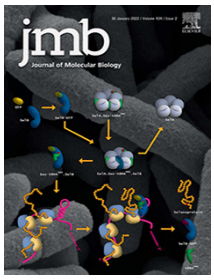

Real-Time Single-Molecule Studies of RNA Polymerase–Promoter Open Complex Formation Reveal Substantial Heterogeneity Along the Promoter-Opening Pathway

Anssi M. Malinen [1,2*], Jacob Bakermans [2], Emil Aalto-Setälä [1], Martin Blessing [2,3], David L. V. Bauer [2,4], Olena Parilova [1], Georgiy A. Belogurov [1], David Dulin [2,5,6] and Achillefs N. Kapanidis [2,7*]

*1 - **Department of Life Technologies,** University of Turku, 20014 Turku, Finland*
*2 - **Biological Physics Research Group,** Clarendon Laboratory, Department of Physics, University of Oxford, Parks Road, Oxford OX1 3PU, UK*
*3 - **Max Planck Institute for the Science of Light,** Staudtstraße 2, 91058 Erlangen, Germany*
*4 - **RNA Virus Replication Laboratory,** The Francis Crick Institute, 1 Midland Road, London NW1 1AT, UK*
*5 - **Junior Research Group 2,** Interdisciplinary Center for Clinical Research, Friedrich-Alexander-University Erlangen-Nürnberg (FAU), Cauerstr. 3, 91058 Erlangen, Germany*
*6 - **Department of Physics and Astronomy,** and LaserLaB Amsterdam, Vrije Universiteit Amsterdam, De Boelelaan 1081, 1081 HV Amsterdam, the Netherlands*
*7 - **Kavli Institute for Nanoscience Discovery,** University of Oxford, Oxford*

Correspondence to Anssi M. Malinen and Achillefs N. Kapanidis: Department of Life Technologies, University of Turku, 20014 Turku, Finland (A.M. Malinen); Clarendon Laboratory, Department of Physics, University of Oxford, Parks Road, Oxford OX1 3PU, UK (A.N. Kapanidis). *anssi.malinen@utu.fi (A.M. Malinen), achillefs.kapanidis@physics.ox.ac.uk (A.N. Kapanidis)*
https://doi.org/10.1016/j.jmb.2021.167383
Edited by Edward Lemke

Abstract

The expression of most bacterial genes commences with the binding of RNA polymerase (RNAP)–σ^{70} holoenzyme to the promoter DNA. This initial RNAP–promoter closed complex undergoes a series of conformational changes, including the formation of a transcription bubble on the promoter and the loading of template DNA strand into the RNAP active site; these changes lead to the catalytically active open complex (RP_O) state. Recent cryo-electron microscopy studies have provided detailed structural insight on the RP_O and putative intermediates on its formation pathway. Here, we employ single-molecule fluorescence microscopy to interrogate the conformational dynamics and reaction kinetics during real-time RP_O formation on a consensus *lac* promoter. We find that the promoter opening may proceed rapidly from the closed to open conformation in a single apparent step, or may instead involve a significant intermediate between these states. The formed RP_O complexes are also different with respect to their transcription bubble stability. The RNAP cleft loops, and especially the β′ rudder, stabilise the transcription bubble. The RNAP interactions with the promoter upstream sequence (beyond −35) stimulate transcription bubble nucleation and tune the reaction path towards stable forms of the RP_O.

Introduction

Transcription initiation is the first and most regulated step in gene expression in all organisms. The expression of most bacterial genes commences with the binding of RNA polymerase (RNAP)–σ^{70} holoenzyme to the promoter DNA.[1] The initial RNAP–promoter closed complex (RP_C) undergoes large conformational changes leading to a RNAP–promoter open complex (RP_O), which is capable of RNA synthesis. These conformational changes are of paramount importance, since their modulation by promoter DNA sequence, protein transcription factors, and small-molecule ligands strongly affects the number of active open complex, and thus the transcription efficiency.[2] Further, several antimicrobials, including clinically used drugs rifampicin[3,4] and fidaxomicin,[5] exert their effect by blocking RNAP from proceeding during a specific step of transcription initiation.[6] However, despite substantial progress in defining the structural basis of transcription initiation mechanism,[7–10] the identity, sequence, and kinetics of conformational changes leading to RP_O formation remain elusive.

