





ARTICLE

DOI: 10.1038/s41467-018-03902-9

OPEN

# Pausing controls branching between productive and non-productive pathways during initial transcription in bacteria

David Dulin <sup>1,2</sup>, David L.V. Bauer <sup>1</sup>, Anssi M. Malinen<sup>1,3</sup>, Jacob J.W. Bakermans <sup>1</sup>, Martin Kaller<sup>1</sup>, Zakia Morichaud<sup>4</sup>, Ivan Petushkov<sup>5</sup>, Martin Depken<sup>6</sup>, Konstantin Brodolin <sup>4</sup>, Andrey Kulbachinskiy<sup>5</sup> & Achillefs N. Kapanidis<sup>1</sup>

Transcription in bacteria is controlled by multiple molecular mechanisms that precisely regulate gene expression. It has been recently shown that initial RNA synthesis by the bacterial RNA polymerase (RNAP) is interrupted by pauses; however, the pausing determinants and the relationship of pausing with productive and abortive RNA synthesis remain poorly understood. Using single-molecule FRET and biochemical analysis, here we show that the pause encountered by RNAP after the synthesis of a 6-nt RNA (ITC6) renders the promoter escape strongly dependent on the NTP concentration. Mechanistically, the paused ITC6 acts as a checkpoint that directs RNAP to one of three competing pathways: productive transcription, abortive RNA release, or a new unscrunching/scrunching pathway. The cyclic unscrunching/scrunching of the promoter generates a long-lived, RNA-bound paused state; the abortive RNA release and DNA unscrunching are thus not as tightly linked as previously thought. Finally, our new model couples the pausing with the abortive and productive outcomes of initial transcription.

<sup>1</sup>Biological Physics Research Group, Clarendon Laboratory, Department of Physics, University of Oxford, Parks Road, Oxford OX1 3PU, UK. <sup>2</sup>Junior Research Group 2, Interdisciplinary Center for Clinical Research, Friedrich-Alexander-University Erlangen-Nürnberg (FAU), Hartmannstrasse 14, 91052 Erlangen, Germany. <sup>3</sup>Department of Biochemistry, University of Turku, 20014 Turku, Finland. <sup>4</sup>Institut de Recherche en Infectiologie de Montpellier (IRIM) UMR9004 CNRS-Université de Montpellier, 1919 Route de Mende, 34293 Montpellier, France. <sup>5</sup>Institute of Molecular Genetics, Russian Academy of Sciences, Moscow 123182, Russia. <sup>6</sup>Department of Bionanoscience, Kavli Institute of Nanoscience, Delft University of Technology, Van der Maasweg 9, 2629 HZ Delft, The Netherlands. These authors contributed equally: David L. V. Bauer, Anssi M. Malinen. Correspondence and requests for materials should be addressed to D.D. (email: [david.dulin@uk-erlangen.de](mailto:david.dulin@uk-erlangen.de)) or to A.N.K. (email: [kapanidis@physics.ox.ac.uk](mailto:kapanidis@physics.ox.ac.uk))

**T**ranscription initiation by DNA-dependent RNA polymerase (RNAP) constitutes the first and often decisive step in gene expression in bacteria. To balance the output of transcription with environmental and cellular needs, an extensive set of molecular mechanisms has evolved to regulate the efficiency and specificity of transcription initiation<sup>1</sup>. The regulatory mechanisms are either directly encoded in the transcribed DNA sequence or mediated by protein transcription factors or small-molecule signals. The target of transcription initiation regulators may be the function of RNAP itself, or the accessibility or affinity of promoters for RNAP. Further regulation occurs in the elongation and termination phases of transcription<sup>2–5</sup>.

To perform promoter-specific transcription initiation, the bacterial RNAP core associates with housekeeping  $\sigma^{70}$  initiation factor (or one of the alternative  $\sigma$ -factors) to form an RNAP holoenzyme<sup>6,7</sup>. The holoenzyme employs sequence-specific interactions between the  $\sigma^{70}$  and the  $-35$  and  $-10$  promoter elements (Fig. 1a) to form an initial RNAP–DNA closed complex, and to isomerize to the catalytically competent RNAP–promoter open complex (RP<sub>O</sub>)<sup>8,9</sup> (Fig. 1b). During initial RNA synthesis, strong interactions with the DNA hold the RNAP at the promoter, resulting in the build-up of “scrunching” of downstream DNA, a conformational change that increases the size of the DNA bubble<sup>10–13</sup>. The eventual break-up of RNAP–promoter contacts and the escape to elongation relax the scrunched DNA<sup>11</sup>. The productive promoter escape pathway competes with abortive initiation, an unproductive pathway wherein the short nascent RNA is thought to dissociate prematurely, resetting the initially transcribing complex (ITC) to RP<sub>O</sub><sup>11,14–18</sup>. Although conformational strain resulting from the DNA scrunching may promote abortive initiation<sup>11</sup>, multiple other factors such as the presence of the  $\sigma_{3,2}$  region (which obstructs the entry to the RNA-exit channel<sup>19–22</sup>), strong RNAP–promoter interactions<sup>9,16,23</sup>, and the initially transcribed sequence<sup>24,25</sup> also contribute.

The step that defines the overall rate of transcription initiation varies between promoters<sup>9,16,23</sup>. In many  $\sigma^{70}$ -dependent promoters, the rate-limiting step is attributed to the half-life of RP<sub>O</sub> or the rate of promoter escape. An extensively studied example of an escape-limited promoter is *lacUV5*<sup>26</sup>, which produces substantial amounts of abortive products; further, transcriptional pausing was identified in the ITC formed on *lac* promoter after the synthesis of 6-nt RNA, in part due to the clash of the 5'-RNA end with the  $\sigma_{3,2}$  region<sup>27,28</sup>.

Recent advances in structural characterization of bacterial transcription initiation complexes have created intriguing hypotheses on how specific molecular interactions and conformational changes drive holoenzyme formation, promoter recognition, open complex formation<sup>29</sup>, and initial RNA synthesis<sup>12,20,30</sup>. Complementing this structural insight with detailed functional analysis is hampered, however, by the multi-step, asynchronous nature of transcription initiation pathways. Single-molecule techniques, which can provide a direct readout for several mechanistic steps and resolve co-existing reaction pathways, are well positioned to overcome the complexity of transcription initiation<sup>31–34</sup>.

Here we combine single-molecule and biochemical analysis of initial transcription to explore the mechanistic basis of the pause encountered by ITC6 on a consensus variant of *lac* promoter, i.e., with consensus  $-35$  and  $-10$  elements, and a 17 bp spacer. We present evidence that the ITC6 pause represents a major control point where the ITCs branch to three competing downstream reaction pathways: pause exit by productive transcription; abortive-RNA release; and slow cycling between DNA conformations with different extents of scrunching but without RNA release. The partitioning between these three paths and their kinetics depends on distinct interactions and structural elements.

The rate of productive pause exit is synergistically controlled by the initial transcribed sequence and the interaction of the 5'-RNA end with  $\sigma_{3,2}$  region, whereas perturbing RNAP interactions with the initially transcribed region favors the entry into the scrunching/unscrunching pathway.

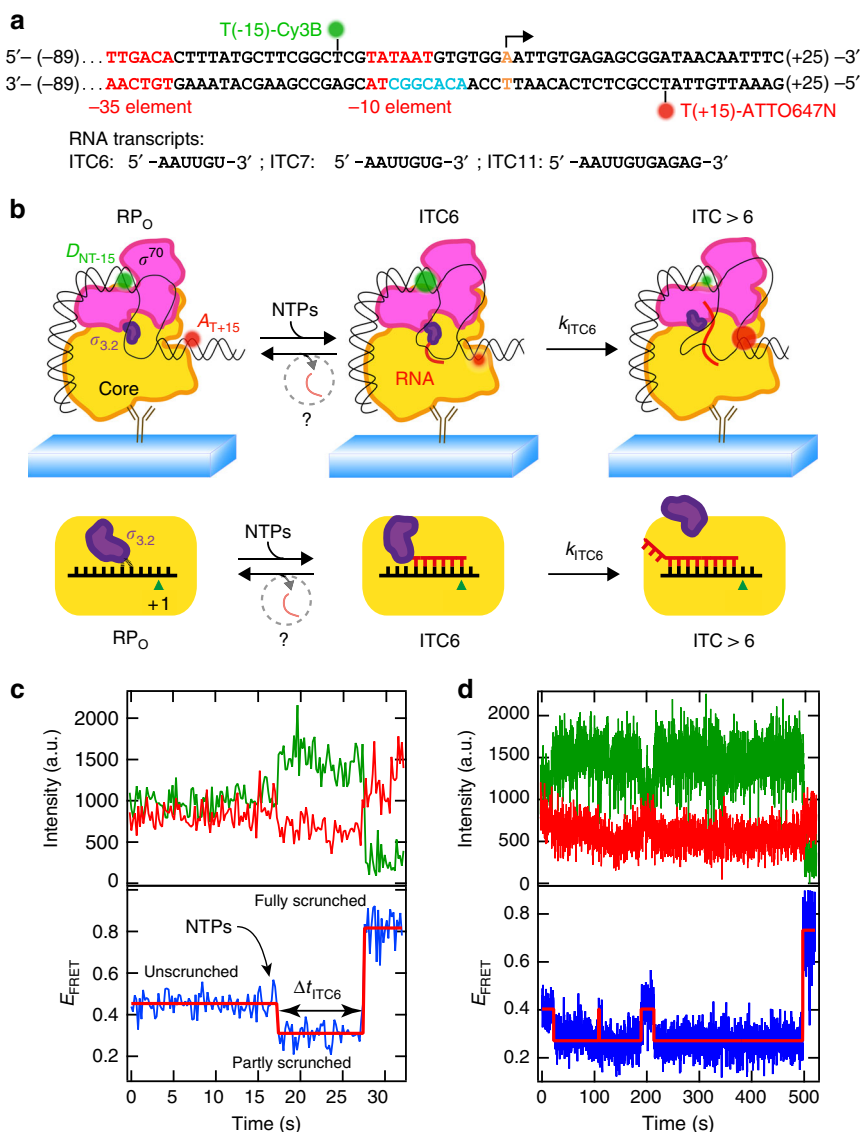
## Results

**High-resolution observation of initial transcription.** To monitor the kinetics of transcription initiation at the single-molecule level, we developed a Förster resonance energy transfer (FRET) sensor with fluorophore-labeled consensus *lac* promoter for real-time imaging of individual transcription-engaged RNAP–promoter complexes. Earlier studies of this consensus *lac* promoter derivative identified a strong pause at ITC6<sup>27,35</sup>. We modified the original promoter design in two ways to allow in-depth biophysical analysis of the rate-limiting ITC6 pause (Fig. 1a and Supplementary Fig. 1a). First, we extended the upstream region of the promoter from  $-39$  to  $-89$ , to enhance RP<sub>O</sub> formation and provide a more native DNA-length context for RNAP–DNA interactions<sup>24,36</sup>. Second, we moved the acceptor dye from position  $+20$  to  $+15$ , to obtain distinct FRET signals for different steps of the initiation pathway. With this configuration, we clearly separated and calibrated (Methods) FRET readouts for three structural states through downstream DNA scrunching<sup>10</sup>: an unscrunched (US) open complex RP<sub>O</sub> ( $E_{\text{FRET}} = 0.49 \pm 0.003$ ), a partly scrunched (PS) paused complex ITC6 ( $E_{\text{FRET}} = 0.37 \pm 0.001$ ), and a fully scrunched (FS) pause-cleared complex at ITC11 ( $E_{\text{FRET}} = 0.80 \pm 0.002$ ) (Fig. 1b and Supplementary Fig. 1).

