamp-domain movement. *Nat. Struct. Mol. Biol.* **21**, 794–802 (2014).
12. Chamberlin, M. J., Nierman, W. C., Wiggs, J. & Neff, N. A quantitative assay for bacterial RNA polymerases. *J. Biol. Chem.* **254**, 10061–9 (1979).
13. Kingston, R. E., Nierman, W. C. & Chamberlin, M. J. A direct effect of guanosine tetraphosphate on pausing of Escherichia coli RNA polymerase during RNA chain elongation. *J. Biol. Chem.* **256**, 2787–97 (1981).
14. Michaelis, J. & Treutlein, B. Single-Molecule Studies of RNA Polymerases. *Chem. Rev.* **113**, 8377–8399 (2013).
15. Dangkulwanich, M., Ishibashi, T., Bintu, L. & Bustamante, C. Molecular Mechanisms of Transcription through Single-Molecule Experiments. *Chem. Rev.* **114**, 3203–3223 (2014).
16. Mayer, A., Landry, H. M. & Churchman, L. S. Pause & go: from the discovery of RNA polymerase pausing to its functional implications. *Curr. Opin. Cell Biol.* **46**, 72–80 (2017).
17. Murakami, K. S. & Darst, S. A. Bacterial RNA polymerases: the wholo story. *Curr. Opin. Struct. Biol.* **13**, 31–39 (2003).
18. Weixlbaumer, A., Leon, K., Landick, R. & Darst, S. A. Structural Basis of Transcriptional Pausing in Bacteria. *Cell* **152**, 431–441 (2013).
19. Abbondanzieri, E. A., Greenleaf, W. J., Shaevitz, J. W., Landick, R. & Block, S. M. Direct observation of base-pair stepping by RNA polymerase. *Nature* **438**, 460–465 (2005).
20. Galburt, E. A. *et al.* Backtracking determines the force sensitivity of RNAP II in a factor-dependent manner. *Nature* **446**, 820–823 (2007).
21. Saba, J. *et al.* The elemental mechanism of transcriptional pausing. *Elife* **8**, (2019).
22. Artsimovitch, I. & Landick, R. Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proc. Natl. Acad. Sci.* **97**, 7090–7095 (2000).
23. Gabizon, R., Lee, A., Vahedian-Movahed, H., Ebright, R. H. & Bustamante, C. J. Pause sequences facilitate entry into long-lived paused states by reducing RNA polymerase transcription rates. *Nat. Commun.* **9**, 2930 (2018).
24. Herbert, K. M. *et al.* Sequence-Resolved Detection of Pausing by Single RNA Polymerase Molecules. *Cell* **125**, 1083–1094 (2006).
25. Lass-Napiorkowska, A. & Heyduk, T. Real-Time Observation of Backtracking by Bacterial RNA Polymerase.

Biochemistry **55**, 647–658 (2016).
26. Erie, D. A. The many conformational states of RNA polymerase elongation complexes and their roles in the regulation of transcription. *Biochim. Biophys. Acta - Gene Struct. Expr.* **1577**, 224–239 (2002).
27. Gamba, P., James, K. & Zenkin, N. A link between transcription fidelity and pausing in vivo. *Transcription* **8**, 99–105 (2017).
28. Kang, J. Y. *et al.* RNA Polymerase Accommodates a Pause RNA Hairpin by Global Conformational Rearrangements that Prolong Pausing. *Mol. Cell* **69**, 802–815.e5 (2018).
29. Belogurov, G. A. & Artsimovitch, I. The Mechanisms of Substrate Selection, Catalysis, and Translocation by the Elongating RNA Polymerase. *J. Mol. Biol.* (2019). doi:10.1016/j.jmb.2019.05.042
30. Sekine, S., Murayama, Y., Svetlov, V., Nudler, E. & Yokoyama, S. The Ratcheted and Ratchetable Structural States of RNA Polymerase Underlie Multiple Transcriptional Functions. *Mol. Cell* **57**, 408–421 (2015).
31. Herbert, K. M. *et al.* Sequence-Resolved Detection of Pausing by Single RNA Polymerase Molecules. *Cell* **125**, 1083–1094 (2006).
32. Neuman, K. C., Abbondanzieri, E. A., Landick, R., Gelles, J. & Block, S. M. Ubiquitous transcriptional pausing is independent of RNA polymerase backtracking. *Cell* **115**, 437–47 (2003).
33. Dalal, R. V. *et al.* Pulling on the Nascent RNA during Transcription Does Not Alter Kinetics of Elongation or Ubiquitous Pausing. *Mol. Cell* **23**, 231–239 (2006).
34. Miropolskaya, N., Esyunina, D. & Kulbachinskiy, A. Conserved functions of the trigger loop and Gre factors in RNA cleavage by bacterial RNA polymerases. *J. Biol. Chem.* **292**, 6744–6752 (2017).