At the initial step of the RP_O formation pathway, the RNAP–σ^{70} holoenzyme recognises the promoter by forming sequence-specific contacts with the −35 element, and sequence-independent contacts upstream from the −35 ("upstream sequence") as well as around the −10 element [reviewed in 2,11] In this RP_C state, the promoter remains fully double-stranded, but is bent by ~17° at the −10 element, thus positioning the downstream promoter DNA above the DNA-binding cleft of the RNAP.[10] Studies using footprinting,[12–15] atomic force microscopy[13] and ensemble FRET[17] have indicated additional extensive bending and wrapping of the promoter upstream sequence (between the −35 element and −82); this bending, which is driven by the C-terminal domains of the two RNAP α-subunits (αCTDs) interacting with the promoter upstream, brings the upstream DNA to the RNAP surface, and strongly facilitates RP_O complex formation.[14,18,19]

The isomerisation of the RP_C towards RP_O complex begins with the flipping of non-template DNA (ntDNA) −11 conserved adenine base from the duplex DNA to a specific pocket in σ^{70}.[20,21] The promoter melting then somehow propagates downstream until the full transcription bubble in the RP_O complex covers positions −11 to +2.[9] The bubble melting is coupled with the loading of downstream DNA duplex into the RNAP cleft, and the loading of single-stranded template DNA (tDNA) into the RNAP active site. Structural[9,10] and biochemical[2] studies have identified several putative intermediates on the path from the RP_C to RP_O; however, the number and structural properties of the intermediates detected appear to heavily depend on the promoter sequence, transcription factors, and experimental conditions.

The mechanism discussed above describes the formation of a uniform RP_O complex on a standard linear reaction pathway. A more complete description of the transcription initiation, however, needs to consider several studies that suggested that individual RP_O molecules are not identical, and they instead differ in functional properties.[22–25] One of the most notable variation among RP_O complexes is their tendency to perform abortive initiation, i.e., the premature release of short RNAs synthesised by promoter-bound RNAP (reviewed in 26). In fact, it has been estimated that >50% of the RP_O complexes are permanently locked into the abortive initiation mode and cannot produce full-length RNA.[22–25] The presence of at least two different RP_O classes – one productive and one non-productive (abortive) – raises the possibility that the RP_O pathway is also not linear, but instead branches to allow the formation of structurally and functionally different RP_O molecules. It has been further suggested that the ratio of productive and non-productive RP_O complexes can be modulated by transcription factors and thus offers a layer for gene regulation in the cell.[27] On the other hand, recent single-molecule studies revealed long-lived pausing, backtracking and arrest of initially transcribing bacterial and mitochondrial RNAPs that could potentially explain the productive and abortive RNA synthesis by a single type of RP_O complexes.[28–30] The RP_O formation pathway branching – its occurrence and mechanism – thus warrants further study.

Here, we utilise single-molecule techniques to resolve asynchronous, multi-step and branched reaction mechanisms during σ^{70}-dependent RP_O formation on a well characterised consensus *lac* promoter. Our results strongly suggest that the RP_O formation pathway is indeed branched both at the step of initial promoter melting and the step of open transcription bubble stabilisation. Furthermore, αCTD interactions with the promoter upstream sequence strongly stimulate bubble initiation and tune the reaction pathway towards more stable RP_O complexes. The RNAP cleft loops (and especially the β′ rudder one), play a key role in stabilising the open transcription bubble.

Results

Direct formation of surface-immobilised catalytically active open complexes

To be able to monitor RNAP–promoter open complex (RP_O) formation in real-time at the single-molecule level, we used FRET to look at the changes in distances between two points, i.e.,

positions −15 and +15 relative to the transcription start site (position +1) on a promoter DNA fragment. A fluorophore pair incorporated in positions −15 (donor) and +15 (acceptor) produces FRET signatures that vary depending on the transcription bubble conformation; this pair has been employed before to monitor conformational changes in populations of single transcription complexes,[31,32] conformational dynamics of RP_O complexes[33] and conformational changes after the formation of RP_O complex[29] on a consensus *lac* promoter (*lacCONS*) (Figure S1(A, B)).