Upon addition of NTPs to RP<sub>O</sub> complexes (either a NTP subset sufficient to reach the ITC11 complex or all NTPs), the FRET signal showed an almost instant transition from the US to the PS state (Fig. 1c, Supplementary Fig. 1b and Supplementary Fig. 1c), suggesting that the transcription complexes synthesized 6-mer RNA and paused. After the pause, the ITCs split into two main populations: the first population comprised “productive” ITCs that resumed transcription and progressed from the PS to the FS state by synthesizing an 11-mer (Fig. 1c). The second population comprised ITCs that returned from the PS to the US state (Fig. 1d); notably, such complexes could cycle multiple times (e.g., at  $\sim 100$  s and  $\sim 200$  s in Fig. 1d) between PS and US states until they eventually reached the FS state (e.g., at  $\sim 500$  s in Fig. 1d).

**Determination of the lifetime of the ITC6 pause.** Two elements appear to contribute to RNAP pausing at ITC6: (i) the clash of 5'-RNA end with the  $\sigma_{3,2}$  region (Fig. 1b), which blocks entry to the RNA-exit channel of RNAP<sup>27</sup>, and (ii) a specific sequence motif (a non-template Y<sup>+</sup>G<sup>+</sup> in the transcribed DNA strand<sup>35</sup>) akin to that causing sequence-specific pausing in elongation<sup>37–39</sup>. We dissected the contributions of these two elements to the ITC6 pause using our single-molecule FRET assay.

To explore the steric-clash hypothesis, we modified the 5'-RNA end of the nascent transcript (and thus its interaction with  $\sigma_{3,2}$ ) by initiating transcription either using a synthetic dinucleotide (ApA) or using ATP, which adds a 5'-triphosphate tail to the 5'-RNA end. To evaluate the effect of the pause sequence motif on the detailed dynamics of initial transcription, we replaced the sequence T<sup>+</sup>G<sup>+</sup> (on non-template DNA) with G<sup>+</sup>T<sup>+</sup>, creating a “ $\Delta$ P promoter” (Supplementary Fig. 1a) that increased the overall rate of initial transcription by shortening the ITC6 pause<sup>35</sup>. In all experiments, the initiating ATP or ApA were held at 500  $\mu$ M, a level significantly above the  $K_M$  of the RP<sub>O</sub> for initiating nucleotides and dinucleotide primers<sup>22</sup>; we also varied the concentration of remaining NTPs (1–500  $\mu$ M).

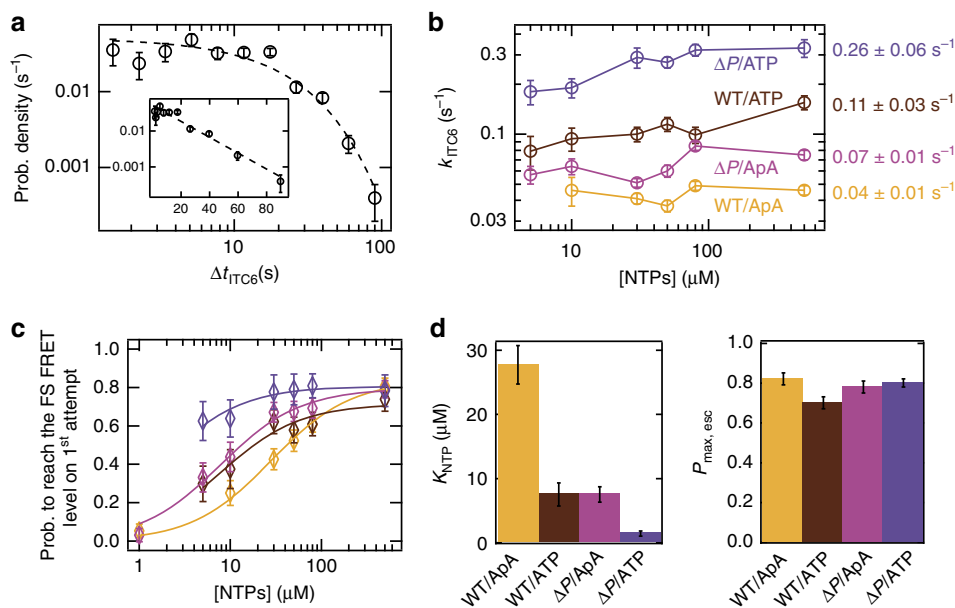


**Fig. 1** Initial transcription monitored at the single-molecule level. **a** Representation of the premelted (turquoise font) WT DNA promoter used in the single-molecule experiments (Supplementary Fig. 1a). The −35 and −10 elements are represented in red. The promoter was donor labeled at −15 position (green sphere) of the non-template DNA strand and acceptor labeled in +15 (red sphere) position of the template DNA strand. An arrow above the base in orange font indicates the +1 position. All the promoters used in the study are described in Supplementary Fig. 1a. **b** Schematic of the initial transcription experiment (Methods). Above: using TIRFM-based smFRET, we monitored the  $E_{FRET}$  variations of the donor-acceptor pair upon NTP addition. The RNAP fluctuates between RP<sub>0</sub>, ITC6, and ITC > 6, to eventually escape the initiation phase toward the elongation phase, or to release the nascent RNA; below: cartoon that magnifies the interactions between the 5'-RNA end and  $\sigma_{3.2}$  and the position of the 5'-RNA end. **c** Fluctuations in the donor (green) and acceptor (red) dyes intensities (above) and the resulting  $E_{FRET}$  (below, blue), showing the variation of  $E_{FRET}$  from an Unscrunched (US) FRET state, followed by the Partly Scrunched (PS) FRET state upon NTP addition, and ending in the Fully Scrunched (FS) FRET state. Experimental conditions: 200 ms time points (100 ms ALEX, Methods), 500  $\mu$ M ApA, and 80  $\mu$ M All NTP. **d** Similar experiment conducted as described in **c**, with the RP failing multiple times before reaching the FS FRET state. Experimental conditions: 200 ms time points, 500  $\mu$ M ApA, and 30  $\mu$ M of all NTPs and WT promoter. The red solid lines in the lower panels in **c** and **d** represent the FRET states extracted from hidden Markov modeling (HMM) (Methods)

We first analyzed the effects of the pause elements on the pause duration at ITC6 ( $\Delta t_{ITC6}$ ) by focusing on the subpopulation of molecules displaying the US→PS→FS scrunching sequence (as in Fig. 1c). The dwell-time distribution for the ITC6 pause was well described by a single exponential (Fig. 2a; see Methods for analysis procedure), where the exit rate  $k_{ITC6}$  was ~1.5-fold lower for the wild-type (WT) promoter compared with the  $\Delta P$  promoter using ApA (Fig. 2b). When we replaced ApA with ATP as the starting substrate and employed the remaining NTPs at above 30  $\mu$ M, the ITC6 pause exit rate increased from  $0.07 \pm 0.01$  to  $0.26 \pm 0.06$  s<sup>−1</sup> for the  $\Delta P$

promoter and from  $0.04 \pm 0.01$  to  $0.11 \pm 0.03$  s<sup>−1</sup> for the WT promoter, i.e., 2.5-fold enhancement in  $k_{ITC6}$  in the absence of the pause motif (Fig. 2b). These experiments demonstrate that the ITC6 pause duration is controlled both by the transcribed sequence and by the structure of the RNA 5'-end, which interacts with  $\sigma_{3.2}$ .

We also noted that the NTP concentration did not influence significantly the  $k_{ITC6}$  (no more than 1.5-fold) for the WT promoter with either ApA or ATP as the starting substrate, or for the  $\Delta P$  promoter with ApA as the starting substrate (Fig. 2b). The rate-limiting step in all these cases is thus neither the intrinsic



**Fig. 2** The pause characteristics are a function of the downstream DNA promoter sequence at the +6 position and of the 5'-RNA end nature. **a** Probability density distribution for the  $\Delta t_{\text{ITC6}}$  for the RP complexes behaving as in Fig. 1c. The dashed line is a single-exponential fit from a MLE. Inset: Log-linear representation of the same data. Experimental conditions: 500  $\mu\text{M}$  ApA starting substrate, 80  $\mu\text{M}$  all NTPs. **b**  $\Delta t_{\text{ITC6}}$  lifetime extracted from single exponential MLE fit similar to **a** for different promoter/starting substrate conditions (as indicated in the panel), different NTP conditions, i.e., all NTPs for WT/ApA (yellow) and ITC11 (Supplementary Fig. 1a) for all others, and different NTP concentrations (Supplementary Table 1). In the ATP-initiated reactions, we did not use NTP concentration below 5  $\mu\text{M}$  to prevent potential misincorporations of ATP (used at 500  $\mu\text{M}$  for initiation purposes)<sup>69</sup>. On the right-hand side is indicated the mean  $\pm$  SD of  $k_{\text{ITC6}}$  for each promoter/starting substrate condition. **c** Probability to reach the fully scrunched (FS) FRET level in a single attempt (Fig. 1c). The solid lines are fits to a binding isotherm of the form  $p(\text{NTP}) = P_{\text{max,esc}} \times [\text{NTP}] / ([\text{NTP}] + K_{\text{NTP}})$ . The error bars are 95% confidence intervals. **d**  $K_{\text{NTP}}$  and  $P_{\text{max,esc}}$  extracted from **c**. Error bars are 1SD extracted from the fit

catalytic activity of the transcription complex nor the binding of the incoming NTP substrate.