35. Pupov, D. V., Barinova, N. A. & Kulbachinskiy, A. V. Analysis of RNA cleavage by RNA polymerases from Escherichia coli and Deinococcus radiodurans. *Biochem.* **73**, 725–729 (2008).
36. Mejia, Y. X., Mao, H., Forde, N. R. & Bustamante, C. Thermal Probing of E. coli RNA Polymerase Off-Pathway Mechanisms. *J. Mol. Biol.* **382**, 628–637 (2008).
37. Ó Maoiléidigh, D., Tadigotla, V. R., Nudler, E. & Ruckenstein, A. E. A unified model of transcription elongation: what have we learned from single-molecule experiments? *Biophys. J.* **100**, 1157–66 (2011).
38. Galburt, E. A., Grill, S. W. & Bustamante, C. Single molecule transcription elongation. *Methods* **48**, 323–332 (2009).
39. Mejia, Y. X., Nudler, E. & Bustamante, C. Trigger loop folding determines transcription rate of Escherichia coli's RNA polymerase. *Proc. Natl. Acad. Sci.* **112**, 743–748 (2015).
40. Imashimizu, M. *et al.* Visualizing translocation dynamics and nascent transcript errors in paused RNA polymerases in vivo. *Genome Biol.* **16**, 98 (2015).
41. Depken, M., Galburt, E. A. & Grill, S. W. The Origin of Short Transcriptional Pauses. *Biophys. J.* **96**, 2189–2193 (2009).
42. Seshasayee, A. S. N., Sivaraman, K. & Luscombe, N. M. An Overview of Prokaryotic Transcription Factors. in *A Handbook of Transcription Factors* (ed. Hughes, T. R.) 7–23 (Springer Netherlands, 2011). doi:10.1007/978-90-481-9069-0_2
43. Tetone, L. E. *et al.* Dynamics of GreB-RNA polymerase interaction allow a proofreading accessory protein to patrol for transcription complexes needing rescue. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E1081–E1090 (2017).
44. Komissarova, N. & Kashlev, M. Transcriptional arrest: Escherichia coli RNA polymerase translocates backward, leaving the 3' end of the RNA intact and extruded. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 1755–60 (1997).
45. Bubunenko, M. G. *et al.* A Cre Transcription Fidelity Reporter Identifies GreA as a Major RNA Proofreading Factor in Escherichia coli. *Genetics* **206**, 179–187 (2017).
46. Sivaramakrishnan, P. *et al.* The transcription fidelity factor GreA impedes DNA break repair. *Nature* **550**, 214–218 (2017).
47. Larson, M. H., Landick, R. & Block, S. M. Single-Molecule Studies of RNA Polymerase: One Singular Sensation, Every Little Step It Takes. *Mol. Cell* **41**, 249–262 (2011).
48. Davenport, R. J. Single-Molecule Study of Transcriptional Pausing and Arrest by E. coli RNA Polymerase. *Science (80-.).* **287**, 2497–2500 (2000).
49. Tolić-Nørrelykke, S. F., Engh, A. M., Landick, R. & Gelles, J. Diversity in the Rates of Transcript Elongation by Single RNA Polymerase Molecules. *J. Biol. Chem.* **279**, 3292–3299 (2004).
50. Imashimizu, M., Shimamoto, N., Oshima, T. & Kashlev, M. Transcription elongation. *Transcription* **5**, e28285 (2014).
51. Dulin, D. *et al.* Elongation-Competent Pauses Govern the Fidelity of a Viral RNA-Dependent RNA Polymerase. *Cell Rep.* **10**, 983–992 (2015).
52. Cnossen, J. P., Dulin, D. & Dekker, N. H. An optimized software framework for real-time, high-throughput tracking of spherical beads. *Rev. Sci. Instrum.* **85**, 103712 (2014).
53. Svetlov, V. & Artsimovitch, I. Purification of Bacterial RNA Polymerase: Tools and Protocols. 13–29 (2015).

54. Janissen, R. *et al.* Global DNA Compaction in Stationary-Phase Bacteria Does Not Affect Transcription. *Cell* **174**, 1188–1199.e14 (2018).
55. Dulin, D., Berghuis, B. A., Depken, M. & Dekker, N. H. Untangling reaction pathways through modern approaches to high-throughput single-molecule force-spectroscopy experiments. *Current Opinion in Structural Biology* **34**, 116–122 (2015).
56. Schwarz, G. Estimating the Dimension of a Model. *Ann. Stat.* **6**, 461–464 (1978).
57. Press, W. H., Teukolsky, S. a, Vetterling, W. T. & Flannery, B. P. *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. *Sample page from NUMBERICAL RECIPES IN C* **1**, (2007).
58. Biernacki, C., Celeux, G. & Govaert, G. Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE Trans. Pattern Anal. Mach. Intell.* **22**, 719–725 (2000).