Here, we modified our previous protocols to detect the nascent RNAP–promoter complex

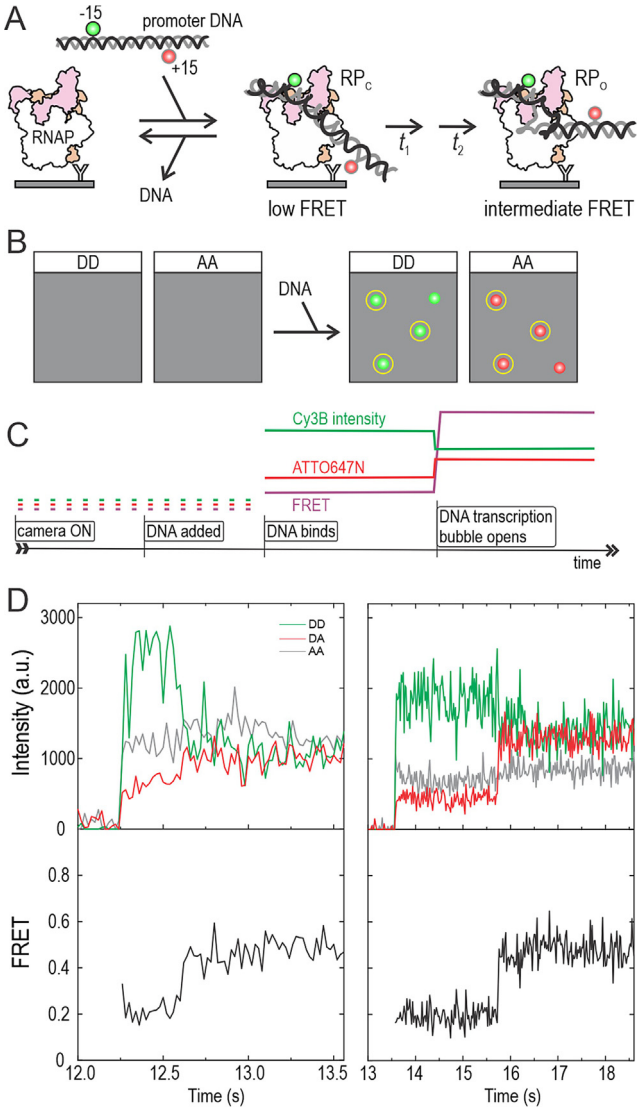

Figure 1. Single-molecule FRET method to monitor the RNAP–promoter open complex formation in real-time. (A) *E.* coli RNAP–σ^{70} holoenzyme is immobilised on the PEGylated microscope coverslip using biotinylated anti-His-tag-antibody. *lacCONS* promoter, which is labelled with a donor fluorophore (D, Cy3B) at non-template DNA position −15 and an acceptor fluorophore (A, ATTO647N) at template DNA position + 15, is added to the reaction buffer. The promoter binds to the RNAP and becomes visible on the coverslip surface. The initial RNAP–promoter closed complex isomerises to the open complex, which decreases the distance between the −15 and +15 dyes and increases the FRET. (B) Schematic microscope field-of-view before and after promoter addition to the reaction buffer. Data on the DD (donor excitation–donor emission) and AA (acceptor excitation–acceptor emission) channels is used to identify the RNAP–promoter complexes containing both the Cy3B and ATTO647N dyes. These molecules are highlighted with yellow circles. (C) Schematic single-molecule trajectory showing promoter binding to the RNAP and subsequent isomerisation to the open complex state. Abrupt increase in the Cy3B and ATTO647N fluorescence intensities defines the moment of promoter binding. The increase in the FRET from the low level to the intermediate level defines the moment of open complex formation. (D) Two experimental trajectories show promoter binding and open complex formation. DA indicates the signal on the donor excitation–acceptor emission channel.

(RP_C) and its subsequent maturation to RP_O (Figure 1). To this end, we attached molecules of the *Escherichia coli* RNA polymerase–σ^{70} holoenzyme to the surface of a coverslip and started imaging the surface using TIRF microscopy (Figure 1(A, B)). Subsequent addition of the dual-labelled promoter DNA to the reaction buffer was expected to lead to the appearance of co-localised fluorescent spots on the donor (Cy3B label) and acceptor (ATTO647N label) detection channels of the microscope upon binding to the surface-attached holoenzyme (Figure 1(B)). The timing of the appearance of the fluorescent spots on the surface (due to DNA binding and formation of RP_C complexes) is precisely defined in the single-molecule trajectories by the simultaneous appearance of Cy3B and ATTO647N fluorescence signals ("DNA binds" time point, Figure 1(C)). The −15/+15 ruler reports low FRET for the RP_C complex, and intermediate FRET for the RP_O complex, since the formation of the transcription bubble decreases the distance between the −15 and +15 positions in the DNA.[33] The $RP_C \rightarrow RP_O$ transition in the trajectories is thus indicated by a sharp FRET increase ("DNA transcription bubble opens" time point, Figure 1(C)), which may occur in one or several steps, depending on the intermediates on the reaction pathway.