We next characterized the probability to exit the ITC6 pause on the first attempt (Fig. 2c). For this purpose, we counted the probability of ITCs to proceed via the reaction path depicted in Fig. 1c (single-scrunch pathway), with or without a detectable ITC6 pause, versus the path in Fig. 1d (cyclic scrunching/unscrunching pathway). ATP-initiated  $\Delta P$  promoter complexes (Fig. 2c) exited the pause on the first attempt at higher probability at all tested (5–500  $\mu\text{M}$ ) NTP concentration compared with the other starting substrate-promoter conditions. The pause-exit probability for the ApA-initiated  $\Delta P$  promoter, and the ATP- or ApA-initiated WT promoter decreased steeply from 0.8 toward 0 at low NTP concentrations (Fig. 2c). By fitting the probability  $p(\text{NTP})$  to exit the ITC6 pause on the first attempt with a descriptive model similar to a binding isotherm (Fig. 2c), we extracted an apparent binding constant  $K_{\text{NTP}}$  and a maximal pause-exit probability  $P_{\text{max,esc}}$  (Fig. 2d). Overall, the WT promoter had a higher  $K_{\text{NTP}}$  compared with  $\Delta P$  promoter complexes ( $28 \pm 3$  vs.  $8 \pm 1$   $\mu\text{M}$ , ApA), whereas ATP-initiated complexes had a lower  $K_{\text{NTP}}$  compared with ApA-initiated ones ( $8 \pm 2$  vs.  $28 \pm 3$   $\mu\text{M}$ , WT promoter;  $1.7 \pm 0.4$  vs.  $8 \pm 2$   $\mu\text{M}$ ,  $\Delta P$  promoter). The probability  $P_{\text{max,esc}}$  was relatively constant, with  $79 \pm 5\%$  of the molecules reaching the FS FRET level on the first attempt at saturating NTP concentration. These results suggest that ITCs can exit a weak ITC6 pause ( $\Delta P$  promoter) efficiently even at low NTP concentration, while overcoming a strong ITC6 pause (WT promoter + ApA at the 5'-RNA end) requires higher NTP concentration (Fig. 2c). Interestingly, we observed that  $98 \pm 2\%$  and  $89 \pm 3\%$  of the complexes paused at ITC6 on WT promoter and  $\Delta P$  promoter, respectively (averaged percentages for all NTP concentrations and ATP starting substrate), indicating that  $T^{+6}G^{+7}$  motif enforces pausing at ITC6 (Supplementary Fig. 2b and Supplementary Note 1).

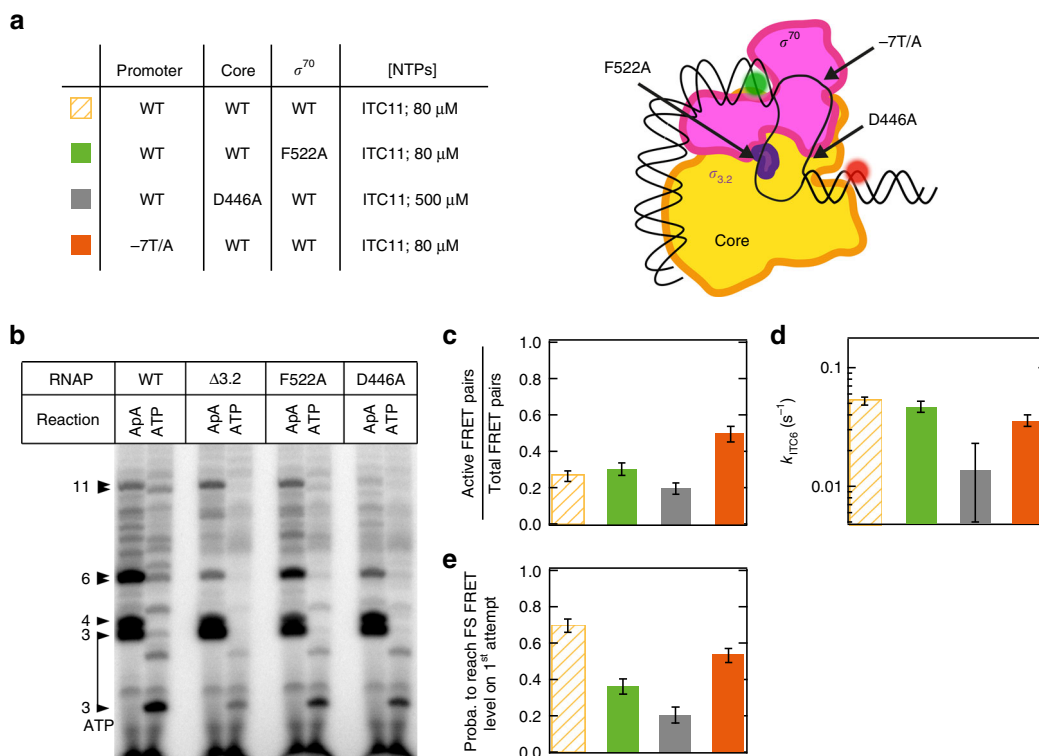
Finally, a fully double-stranded promoter (dsWT, Supplementary Fig. 1a) did not modify the ITC6 pause exit rate both for ApA and ATP starting substrates (Supplementary Fig. 2c), whereas the probability to reach the FS state during the first attempt on this promoter was also strongly decreased in the absence of a 5'-RNA end triphosphate ( $14 \pm 4\%$  vs.  $58 \pm 5\%$ , Supplementary Fig. 2d), suggesting again that the 5'-RNA end triphosphate assists in the ITC6 pause exit.

### Structural determinants of transcription pathway partitioning.

Our single-molecule reaction trajectories demonstrated (Fig. 1) that the transcription complexes paused at ITC6 may either resume RNA extension or cycle between differently scrunched paused states. To establish the interactions contributing to the pathway partitioning, we engineered structural changes (Fig. 3a) in RNAP,  $\sigma^{70}$ , and nucleic acids, and characterized the impacts on the function of ITCs.

We first explored the significance of  $\sigma_{3,2}$ -template-strand DNA interaction by using F522A substituted  $\sigma^{70}$ , which is deficient in an interaction between the -4 template DNA base and  $\sigma_{3,2}$ <sup>29</sup>, and affects initial transcription<sup>22</sup>. During transcription initiation, the F522A substitution decreased the amounts of  $\leq 6$  nt RNA products, likely because it decreased the physical barrier from their clash with the  $\sigma_{3,2}$  region<sup>22</sup>, but had little effect on the ITC11 formation (Fig. 3b). In the FRET assay, the F522A  $\sigma^{70}$  derivative retained similar activity (Fig. 3c and Supplementary Fig. 2e) and  $k_{\text{ITC6}}$  as the WT (Fig. 3d and Supplementary Fig. 2f). Instead, the substitution significantly decreased the fraction of complexes exiting the pause on the first attempt ( $70 \pm 4\%$  to  $37 \pm 4\%$  for ApA starting substrate), independently of the use of ApA (Fig. 3e) or ATP (Supplementary Fig. 2g) as the starting substrate. The weakening of  $\sigma_{3,2}$  interaction with the template-strand DNA thus destabilizes the PS promoter conformation and biases the paused





**Fig. 3** Core and  $\sigma^{70}$  mutants affect ITC6 pause exit probability. **a** Experimental conditions studied here and schematic of the different RP complex variants, all with ApA starting substrate. The same experiments have been performed with ATP starting substrate and are presented in Supplementary Fig. 2e, f, g. **b** Bulk reaction performed with a consensus *lac* promoter (Supplementary Fig. 1a) and different holoenzyme mutants, in ITC11 NTP conditions (500  $\mu$ M ApA and 80  $\mu$ M ATP/UTP/GTP or 500  $\mu$ M ATP and 80  $\mu$ M UTP/GTP) and 1 min incubation at 37  $^{\circ}$ C (the experimental procedure is described in the Supplementary Information). The band assignment is indicated on the left hand side of the gel. **c** The fraction of transcriptionally active RPs, which displayed NTP-dependent  $E_{\text{FRET}}$  changes, of all surface-immobilized RPs (see also, Methods: FRET pair localization and detection), in the experimental conditions described in **a**. **d** ITC6 pause exit rate for the experimental conditions described in **a**. **e** Probability to reach the FS FRET level on the first attempt for the experimental conditions described in **a**. Error bars are either 1SD from 1000 bootstraps procedure **d** or 95% confidence interval **c**, **e**

ITC6 toward the scrunching/unscrunching pathway. In comparison, the  $\sigma_{3.2}$  region deletion has a much more severe defects in transcription initiation and strictly requires dinucleotide primers for 11-mer synthesis<sup>22,27</sup> (Fig. 3b).

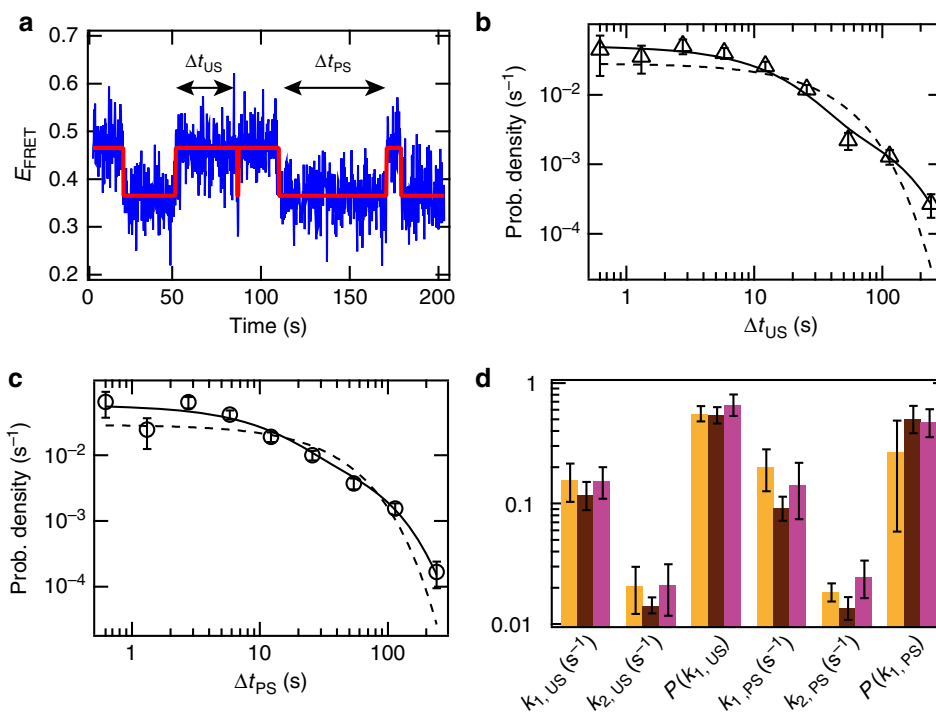
We next destabilized by  $\beta$ D446A substitution the binding of non-template guanine in the “CRE-pocket” (Core Recognition Element-pocket) of RNAP<sup>29,37</sup>. Our results align with an earlier observation of the CRE-pocket being involved in open complex formation and transcription start-site selection<sup>40</sup>; we observed three-fold less active RNAP–promoter complexes (ApA starting substrate, Supplementary Fig. 2e), which is only partly recovered when increasing the NTP concentration to 500  $\mu$ M (Fig. 3c). We also observed, similar to the consensus elongation pause<sup>37</sup>, a reduced escape rate ( $0.10 \pm 0.01$  vs.  $0.06 \pm 0.01$  s<sup>-1</sup>) from the ITC6 pause (Supplementary Fig. 2f). Both pauses are characterized by the presence of Y<sup>-1</sup>G<sup>+1</sup> ntDNA motif. Our third observation reveals a potentially novel function for the CRE-pocket in helping to maintain the scrunching promoter DNA conformation; fourfold less ( $20 \pm 5\%$  vs.  $70 \pm 4\%$ ) complexes managed to escape the pause on the first attempt (Fig. 3e). In bulk transcription reactions, only a minor fraction of complexes formed by the D446A RNAP could extend RNA beyond six nucleotides, suggesting that the contacts of D446 with +7G are critically important to exit the paused state (Fig. 3b)<sup>35</sup>.