59. Forde, N. R., Izhaky, D., Woodcock, G. R., Wuite, G. J. L. & Bustamante, C. Using mechanical force to probe the mechanism of pausing and arrest during continuous elongation by Escherichia coli RNA polymerase. *Proc. Natl. Acad. Sci.* **99**, 11682–11687 (2002).
60. Harrington, K. J., Laughlin, R. B. & Liang, S. Balanced branching in transcription termination. *Proc. Natl. Acad. Sci.* **98**, 5019–5024 (2001).
61. Turtola, M., Mäkinen, J. J. & Belogurov, G. A. Active site closure stabilizes the backtracked state of RNA polymerase. *Nucleic Acids Res.* **46**, 10870–10887 (2018).
62. Abbondanzieri, E. A., Greenleaf, W. J., Shaevitz, J. W., Landick, R. & Block, S. M. Direct observation of base-pair stepping by RNA polymerase. *Nature* **438**, 460–465 (2005).
63. Larson, M. H., Greenleaf, W. J., Landick, R. & Block, S. M. Applied Force Reveals Mechanistic and Energetic Details of Transcription Termination. *Cell* **132**, 971–982 (2008).
64. Ostrofet, E., Papini, F. S. & Dulin, D. Correction-free force calibration for magnetic tweezers experiments. *Sci. Rep.* **8**, 15920 (2018).
65. Gusarov, I. & Nudler, E. The Mechanism of Intrinsic Transcription Termination. *Mol. Cell* **3**, 495–504 (1999).
66. Sosunova, E., Sosunov, V., Epshtein, V., Nikiforov, V. & Mustaev, A. Control of Transcriptional Fidelity by Active Center Tuning as Derived from RNA Polymerase Endonuclease Reaction. *J. Biol. Chem.* **288**, 6688–6703 (2013).
67. Mishanina, T. V., Palo, M. Z., Nayak, D., Mooney, R. A. & Landick, R. Trigger loop of RNA polymerase is a positional, not acid–base, catalyst for both transcription and proofreading. *Proc. Natl. Acad. Sci.* 201702383 (2017). doi:10.1073/pnas.1702383114
68. Wang, B., Opron, K., Burton, Z. F., Cukier, R. I. & Feig, M. Five checkpoints maintaining the fidelity of transcription by RNA polymerases in structural and energetic details. *Nucleic Acids Res.* **43**, 1133–1146 (2015).
69. Sivaramakrishnan, P. *et al.* The transcription fidelity factor GreA impedes DNA break repair. *Nature* **550**, 214–218 (2017).
70. Borukhov, S., Sagitov, V. & Goldfarb, A. Transcript cleavage factors from E. coli. *Cell* **72**, 459–66 (1993).
71. Roldán, É., Lisica, A., Sánchez-Taltavull, D. & Grill, S. W. Stochastic resetting in backtrack recovery by RNA polymerases. *Phys. Rev. E* **93**, 062411 (2016).
72. Abbondanzieri, E. A., Shaevitz, J. W. & Block, S. M. Picocalorimetry of Transcription by RNA Polymerase. *Biophys. J.* **89**, L61–L63 (2005).
73. 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**, 684–687 (2003).
74. Malinen, A. M. *et al.* Active site opening and closure control translocation of multisubunit RNA polymerase. *Nucleic Acids Res.* **40**, 7442–7451 (2012).
75. Nedialkov, Y. A. *et al.* The RNA polymerase bridge helix YFI motif in catalysis, fidelity and translocation. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1829**, 187–198 (2013).
76. Bar-Nahum, G. *et al.* A Ratchet Mechanism of Transcription Elongation and Its Control. *Cell* **120**, 183–193 (2005).
77. Turtola, M., Mäkinen, J. J. & Belogurov, G. A. Active site closure stabilizes the backtracked state of RNA polymerase. *Nucleic Acids Res.* (2018). doi:10.1093/nar/gky883
78. Foster, J. E., Holmes, S. F. & Erie, D. A. Allosteric Binding of Nucleoside Triphosphates to RNA Polymerase Regulates Transcription Elongation. *Cell* **106**, 243–252 (2001).
79. Herbert, K. M., Greenleaf, W. J. & Block, S. M. Single-Molecule Studies of RNA Polymerase: Motoring Along. *Annu. Rev. Biochem.* **77**, 149–176 (2008).
80. Adelman, K. *et al.* Single molecule analysis of RNA polymerase elongation reveals uniform kinetic behavior. *Proc. Natl. Acad. Sci.* **99**, 13538–13543 (2002).
81. Zhang, J., Palangat, M. & Landick, R. Role of the RNA polymerase trigger loop in catalysis and pausing. *Nat. Struct. Mol. Biol.* **17**, 99–104 (2010).

82. Miropolskaya, N. *et al*. Interplay between the trigger loop and the F loop during RNA polymerase catalysis. *Nucleic Acids Res.* **42**, 544–552 (2014).