Our experimental single-molecule trajectories indeed show the expected fluorescence intensity and FRET signatures of RP_C complex formation and isomerisation to the RP_O state (Figure 1(D)). The moment of RNAP-promoter complex formation was precisely defined by the simultaneous appearance of Cy3B and ATTO647N fluorescence in the single-molecule trajectories (e.g., at 12.25 s and 13.5 s in the left and right panels, respectively, of Figure 1(D)). The apparent FRET efficiency (E^*) of the first stable complexes was $E^* \sim 0.2$ (Figure 1(D)), a value identical to that we obtained previously for the closed transcription bubble state.[33] After a short time, the FRET increased to $E^* \sim 0.45$ (at ~12.6 s and ~15.7 s in the traces of Figure 1(D)), a value identical to that we obtained previously for the open transcription bubble state.[33] DNA binding to the coverslip surface was strictly mediated by the RNAP, since the number of non-specific DNA binding events was negligible in the absence of RNAP on the surface (cf. Figure S2(A, B)). On the population level, the newly formed RNAP–promoter complexes displayed a bimodal FRET distribution, with mean FRET values ~0.2 and ~0.45 (Figure S2(C)) contrasting with the unimodal FRET distribution (mean ~ 0.18) of the protein-free immobilised promoter DNA (Figure S2(D)). To test whether the ~0.45 FRET state is indeed a catalytically competent RP_O complex, we added NTPs to the sample buffer; this addition almost eliminated the ~0.45 FRET state, as expected if RP_O complexes engage RNA synthesis and escape the promoter (Figure S2(C)).

To provide further proof for the formation of catalytically active RP_O complexes *in situ* on the coverslip surface, we performed experiments using a promoter with a different labelling scheme, which is very effective in monitoring the progress of initial transcription (dyes at positions −15 and +20; Figure S1(C)). The scrunching of the downstream DNA towards the RNAP during initial RNA synthesis leads to a stepwise increase in FRET until RNAP escape from the promoter returns the FRET to a low level (Figure S3(A, B)).[24] Example trajectories in Figure S3(C) demonstrate abortive initiation and promoter-escape events occurring shortly after the formation of the RNAP–promoter complexes. However, we note that some RNAP–promoter complexes (typically 20–50% of all complexes) on the surface neither form RP_O nor engage RNA synthesis; these molecules remain in stable low FRET state (~0.2) and may thus represent unproductive complexes resulting, e.g., from RNAP binding to the ends of the promoter DNA fragment (Figure S3(D)). HMM analysis of these complexes (on both −15/+15 and −15/+20 labelled promoters) did not produce Viterbi changes, further corroborating our interpretation that the transition from low FRET state ($E^* \sim 0.2$) to a long-lived higher FRET state ($E^* \sim 0.45$) on −15/+15 promoter indeed indicates a RP_O formation event instead of spurious fluorescence fluctuation. However, because the FRET sensitivity is not sufficient to confidently distinguish RP_C complex from non-specific RNAP–DNA complexes, we decided to analyse further only the RNAP–promoter complexes which directly show the appearance of the FRET signature of the RP_O complex (i.e., the ~0.45 FRET state) on the −15/+15 labelled promoter.

Extended promoter upstream sequence stimulates RP_O formation

To study the kinetics of RP_O complex formation in real-time and its modulation by the αCTD–promoter upstream interactions, we performed experiments using a long (DNA extending from position −89 to position +25) and short (DNA extending from position −39 to position + 25) version of the *lacCONS* promoter (Figure 2(A)). We also examined the kinetics of open complex formation by using additional versions of the promoter DNAs, which were either fully double-stranded (dsLC2 promoter; Figure S1(A)) or contained a mismatch in the promoter region from −10 to −4 (a.k.a. pre-melted promoter, pmLC2; Figure S1(B)).