To probe the effects of weakened interactions between  $\sigma$  region 2 and the -10 promoter element, we replaced the consensus -7 thymine in the non-template DNA by an adenine (-7T/A, Supplementary Fig. 1a). The -7 thymine is inserted into a pocket of  $\sigma$  in RP<sub>O</sub><sup>29,41,42</sup>. However, as -7T/A substitution resulted in

only small changes in the ITC6 pause exit rate and the fraction of complexes exiting the pause on the first attempt (Fig. 3c, d), -7T- $\sigma$  interaction appears to have a minor role after RNA length is  $\geq 6$ .

**Kinetics of promoter unscrunching/scrunching.** We next analyzed the molecules whose pausing at ITC6 was followed by cyclic unscrunching/scrunching events. These molecules may cycle for tens to hundreds of seconds between the PS and US states until they reach the FS FRET level or the dyes bleached (Figs. 1d and 4a). For this cycling population, we generated probability density distributions for the dwell times in PS ( $\Delta t_{\text{PS}}$ ) and US ( $\Delta t_{\text{US}}$ ) states (Fig. 4b, c). Both PS and US distributions showed a similar trend, with dwell times varying from  $\sim 0.4$  s to  $\sim 200$  s. The distributions were fitted well by a two-exponential probability distribution (solid lines, Fig. 4b, c; dashed lines depict a single-exponential function) (Methods). The fit thus defines for the US and PS states the exit rates  $k_1$  and  $k_2$ , as well as the probability  $P(k_1)$  to exit a state with rate  $k_1$  (Supplementary Fig. 2e, g).

We applied this data analysis to WT promoter reactions initiated with ApA or ATP, and the  $\Delta P$  promoter initiated with ApA (Fig. 2b). We did not include ATP-initiated  $\Delta P$  promoter results because most complexes exited the ITC6 pause directly to the FS state (Fig. 2c). We first noted that the exit rates  $k_1$  and  $k_2$ , as well as the  $P(k_1)$  probabilities of PS and US states, remained fairly constant in all used NTP concentrations (Supplementary Fig. 2h, j). We observed a single exception with the ITC on the ApA-initiated WT promoter, which showed a decreased probability  $P(k_1, \text{PS})$  at higher NTP concentrations (right panel, Supplementary Fig. 2h). We observed, in average that the US and



**Fig. 4** RP complexes that do not reach the FS FRET level alternate between US and PS FRET levels for long period of time. **a** Typical  $E_{\text{FRET}}$  trace where the RP alternates between unscrunched (US) and partly scrunched (PS) DNA promoter FRET levels. The red solid line represents the FRET levels extracted from empirical Bayesian probability hidden Markov model. We collect the dwell times  $\Delta t$  for each FRET level. Experimental condition: WT DNA promoter; NTP start: ApA; NTP: U/G/ATP 80  $\mu\text{M}$ . **b, c** Probability density distribution of the dwell times  $\Delta t_{\text{US}}$  and  $\Delta t_{\text{PS}}$ , respectively, for the experimental conditions described in **a**, with its single and double exponential MLE fit (dashed and solid line, respectively, Methods for MLE fit procedure). **d** Average for all the NTP concentration (Supplementary Fig. 2e–g) of the double exponential MLE parameters ( $k_1$  and  $k_2$  exit rates, and  $P$  probability of being in the exponential with  $k_1$  exit rate) for US and PS FRET states averaged over all NTP concentrations, for the conditions described in Fig. 2a and Supplementary Table 1. Color scheme as in Fig. 2

PS states had nearly identical kinetic parameters:  $k_1 \sim 0.15 \text{ s}^{-1}$ ,  $k_2 \sim 0.02 \text{ s}^{-1}$ , and  $P(k_1) \sim 0.6$ . Notably, these values were also independent of the NTP subset used (allowing maximal transcript lengths 7 or 11), the nature of the RNA 5'-end, or any of the tested RNAP-promoter variations (Fig. 4d and Supplementary Fig. 2k, l, m).

The remarkable insensitivity of the unscrunching/scrunching kinetics to the tested parameters, and in particular to the NTP concentration, strongly suggests that the complexes entering the unscrunching/scrunching pathway are catalytically inactive. Even though the complexes are ultimately able to resume RNA synthesis, the unscrunching/scrunching pathway delays the promoter clearance by tens of seconds (Fig. 1d).

**Promoter unscrunching and RNA release.** The discovery of extensive unscrunching/scrunching cycling raises intriguing questions about its relationship with abortive initiation. Is the nascent RNA released or retained in unscrunching event, and is the subsequent re-scrunching driven by the synthesis of a new RNA (Fig. 5a)? To address the questions, we devised a three-step assay (Fig. 5b) in several conditions (Fig. 5c): (i) RNA synthesis was triggered for  $\sim 10 \text{ s}$  with ITC11 (Fig. 5e, g) or ITC7 (Supplementary Fig. 3a) NTP subsets, (ii) the surface was extensively rinsed to remove NTPs, (iii) surface-bound complexes were re-imaged. To our surprise,  $28 \pm 4\%$  (ApA, ITC11) and  $18 \pm 2\%$  (ATP, ITC11) of initially active complexes cyclically unscrunched/scrunched also in the absence of NTPs (Fig. 5d). The fraction of unscrunching/scrunching molecules was comparable to those observed in the continuous presence of NTPs ( $27 \pm 3\%$ , ApA starting substrate;  $42 \pm 4\%$  ATP

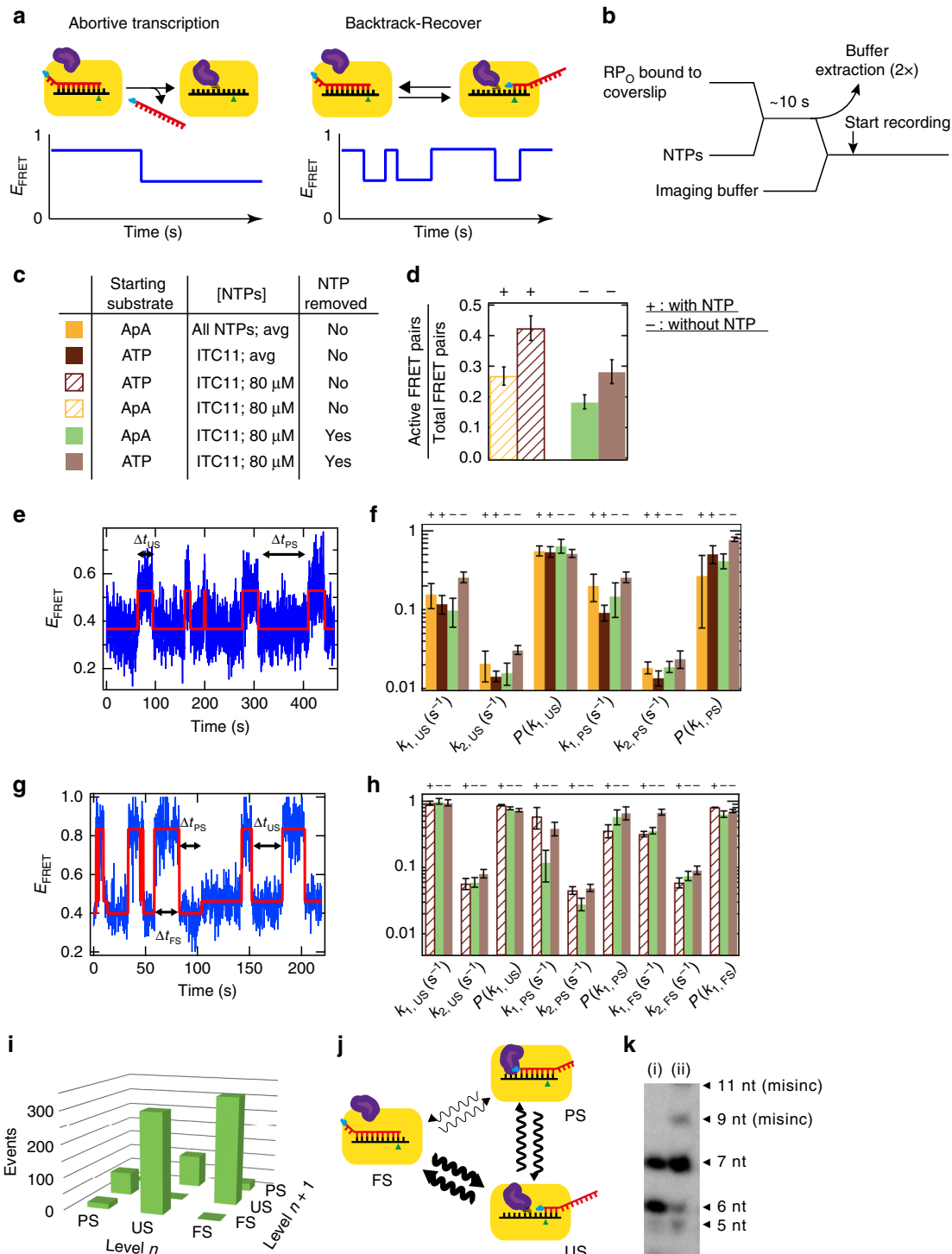
(Fig. 5d). Scrunching/unscrunching by the NTP-depleted complexes lasted for hundreds of seconds, being only limited by dye bleaching (Fig. 5e, g). Consistent with the maximal RNA length, complexes pulsed with ITC7 NTPs sampled only US and PS states (Supplementary Fig. 3a), whereas complexes pulsed with ITC11 NTPs could additionally occupy the FS state (Fig. 5g). This confirms that we did not observe re-synthesis of a 11mer transcript, as we observed earlier that the complex could not escape the ITC6 pause at low NTP concentration (Fig. 2c). Our results clearly demonstrate that the extended cycling in different scrunching states does thus not require active RNA synthesis.

Analysis of the complexes pulsed with ITC11 NTPs identified two types of unscrunching/scrunching molecules: the first cycled between US and PS FRET levels only (Fig. 5e), whereas the second cycled between US, PS, and FS FRET levels (Fig. 5g). The US/PS subpopulation included  $48 \pm 6\%$  (ApA starting substrate) or  $39 \pm 5\%$  (ATP starting substrate) of all cycling molecules, respectively (Supplementary Table 1). The US/PS and US/PS/FS subpopulations likely represent the ITCs, which at the moment of NTP withdrawal had synthesized 6- and 11-nt RNAs, respectively. The US/PS subpopulation showed similar unscrunching/scrunching kinetics to what was observed with NTPs, i.e.,  $k_1 = 0.16 \pm 0.07 \text{ s}^{-1}$ ,  $k_2 = 0.02 \pm 0.004 \text{ s}^{-1}$ , and  $P(k_1) = 0.57 \pm 0.05$  (Fig. 5f). The US/PS/FS subpopulation instead sampled all scrunching states almost an order of magnitude faster compared to the US/PS subpopulation (i.e.,  $k_1 = 0.96 \pm 0.04 \text{ s}^{-1}$ ,  $k_2 = 0.07 \pm 0.01 \text{ s}^{-1}$ , and  $P(k_1) = 0.79 \pm 0.07$  for US, Fig. 5h), independently of using ATP or ApA for initiation (Fig. 5f, h).

Close inspection of the trajectories belonging to the US/PS/FS subpopulation revealed that the two most frequently encountered state transitions were FS→US and its reversal US→FS (Fig. 5i and Supplementary Fig. 3f); this was also the case in the continuous presence of NTP (Supplementary Fig. 3g). The US→PS and PS→US transitions were about 4-fold less frequent, whereas PS→FS or FS→PS transitions were only rarely observed. This data clearly indicate that RPs engaged in the unscrunching/scrunching pathway do not share the same linear US→PS→FS reaction coordinate of ITCs engaged in productive transcription (Fig. 5j). We also note the absence of any temporal correlation between two successive state dwell times ( $dt_n$  and  $dt_{n+1}$ ), independently

of the scrunching state they originate from (right hand side, Supplementary Fig. 3b, c, d), supporting a memory-less transition from one state to the next.

**Paused ITC may undergo abortive initiation or hold RNA.** Our FRET assay monitors the conformation of the promoter DNA and thus does not provide a direct readout for the presence of RNA in the ITCs. As pulsed RNA synthesis generated ITCs that cycle for several min between scrunching states, we assumed that these ITCs retain the nascent RNA in the transcription bubble. The assumption generates two testable hypotheses: first, RNA is



slowly released from NTP-deprived ITCs; second, RNAs retained in ITCs are extendable upon NTP reintroduction.

To determine the profile and time dependence of RNA release from ITCs, we immobilized biotinylated RP<sub>O</sub> complexes to streptavidin-coated magnetic beads. The complexes were pulsed for 10 s with the ITC7 NTP subset (containing  $\alpha^{32}\text{P}$ -UTP), pulled down, washed, and immersed into NTP-free reaction buffer; beads and supernatant were then analyzed at specified times to obtain the time-dependent profile of retained and released RNAs (Supplementary Fig. 4a, b). Our results showed that the RNA-release kinetics was strikingly biphasic: many ITCs released their RNA within the first 2 min, the release being almost quantitative for the shortest RNAs (~95% of 3–4-mers) and less efficient for 5-, 6-, and 7-nt RNAs (45%, 80%, and 80%, respectively; Supplementary Fig. 4b). After the rapid initial phase, the amount of released 6- or 7-nt RNA increased only marginally. After 15 min, still ~20% of 6–7-nt RNA remained bound in the ITCs. This amount is two-fold lower than what we measured in similar NTP-pulsed single-molecule experiments, where most of the active ITCs were sampling the unscrunching/scrunching states for several min (Fig. 5d). We also tested the retention of short RNAs in transcription complexes formed under the ITC11 conditions and found that comparable amounts of 6-nt transcripts remained bound to RNAP even after prolonged incubation of the complexes in the absence of NTPs (Supplementary Fig. 4c). This is in agreement with the PS and FS FRET states observed in the absence of NTPs (Fig. 5e, g).

To probe whether the stalled ITCs retaining 6-nt RNA for an extended period of time can resume active transcription, we chased the immobilized and washed ITCs with the next incoming nucleotide (GTP). We observed that the 6-nt RNA became converted quantitatively to 7-nt RNA (Fig. 5k; longer products appear due to mis-incorporation), indicating that the ITCs both retain the nascent RNA in the transcription bubble, and can access the catalytically active conformation.

In summary, the biochemical analysis revealed two populations of stalled ITCs: 70–80% of ITCs that enter the abortive initiation pathway (rapidly releasing the nascent RNAs) and 20–30% of ITCs that retain 6–7-nt RNA products and catalytic competence for tens of min after NTP depletion. These results clearly show that the nascent RNA can be stably trapped within the cyclically unscrunching/scrunching RP complex, until being eventually elongated.

## Discussion

In this study, we employed a refined, high contrast single-molecule FRET assay to quantitatively dissect the reaction pathway and kinetics of the ITCs on the consensus *lac* promoter. We specifically examined the role of the  $\sigma_{3,2}$  region, the nature of

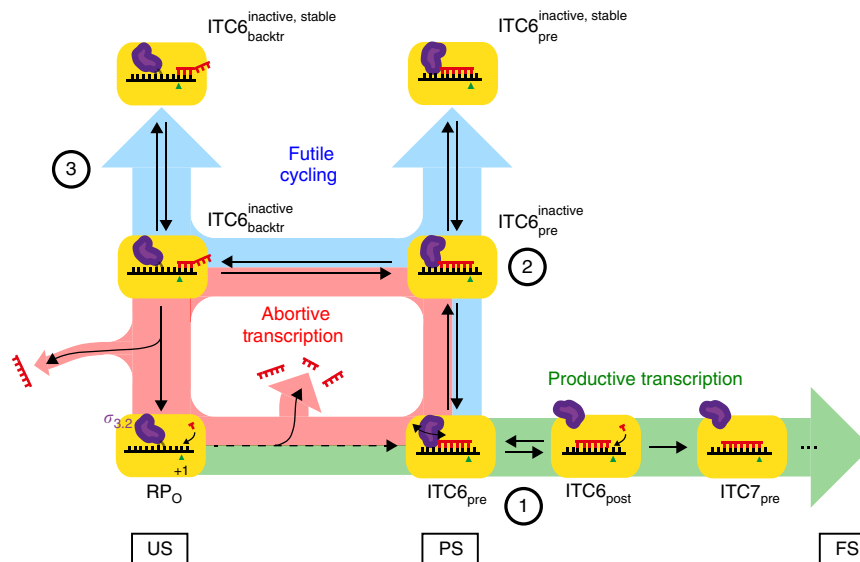
pausing, and pausing-related conformational changes such as scrunching/unscrunching in the presence and absence of RNA release.

The  $\sigma_{3,2}$  region has been described as a good candidate for causing pausing, and structural, biochemical and single-molecule biophysical studies have confirmed that by occluding the RNA-exit channel of RNAP the  $\sigma_{3,2}$  region forms a barrier for the elongation of the nascent transcript past 5–6 nt<sup>12,19,20,27</sup> (Figs. 1b and 6). Consistent with these results, we recently showed that partial deletion of  $\sigma_{3,2}$  significantly diminished pausing at ITC6<sup>27</sup>. However, since the same  $\sigma_{3,2}$  derivative was associated with accelerated conformational dynamics in the open complex<sup>43</sup>, a possibility existed that some of the inferred effects on pause kinetics were indirect (e.g., due to instability of the template strand conformation in the DNA binding cleft leading to increased abortive RNA release and shortening of the ITC6 pause). Our new finding that the triphosphate moiety at the 5'-RNA end, which specifically interacts with the  $\sigma_{3,2}$ <sup>12,20</sup>, both shortens the half-life of ITC6 pause and increases the probability of productive pause exit further suggests that  $\sigma_{3,2}$  is a major pause determinant in initial transcription. We also demonstrated that substitution of the  $\sigma_{3,2}$  residue F522, which contacts template DNA upstream of the active site and represents a barrier for initial RNA synthesis<sup>22,29</sup>, decreases the probability of successful RNA extension beyond the PS state. Thus, competition between the RNA 5'-end and residues from the  $\sigma_{3,2}$  region is required for efficient promoter escape, likely by promoting  $\sigma_{3,2}$  extrusion.

The 80–90% probability to enter the pause may reflect the presence of transcriptionally non-permissive (pausing RPs) and permissive (non-pausing RPs)  $\sigma_{3,2}$  conformations present in different ITC6 complexes. Based on the structural considerations<sup>12,20</sup> and the stage of initial transcription (ref. 27; this study), the clash between  $\sigma_{3,2}$  and RNA 5'-end may hamper the movement of the template DNA and/or RNA to the post-translocated register, and therefore stabilize the pre-translocated state<sup>27,35</sup>. We provide an additional evidence in favor of this hypothesis by showing that  $\beta\text{D446A}$  RNAP (which de-stabilizes the post-translocated<sup>37</sup>) displayed twofold decreased ITC6 pause exit rate. However, because the pause exit rate did not strongly depend on the NTP concentration, the pause is not directly controlled by the thermodynamic equilibrium between the pre- and post-translocated states of ITC6. By similar reasoning, the pause is also not controlled by the catalytic rate of post-translocated ITC6. We thus postulate that the pause-controlling step is kinetic and involves relatively slow repositioning of the  $\sigma_{3,2}$  tip in a way that the barrier to forward translocation is removed. The highest observed pause exit rate on the promoter variant lacking the consensus pause motif ( $0.3\text{ s}^{-1}$  for the  $\Delta\text{P}$  promoter, Fig. 2b) may reflect the rate of ITC6 pre→post translocation that we suggest to

**Fig. 5** Post-RNA synthesis DNA promoter scrunching kinetics are independent of the presence of NTP. **a** Schematic of the RPs possible behavior and the corresponding FRET signal. **b** Schematic of the experiment. **c** Table presenting the experimental variables probed in Fig. 5. Note that the data represented in brown and yellow are from Fig. 2 and are averaged (avg) values for all the concentration of NTP probed in these experiments. Here, we used the WT DNA promoter (Supplementary Fig. 1) and wild-type holoenzyme (Methods). **d** Ratio of the numbers of active FRET pairs, i.e., that display scrunching/unscrunching cycles, over the total number of FRET pairs after selection (Methods: FRET pair localization and detection). The + and - above the bars indicate the presence and the absence, respectively, of NTPs during the experiment. An identical notation is used in **f** and **h**. **e** Experimental FRET trace after NTP removal **b** showing an RP complex alternating between US and PS, but not in the FS FRET level. The experimental conditions used for the acquisition of this trace correspond to the light green color code in **c**. **f** Scrunching kinetics ( $k_1$ ,  $k_2$ , and  $P(k_1)$ ) extracted from a double-exponential MLE fit analysis of the  $\Delta t_{\text{US}}$  and  $\Delta t_{\text{PS}}$  distributions from the traces alike **e**. **g** Experimental FRET trace after NTP removal (as in **b**) showing an RP complex alternating between US, PS and FS FRET level. **h** Scrunching kinetics ( $k_1$ ,  $k_2$ , and  $P(k_1)$ ) extracted from a double-exponential MLE fit analysis of the  $\Delta t_{\text{US}}$ ,  $\Delta t_{\text{PS}}$ , and  $\Delta t_{\text{FS}}$  distributions from the traces alike **g** for the conditions described in **c**. **i** 3D histogram showing the number of transition between two different FRET levels for two consecutive dwell times  $n$  and  $n + 1$  for the traces acquired as described in **c**. **j** Schematic representing the results of **i**; we use wavy arrows to highlight that these are complex transitions involving two timescales; the thicker the arrow, the more likely the transition. **k** Transcript retention experiment using magnetic-bead-attached RP complexes incubated in ITC7 conditions and rinsed (i), then restarted after 15 s by chasing 80  $\mu\text{M}$  GTP (ii) (Supplementary Protocol in Supplementary Methods). The 9mer and 11mer originated from GTP misincorporations