RP_O formation was inefficient in the case of short dsLC2; in fact, we could identify only 5 real-time promoter-binding events indicating RP_O complex formation (3% of all promoter-binding events,

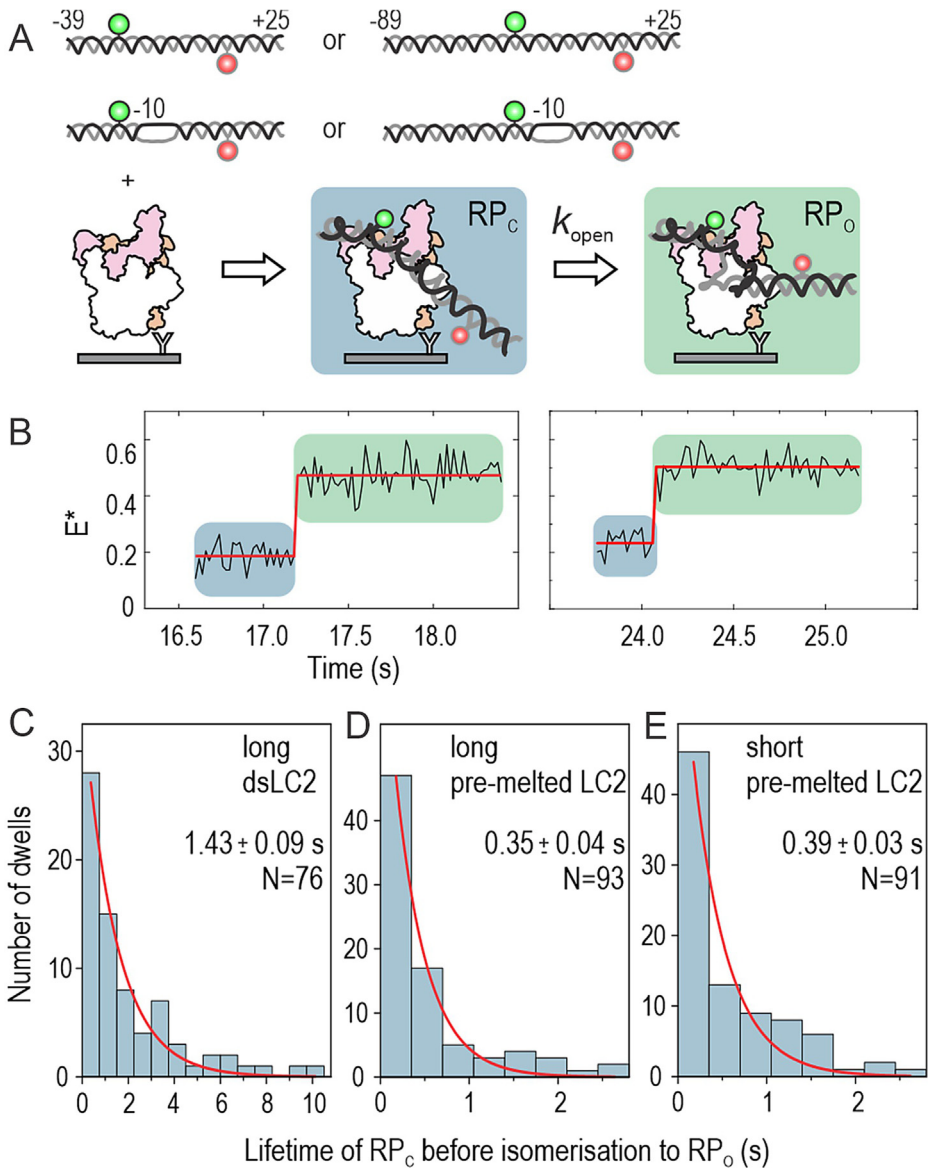

Figure 2. Rate of RP$_O$ formation. (A) Schematic representation of the real-time RP$_O$ formation experiment. The promoter has donor and acceptor labels at positions −15 and +15, respectively. Promoters were employed as short (promoter span −39/+25), long (−89/+25), fully double-stranded and pre-melted (mismatch region −10/−4) versions. (B) Example trajectory on the left demonstrates promoter binding to the surface-immobilised RNAP and the formation of RP$_C$ complex at 16.6 s. The RP$_C$ isomerises to RP$_O$ complex at 17.2 s. Another example trajectory of real-time RP$_O$ complex formation is shown on the right. Dwell-times in the RP$_C$ state were fit to mono-exponential function to obtain the rate constant of RP$_O$ complex formation (C) on the long dsLC2 promoter, (D) long pre-melted LC2 promoter and (E) short pre-melted LC2 promoter.