**Fig. 6** Model for initial bacterial transcription. The progress of initial transcription is illustrated by depicting RNAP (yellow block) at key points of the inferred mechanism. The mechanism includes three competing reaction pathways, which the ITC can embark on. Productive Transcription pathway (highlighted in green) results in promoter escape and synthesis of full-length RNA. Abortive Initiation pathway (highlighted in red) leads to the synthesis and dissociation of short RNA products. Futile Cycling (highlighted in blue) temporarily traps the ITC6 into catalytically inactive interconverting pre-translocated and backtracked states, respectively. Purple finger shows the different conformations of  $\sigma_{3,2}$ . Green triangle marks the template base for the next incoming nucleotide in the active site of RNAP. Red and black strands represent the nascent RNA and template DNA, respectively. The multi-timescale transitions between PS and US in Fig. 5j have here been broken up to suggest a simple composition that would give rise to the two timescales observed in Fig. 4b, c. The numeration (1, 2, and 3) indicates the three significant molecular mechanisms described by the model: the initial barrier imposed by  $\sigma_{3,2}$  to the transcript elongation, the subsequent loss of catalytic conformation and the RNA-dependent reversible backtracking, respectively. The US, PS, and FS FRET levels observed during the experiments are indicated at the bottom of the schematic

be controlled by  $\sigma_{3,2}$  repositioning, increasing dramatically the lifetime of the pre-translocated state in comparison to transcription elongation<sup>44</sup> (~265-fold: 2.3 s vs. 8.7 ms). Importantly, several studies of pausing during transcription elongation have shown the predisposition of the pre-translocated RNAP to isomerize into a catalytically inactive off-pathway state, known as the elemental pause<sup>2,37,38,45</sup>. The  $\sigma_{3,2}$ -dependent translocation barrier encountered during initial transcription may thus act, by accumulating the pre-translocated ITC6, to increase the probability to isomerize into an elemental pause-like state (Fig. 6).

We further noticed that the pyrimidine/guanine [(Y)/G] motif, first identified in elongation consensus pauses<sup>37,38</sup>, also affects pausing during transcription initiation (ref. 35 and this study). According to our current data the pause exit rate during initiation (~0.3 s<sup>-1</sup>) is similar to that during elongation (~0.5 s<sup>-1</sup>)<sup>38</sup>. The substitution of the motif increased both the pause exit rate and the probability to exit the pause. In contrast, substitution D446A in the CRE-pocket, which was previously shown to increase consensus pausing<sup>37</sup>, also impaired RNA extension in the ITC6. Overall, it appears that the first events leading to a pause during initiation and elongation phases of transcription are similar: an energetic (transcribed sequence in elongation) or physical ( $\sigma_{3,2}$  in initial transcription) barrier to translocation delays RNAP in the pre-translocated register<sup>35</sup> from where the protein can, with sequence-dependent efficiency, branch-off to a catalytically inactive elemental pause state (Fig. 6).

Although the entry of ITC6 into the elemental pause was nearly obligatory (80–90% of trajectories showed the pause, Supplementary Fig. 2b), a significant fraction (~20% at saturating NTP concentration, Fig. 2d) of the RNAP complexes did not exit this pause on the first attempt, but instead embarked on another reaction pathway involving cyclic unscrunching/scrunching events. Unscrunching mechanistically resembles backtracking and leads to the displacement of 3'-RNA end from the pre-

translocated register (predominant in PS state) towards the NTP-entry channel in US state. The net effect is a long-duration catalytic inactivation of ITC6 (Fig. 4). Interestingly, though promoter scrunching has been associated with intermediate stressed states, we did not observe a different lifetime for US, PS, and FS states (Fig. 5h), suggesting that the transitions connecting the scrunch states are not dominated energetically by the possible intermediate stressed states. Backtracking during initial transcription was also observed using magnetic tweezers<sup>28</sup> (see Supplementary Discussion). The probability to enter the unscrunching/scrunching pathway inversely correlated with NTP concentration (Fig. 2c); at periods of low cellular NTP pool, the unscrunching/scrunching mechanism may thus efficiently inhibit promoter escape and transcript levels. Furthermore, perturbation of RNAP interactions with the DNA template and RNA transcript, e.g., by  $\sigma_{3,2}$ -F522A or  $\beta$ D446A substitutions, favored the partitioning of ITC6 into the unscrunching/scrunching pathway (Fig. 3d and Supplementary Fig. 2g). This finding may further imply that native promoter and initially transcribed sequences encode efficient promoter-escape kinetics because they disfavor ITC6 partitioning into the non-productive unscrunching/scrunching pathway. Consistently, Record and colleagues<sup>46</sup> recently reported the correlation of stronger holoenzyme-discriminator (promoter sequence between the -10 element and transcription start site) interaction with the production of longer abortive RNAs, while producing a higher yield of promoter escape.

Previous single-molecule studies assumed a direct link between unscrunching and abortive transcription<sup>11,17,27,28</sup>. However, those studies focused on the DNA conformation and did not evaluate the retention of RNA in the transcription complexes. Our data demonstrated that brief pulsing of open complexes with NTPs resulted in a population of ITCs that kept on cycling between US/PS/FS states for an extended period of time (Fig. 5),

and that the transcripts could remain stably attached to the ITC for at least 45 min (Supplementary Fig. 4). Backtracking to the US state shortens the template DNA–RNA hybrid to  $\leq 5$  bp, and should reduce the hybrid lifetime. However, the possibility of a short hybrid that locks the transcript into the DNA-binding cleft is supported by the observation of a 4-nt RNA bound to a bacterial open promoter complex in crystals<sup>41</sup>. Furthermore, the backtracked RNA potentially forms interactions in the NTP-entry channel, as observed in the yeast RNAP II<sup>47</sup>. Taken together, the stability of the short hybrid within the complex and the positive interactions between the RNA and the NTP entrance channel may support transcript retention upon promoter unscrunching.

Recent work<sup>46</sup> has also noted that RP<sub>O</sub> complexes on  $\lambda$ P<sub>R</sub> and T7A1 promoters were divided into two populations upon NTP addition: a first population (30–45% of all complexes) that rapidly (within 10 sec) synthesized long RNA (ITC > 10) and a second population that was stalled in early transcription (ITC < 10), and that released RNA slowly, similarly to moribund complexes<sup>48</sup>. We propose that these two populations, i.e., the population producing quickly long RNAs and the moribund complexes, are consistent with the two populations we described here, i.e., the RP complexes that exited the ITC6 pause on the first attempt (Fig. 1c), and the population that entered the cyclic unscrunching/scrunching state from the ITC6 pause (Fig. 1d), respectively (Supplementary Discussion).

Our new findings are summarized in a kinetic model of the transition to productive transcription (Fig. 6 and Supplementary Discussion).

The dependence of entry and recovery from the pause states implies wide variation in the kinetics of initial transcription across the bacterial promoter sequence space. On the consensus *lac* and similar promoters, the molecular mechanism of pausing sensitizes the efficiency of promoter escape to NTP concentration, potentially trapping the RNAP to the promoter in a “ready-to-fire” or “poised” mode until improved growth conditions lead to the replenishing of cellular NTP pool<sup>49–51</sup>. The trapping of poised RNAPs at or near the promoter thus emerges as a common transcription regulation strategy achievable by different molecular mechanisms. For example, the  $\sigma^{54}$ -RNAP holoenzyme forms an inactive, stable closed complex in bacteria<sup>52</sup>, whereas negative elongation factors cause RNAP to stall within 20–60 bases downstream of the transcription start site in many metazoan genes<sup>53</sup>. In all cases, inhibited RNAPs are ready-to-fire when activating signals arrive from relevant signal-transduction cascades thus avoiding promoter search, binding, melting, and activation.

## Methods

**Glass coverslips preparation for single-molecule experiments.** Borosilicate glass coverslips (1.5 MenzelGläser, Germany) were sonicated for 30 min in a 2% (V/V) solution of Hellmanex III (Helma Analytics, Germany)/deionized water. After being thoroughly rinsed with deionized water, the coverslips were dried, disposed into a plasma cleaner (Harrick Plasma, NY, USA), and exposed to a nitrogen plasma for 30 min. The coverslips are subsequently immersed into a 1% (V/V) solution of Vectabond (product code SP-1800, Vector Labs, CA, USA)/acetone for 10 min. The coverslips were then rinsed in deionized water and dried with a stream of nitrogen gas. After disposing a silicone gasket (103280, Grace Bio-Labs, OR, USA) on each coverslip, each well was filled with 20  $\mu$ l of a 100 mg/ml solution of methoxy-PEG (5 kDa)-SVA/ biotin-PEG (5 kDa)-SC (2.5% (w/w) (Laysan Bio, AL, USA) in 50 mM MOPS-KOH buffer, pH 7.5. The wells were incubated for ~1.5 h and thoroughly rinsed with a 1  $\times$  phosphate-buffered saline (PBS; Sigma Aldrich, UK) solution. The coverslips were stored at 4 °C up to 2 weeks before use.

**Protein immobilization protocol.** The pegylated coverslips were incubated for ~10 min with a solution 0.5 mg/ml of Neutravidin (31050, Sigma Aldrich) in 0.5  $\times$  PBS and subsequently rinsed with 1  $\times$  PBS. Preceding observation on the microscope, the coverslips were incubated for ten min with a 3% (V/V) solution of Penta-His biotin conjugate antibody (34440, QIAGEN, UK) in reaction buffer (40

mM HEPES buffer pH 7.3 (ThermoFisher Scientific, UK), 100 mM potassium glutamate, 10 mM MgCl<sub>2</sub>, 1 mM dithiothreitol (DTT), 1 mM cysteamine hydrochloride, 5% glycerol (V/V), 0.5 g/l bovine serum albumin) and subsequently rinsed with reaction buffer. After adjusting the coverslips on the microscope, 100  $\mu$ M of His-tagged protein–DNA complex, e.g., RP<sub>O</sub>, was incubated in the observed well until the desired density of molecules on the coverslips surface was reached, followed by one-step rinsing with reaction buffer.

**Core RNAP and  $\sigma^{70}$  preparation.** The expression and purification of the core bacterial RNAP have previously been described in ref. 54. The expression and purification of the WT  $\sigma^{70}$  have been previously described in ref. 21.

**Holo-RNAP and RP<sub>O</sub> assembly.** Core RNAP (0.5  $\mu$ M) was mixed with 0.6  $\mu$ M of  $\sigma^{70}$  in 20 mM Tris-HCl pH 7.9, 150 mM NaCl, 0.1 mM EDTA, 50% (V/V) Glycerol, and 0.1 mM DTT, and incubated at 30 °C for 30 min. The resulting holo-complex was stored at ~20 °C.

Holo-complex (5 nM) was mixed with 2.5 nM of DNA promoter in reaction buffer and incubated for 10 min at 37 °C to form the RP<sub>O</sub> complex.

**DNA constructs preparation.** The DNA constructs preparation is described in detail in the Supplementary Methods.

**Microscope and single-molecule experiments.** The single-molecule TIRF microscope for FRET experiments has been previously described in ref. 55. Shortly, the 532 nm and the 642 nm wavelength laser beams (donor laser excitation and acceptor laser excitation, respectively) were focused in the back focal plane of an oil immersion objective (Olympus, N.A. 1.4) and illuminate alternatively the field of view, i.e., ALEX mode<sup>56</sup>. The TIRF-reflected beams were directed toward a position sensor to control the objective focal plane distance to the sample at a fixed position (MS-2000 stage, ASI, OR, USA). The photons resulting from the de-excitation of the dyes molecules, i.e., fluorescence, were separated from the excitation laser beams with a dichroic mirror and spectrally splitted in two channels, e.g., donor and acceptor that are imaged on the same electron-modifying charge-coupled device camera (iXon, Andor, Ireland). For 100 ms ALEX illumination, i.e., 200 ms frame time acquisition, the laser power measured preceding the dichroic mirror is ~0.4 mW for donor excitation laser and ~0.09 mW for the acceptor excitation laser. For 40 ms ALEX illumination (only used to acquire the data with the  $\Delta$ P promoter and ATP starting substrate experimental condition), i.e., 80 ms frame time, the laser power measured preceding the dichroic mirror is ~1 mW for donor excitation laser and ~0.25 mW for the acceptor excitation laser.

The imaging buffer contained the reaction buffer completed with 1 mM Trolox, 1 mM COT, 1% (w/v) glucose, 0.4  $\mu$ g/ml of catalase, and 1 mg/ml of glucose oxidase (Sigma Aldrich). The catalase and the glucose oxidase were pre-mixed together in a solution of 50 mM KCl and 50 mM Tris-OAc buffer pH 7.3 at 100  $\times$  concentration<sup>57,58</sup>.

The data were acquired after immobilization of the RP<sub>O</sub> complex to the surface. After ~200 frames (~20 s), the imaging buffer is spiked with a 12.5  $\times$  NTP solution and the reaction is observed for the remaining ~5800 frames (total time).

For the post-RNA synthesis rinsing experiments, the RP<sub>O</sub> was incubated with NTPs in the reaction buffer for 10 s before the reaction buffer was exchanged twice and finally replaced with imaging buffer, followed by the start of the acquisition. The buffer exchange procedure takes ~40 sec to be completed before the start of the acquisition.

All single-molecule FRET experiments were performed at 22 °C.

**Single-molecule data analysis.** FRET pair localization and detection: The movies recorded on the camera were offline analyzed using the home-built Matlab routine Twotone-ALEX<sup>59</sup> to extract the intensities of co-localized donor and acceptor, i.e., FRET-pair. The following parameters from Twotone-ALEX were used to select only the FRET pairs formed by a single ATTO647N acceptor dye and a single Cy3b donor dye: channel filter as DexDem&&AexAem&&DexAem (colocalization of the donor dye signal upon donor laser excitation, the acceptor dye signal upon acceptor laser excitation, and the acceptor dye signal upon donor laser excitation), a width limit between the donor and the acceptor between 1 and 2 pixels, a nearest-neighbor limit of 6 pixels, and a maximal ellipticity of 0.6 (ellipticity is defined as the ratio of the minor and the major axis of the ellipse). The traces extracted from the Twotone-ALEX analysis were then sorted to remove all the traces that displayed extensive blinking or multisteps photobleaching, i.e., that contain more than one donor or acceptor dye in the same diffraction limited intensity spot.

**Calibration of the FRET sensor.** We calibrated the FRET sensor by measuring  $E_{\text{FRET}}$  for initial transcribing complexes in the presence of different subsets of nucleotides that allowed maximal transcript lengths of 6, 7, or 11 nucleotides, and compared their FRET profiles with RP<sub>O</sub>. These experiments allowed unambiguous assignment of the RP<sub>O</sub> and ITC11 FRET levels ( $E_{\text{FRET}}$ ) as ~0.5 and ~0.76, respectively (Fig. 1c; see also Supplementary Fig. 1). For clarity, we define these FRET states as “US” and “FS”, respectively. Although it is commonly accepted that RNAP escapes *lac* promoters after synthesizing an 11-mer<sup>9,60</sup>, the high FRET signal

after forming a 11-nt long transcript (Fig. 1c) is consistent with extended scrunching of the downstream DNA in an ITC (and not an elongation complex). On the other hand, complexes with maximal transcript lengths of 6 and 7 nucleotides yielded  $E_{\text{FRET}}$  0.38 and 0.37, respectively. We observed here a decrease in  $E_{\text{FRET}}$ , i.e., an increase in the inter-dye distance as a consequence of further transcript extension as the downstream dye at some positions moves farther along from the donor due to the rotation of the downstream dsDNA<sup>27,61</sup>. The assignment of this  $E_{\text{FRET}}$  level to an ITC6 pause state, or “PS,” was based on our previous observations showing that ITC6 (but not ITC7) accumulates in significant amounts in conditions that allow continuous transcription past the 6–7-mer RNA and that ITCs do not pause again until reaching the position +12 of the DNA template used here<sup>27</sup>.

**FRET efficiency and hidden Markov modeling.** The FRET efficiency dynamics for each FRET-pair was calculated with the standard formula  $E = \frac{I_{\text{DA}}}{I_{\text{DA}} + I_{\text{DD}}}$ , with  $I_{\text{DA}}$  and  $I_{\text{DD}}$  being respectively the intensity of the acceptor and of the donor upon donor excitation<sup>62</sup>. The traces were analyzed through a modified version of the hidden Markov model ebFRET software from ref. <sup>63</sup> (the modified code is available from the corresponding authors on reasonable request), where only steps longer than two frames and separated from the subsequent step by more than twice the Allan deviation estimated at five frames were conserved<sup>64</sup> to be assembled into dwell time. The first dwell time, i.e., preceding NTP addition, and the last dwell time, i.e., preceding photobleaching or transition to FS FRET state, of each trace were removed from the dwell time distribution.

**Characterization of the dwell time distributions.** A detailed analysis of the dwell time distributions is provided in ref. <sup>65</sup>. Shortly, the distribution of  $\tau$  are described by a probability distribution function with  $m$  exponentials:

$$p_i(\tau) = \sum_{n=1}^m p_n \cdot k_n \cdot e^{-k_n \tau}, \quad (1)$$

where  $k_n$  and  $p_n$  are the characteristic rate of the  $m^{\text{th}}$  exponential and its probability, respectively. The minimum number of exponential to fit the distributions was determined for each distribution by using the Bayes Schwarz Information Criterion<sup>66</sup>. We calculate the maximum likelihood estimate of the parameters (MLE)<sup>67</sup> by maximizing:

$$L = \sum_{i=1}^N \ln(p_i(\tau_i)) \quad (2)$$

over the parameter set. Here, the  $\tau_i$  are the experimentally measured dwell times and  $N$  is the number of collected dwell times  $\tau_i$ . The error bars for each fitting parameters are one standard deviation extracted from 1000 bootstrap procedures<sup>68</sup>. The ebFRET software<sup>63</sup> was also used to extract the peak positions of each FRET level, subsequently fitted with a Gaussian function, with the peak center and the SD as free parameters (Supplementary Figs. 1d and 3a-d).

**Data availability.** The data sets generated and analyzed during the current study are available from the corresponding author on reasonable request.

Received: 17 June 2017 Accepted: 20 March 2018

Published online: 16 April 2018

## References

- Browning, D. F. & Busby, S. J. Local and global regulation of transcription initiation in bacteria. *Nat. Rev. Microbiol.* **14**, 638–650 (2016).
- Belogurov, G. A. & Artsimovitch, I. Regulation of transcript elongation. *Annu Rev. Microbiol.* **69**, 49–69 (2015).
- Roberts, J. W., Shankar, S. & Filter, J. J. RNA polymerase elongation factors. *Annu Rev. Microbiol.* **62**, 211–233 (2008).
- Washburn, R. S. & Gottesman, M. E. Regulation of transcription elongation and termination. *Biomolecules* **5**, 1063–1078 (2015).
- Ray-Soni, A., Bellecourt, M. J. & Landick, R. Mechanisms of bacterial transcription termination: all good things must end. *Annu. Rev. Biochem.* **85**, 319–347 (2016).
- Browning, D. F. & Busby, S. J. The regulation of bacterial transcription initiation. *Nat. Rev. Microbiol.* **2**, 57–65 (2004).
- Saecker, R. M., Record, M. T. Jr & Dehaseth, P. L. Mechanism of bacterial transcription initiation: RNA polymerase-promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis. *J. Mol. Biol.* **412**, 754–771 (2011).
- Ruff, E. F., Record, M. T. Jr & Artsimovitch, I. Initial events in bacterial transcription initiation. *Biomolecules* **5**, 1035–1062 (2015).
- Murakami, K. S. & Darst, S. A. Bacterial RNA polymerases: the whole story. *Curr. Opin. Struct. Biol.* **13**, 31–39 (2003).
- Kapanidis, A. N. et al. Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science* **314**, 1144–1147 (2006).
- Revyakin, A., Liu, C., Ebright, R. H. & Strick, T. R. Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science* **314**, 1139–1143 (2006).
- Zuo, Y. & Steitz, T. A. Crystal structures of the *E. coli* transcription initiation complexes with a complete bubble. *Mol. Cell* **58**, 534–540 (2015).
- Winkelman, J. T. et al. Crosslink mapping at amino acid-base resolution reveals the path of scrunched DNA in initial transcribing complexes. *Mol. Cell* **59**, 768–780 (2015).
- Carpousis, A. J. & Gralla, J. D. Cycling of ribonucleic acid polymerase to produce oligonucleotides during initiation in vitro at the lac UV5 promoter. *Biochemistry* **19**, 3245–3253 (1980).
- Gralla, J. D., Carpousis, A. J. & Stefano, J. E. Productive and abortive initiation of transcription in vitro at the lac UV5 promoter. *Biochemistry* **19**, 5864–5869 (1980).
- Hsu, L. M. Monitoring abortive initiation. *Methods* **47**, 25–36 (2009).
- Margeat, E. et al. Direct observation of abortive initiation and promoter escape within single immobilized transcription complexes. *Biophys. J.* **90**, 1419–1431 (2006).
- Goldman, S. R., Ebright, R. H. & Nickels, B. E. Direct detection of abortive RNA transcripts in vivo. *Science* **324**, 927–928 (2009).
- Samanta, S. & Martin, C. T. Insights into the mechanism of initial transcription in *Escherichia coli* RNA polymerase. *J. Biol. Chem.* **288**, 31993–32003 (2013).
- Basu, R. S. et al. Structural basis of transcription initiation by bacterial RNA polymerase holoenzyme. *J. Biol. Chem.* **289**, 24549–24559 (2014).
- Kulbachinskiy, A. & Mustaev, A. Region 3.2 of the sigma subunit contributes to the binding of the 3'-initiating nucleotide in the RNA polymerase active center and facilitates promoter clearance during initiation. *J. Biol. Chem.* **281**, 18273–18276 (2006).
- Pupov, D., Kuzin, I., Bass, I. & Kulbachinskiy, A. Distinct functions of the RNA polymerase sigma subunit region 3.2 in RNA priming and promoter escape. *Nucleic Acids Res.* **42**, 4494–4504 (2014).
- Hsu, L. M. Promoter clearance and escape in prokaryotes. *Biochim. Et. Biophys. Acta* **1577**, 191–207 (2002).
- Vo, N. V., Hsu, L. M., Kane, C. M. & Chamberlin, M. J. In vitro studies of transcript initiation by *Escherichia coli* RNA polymerase. 3. Influences of individual DNA elements within the promoter recognition region on abortive initiation and promoter escape. *Biochemistry* **42**, 3798–3811 (2003).
- Carpousis, A. J., Stefano, J. E. & Gralla, J. D. 5' nucleotide heterogeneity and altered initiation of transcription at mutant lac promoters. *J. Mol. Biol.* **157**, 619–633 (1982).
- Stefano, J. E. & Gralla, J. Lac UV5 transcription in vitro. Rate limitation subsequent to formation of an RNA polymerase-DNA complex. *Biochemistry* **18**, 1063–1067 (1979).
- Duchi, D. et al. RNA polymerase pausing during initial transcription. *Mol. Cell* **63**, 939–950 (2016).
- Lerner, E. et al. Backtracked and paused transcription initiation intermediate of *Escherichia coli* RNA polymerase. *Proc. Natl Acad. Sci. USA* **113**, E6562–E6571 (2016).
- Zhang, Y. et al. Structural basis of transcription initiation. *Science* **338**, 1076–1080 (2012).
- Liu, B., Zuo, Y. & Steitz, T. A. Structures of *E. coli* sigmaS-transcription initiation complexes provide new insights into polymerase mechanism. *Proc. Natl Acad. Sci. USA* **113**, 4051–4056 (2016).
- Xie, S. N. Single-molecule approach to enzymology. *Single Mol.* **2**, 229–236 (2001).
- Tinoco, I. Jr. & Gonzalez, R. L. Jr. Biological mechanisms, one molecule at a time. *Genes Dev.* **25**, 1205–1231 (2011).
- Dulin, D., Lipfert, J., Moolman, M. C. & Dekker, N. H. Studying genomic processes at the single-molecule level: introducing the tools and applications. *Nat. Rev. Genet.* **14**, 9–22 (2013).
- Kapanidis, A. N. & Strick, T. Biology, one molecule at a time. *Trends Biochem. Sci.* **34**, 234–243 (2009).
- Bauer, D. L. V., Duchi, D. & Kapanidis, A. N. *E. coli* RNA polymerase pauses during initial transcription. *Biophys. J.* **110**, 21a–21a (2016).
- Ross, W. & Gourse, R. L. Sequence-independent upstream DNA-alphaCTD interactions strongly stimulate *Escherichia coli* RNA polymerase-lacUV5 promoter association. *Proc. Natl Acad. Sci. USA* **102**, 291–296 (2005).
- Vvedenskaya, I. O. et al. Interactions between RNA polymerase and the “core recognition element” counteract pausing. *Science* **344**, 1285–1289 (2014).
- Larson, M. H. et al. A pause sequence enriched at translation start sites drives transcription dynamics in vivo. *Science* **344**, 1042–1047 (2014).



39. Imashimizu, M. et al. Visualizing translocation dynamics and nascent transcript errors in paused RNA polymerases in vivo. *Genome Biol.* **16**, 98 (2015).
40. Vvedenskaya, I. O. et al. Interactions between RNA polymerase and the core recognition element are a determinant of transcription start site selection. *Proc. Natl Acad. Sci. USA* **113**, E2899–E2905 (2016).
41. Bae, B., Feklistov, A., Lass-Napiorkowska, A., Landick, R. & Darst, S. A. Structure of a bacterial RNA polymerase holoenzyme open promoter complex. *eLife*. **4**, 2015.
42. Feklistov, A. & Darst, S. A. Structural basis for promoter-10 element recognition by the bacterial RNA polymerase sigma subunit. *Cell* **147**, 1257–1269 (2011).
43. Duchi, D. et al. Conformational heterogeneity and bubble dynamics in single bacterial transcription initiation complexes. *Nucleic Acids Res.* **46**, 677–688 (2017).
44. Malinen, A. M. et al. Active site opening and closure control translocation of multisubunit RNA polymerase. *Nucleic Acids Res.* **40**, 7442–7451 (2012).
45. Landick, R. The regulatory roles and mechanism of transcriptional pausing. *Biochem Soc. Trans.* **34**, 1062–1066 (2006).
46. Henderson, K. L. et al. Mechanism of transcription initiation and promoter escape by *E. coli* RNA polymerase. *Proc. Natl Acad. Sci. USA* **114**, E3032–E3040 (2017).
47. Cheung, A. C. & Cramer, P. Structural basis of RNA polymerase II backtracking, arrest and reactivation. *Nature* **471**, 249–253 (2011).
48. Kubori, T. & Shimamoto, N. A branched pathway in the early stage of transcription by *Escherichia coli* RNA polymerase. *J. Mol. Biol.* **256**, 449–457 (1996).
49. Belenky, P. et al. Bactericidal antibiotics induce toxic metabolic perturbations that lead to cellular damage. *Cell Rep.* **13**, 968–980 (2015).
50. Buckstein, M. H., He, J. & Rubin, H. Characterization of nucleotide pools as a function of physiological state in *Escherichia coli*. *J. Bacteriol.* **190**, 718–726 (2008).
51. Turnbough, C. L. Jr. Regulation of bacterial gene expression by the NTP substrates of transcription initiation. *Mol. Microbiol.* **69**, 10–14 (2008).
52. Zhang, N., Darbari, V. C., Glyde, R., Zhang, X. & Buck, M. The bacterial enhancer-dependent RNA polymerase. *Biochem. J.* **473**, 3741–3753 (2016).
53. Adelman, K. & Lis, J. T. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat. Rev. Genet.* **13**, 720–731 (2012).
54. Svetlov, V. & Artsimovitch, I. Purification of bacterial RNA polymerase: tools and protocols. *Methods Mol. Biol.* **1276**, 13–29 (2015).
55. Holden, S. J. et al. Defining the limits of single-molecule FRET resolution in TIRF microscopy. *Biophys. J.* **99**, 3102–3111 (2010).
56. Kapanidis, A. N. et al. Fluorescence-aided molecule sorting: analysis of structure and interactions by alternating-laser excitation of single molecules. *Proc. Natl Acad. Sci. USA* **101**, 8936–8941 (2004).
57. Aitken, C. E., Marshall, R. A. & Puglisi, J. D. An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments. *Biophys. J.* **94**, 1826–1835 (2008).
58. Dave, R., Terry, D. S., Munro, J. B. & Blanchard, S. C. Mitigating unwanted photophysical processes for improved single-molecule fluorescence imaging. *Biophys. J.* **96**, 2371–2381 (2009).
59. Holden, S. J., Uphoff, S. & Kapanidis, A. N. DAOSTORM: an algorithm for high-density super-resolution microscopy. *Nat. Methods* **8**, 279–280 (2011).
60. Mukhopadhyay, J. et al. Translocation of sigma(70) with RNA polymerase during transcription: fluorescence resonance energy transfer assay for movement relative to DNA. *Cell* **106**, 453–463 (2001).
61. Robb, N. C. et al. The transcription bubble of the RNA polymerase-promoter open complex exhibits conformational heterogeneity and millisecond-scale dynamics: implications for transcription start-site selection. *J. Mol. Biol.* **425**, 875–885 (2013).
62. Ha, T. et al. Probing the interaction between two single molecules: fluorescence resonance energy transfer between a single donor and a single acceptor. *Proc. Natl Acad. Sci. USA* **93**, 6264–6268 (1996).
63. van de Meent, J. W., Bronson, J. E., Wiggins, C. H. & Gonzalez, R. L. Jr. Empirical Bayes methods enable advanced population-level analyses of single-molecule FRET experiments. *Biophys. J.* **106**, 1327–1337 (2014).
64. Dulin, D. et al. High spatiotemporal-resolution magnetic tweezers: calibration and applications for DNA dynamics. *Biophys. J.* **109**, 2113–2125 (2015).
65. Dulin, D., Berghuis, B. A., Depken, M. & Dekker, N. H. Untangling reaction pathways through modern approaches to high-throughput single-molecule force-spectroscopy experiments. *Curr. Opin. Struct. Biol.* **34**, 116–122 (2015).
66. Schwarz, G. Estimating dimension of a model. *Ann. Stat.* **6**, 461–464 (1978).
67. Cowan, G. *Statistical Data Analysis*. (Oxford University Press, Gret Clarendon Street, Oxford, 1998).
68. Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. *Numerical Recipes in C: The Art of Scientific Computing*. (Cambridge Univ. Press, New York, NY, USA, 1992).
69. James, K., Gamba, P., Cockell, S. J. & Zenkin, N. Misincorporation by RNA polymerase is a major source of transcription pausing in vivo. *Nucleic Acids Res.* **45**, 1105–1113 (2017).

## Acknowledgements

We thank Dr. Francesco Pedaci, Dr. Mohamed Fareh, and Tao Ju Cui for carefully reading the manuscript. We also thank Dr. Jan Lipfert for discussion and Dr. Jun Fan for experimental support during the revision process. This work was supported by grants to ANK from the European Research Council (261227), the Wellcome Trust (110164/Z/15/Z), and the UK BBSRC (BB/H01795X/1 and BB/J00054X/1). D.L.V.B. was supported by an US National Science Foundation grant (1309306) and by an EPA Cephalosporin Junior Research Fellowship at Linacre College, Oxford. D.D. was supported by the Interdisciplinary Center for Clinical Research (IZKF) at the University Hospital of the University of Erlangen-Nuremberg. A.M.M. was supported by the Instrumentarium Science Foundation, Finnish Cultural Foundation and Alfred Kordelin Foundation. I.P. and A.K. were supported by the Russian Foundation for Basic Research grant 17-54-150009.

## Author contributions

D.D. and A.N.K. designed the research. D.D. designed, performed, and analyzed the single-molecule experiments. M.K. made the DNA promoter constructs. D.D. and J.J.W.B. designed and implemented the custom analysis routines. D.L.V.B. and I.P. performed and analyzed the gel electrophoresis experiments. Z.M. and K.B. provided the  $\sigma^{70}$  mutant. D.D., A.M.M., A.K., and A.N.K. discussed the data. D.D., A.M.M., and M.D. discussed the model. D.D., A.M.M., and A.N.K. wrote the article.

## Additional information

**Supplementary Information** accompanies this paper at <https://doi.org/10.1038/s41467-018-03902-9>.

**Competing interests:** The authors declare no competing interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018