N = 167); even after prolonged incubation of the RNAP–promoter complexes (∼5 min) on the surface, the RP$_O$ complex (i.e., the FRET species with E^* ∼ 0.45) was nearly absent from the population histogram (Figure S4(A)). In contrast, the RP$_O$ complex formed efficiently on the long promoters, as well as on the short pre-melted promoter, as seen in the E^* histograms (Figure S4 (B–D)) and individual trajectories (Figure 2(B)).

We then performed Hidden Markov modelling (HMM) of the trajectories to extract the dwell times in the RP$_C$ state (E^* ∼ 0.2) before transcription bubble opening and RP$_O$ complex formation (Figure 2(B)). The observed distribution of dwell times in the RP$_C$ state for the long dsLC2 promoter (Figure 2(C)) was fitted to a mono-exponential decay function to determine a mean lifetime for the RP$_C$ complex of 1.43 ± 0.09 s (±SE; amplitude parameter was 35.2 ± 1.5). We

also tried to fit the dwell-time distribution using a bi-exponential equation, but rejected this more complex kinetic model because the fit parameters were poorly defined as evident from large SE (27% for lifetimes and 13–45% for amplitudes). Using a similar analysis, we estimated the RP_C complex lifetime as 0.35 ± 0.04 s on the long pre-melted LC2 (Figure 2(D)) and 0.39 ± 0.03 s on the short pre-melted LC2 (Figure 2(E)), respectively. These values indicate that the αCTD interactions with the upstream sequence (−89 to −40) significantly enhance the isomerisation rate of the RP_C to RP_O complex; however, this happens only on a fully double-stranded promoter. Because the introduction of the pre-melted region (−10/−4) to the promoter nearly equalised the rate of RP_C isomerisation to the RP_O complex on the short and long promoters, the αCTD–promoter interactions appear to predominantly contribute to the lowering of the activation energy of initial transcription bubble nucleation.

A subpopulation of RP_O complexes form via a kinetically significant intermediate

Close inspection of the HMM fit to the RP_O formation FRET trajectories revealed that, even though most bubble-opening events were described by a two-state model, (i.e., the promoter conformation in the initial complex changed to the RP_O state in a single step; Figure 2(B)), a subpopulation included an intermediate state (hereafter, RP_i complex) between the initial RP_C and final RP_O states (Figure 3(A)).

Specifically, the RP_i complex was identified in 20% (exact 95% binomial confidence interval[34]: 11–30%), in 14% (8–23%) and in 14% (8–23%) of all trajectories in the case of long dsLC2, long pmLC2 and short pmLC2 promoter, respectively. The arithmetic mean lifetime of the RP_i state was 0.32 s (95% CI by bootstrapping (10,000 iterations): 0.18–0.49 s, $N = 15$), 0.15 s (0.07–0.26 s, $N = 13$) and 0.17 s (0.10–0.23 s, $N = 13$) on the long dsLC2, long pmLC2 and short pmLC2 promoter, respectively. Similarity of the estimates suggests that the RP_i lifetime does not strongly depend on the used promoter type.

The sporadic appearance of the RP_i in the trajectories could indicate either heterogeneous reaction step (bubble opening with or without RP_i intermediate), failure to detect most of the RP_i states (if RP_i was an obligatory intermediate) or false positive HMM state assignment (if RP_i does not really exist). To evaluate these scenarios, we used simulated smFRET trajectories and estimated the efficiency of RP_i detection by the HMM routine. To obtain a conservative estimate, the trajectories were simulated using a short mean RP_C lifetime (0.16 s) and FRET signal noise that was 15–50% higher than in experimental data (Figure S5(A–F), example trajectories; Figure S5 (G–L), trajectory noise). As expected, the detection efficiency increased with the length of the RP_i dwell: 1, 2–4 and ≥5 frame dwells (1 frame = 20 ms) were detected with ∼15%, ∼60% and ∼80–90% efficiency, respectively (Figure S5 (M)). The false positive rate of RP_i on trajectories simulated without this state was only 1% (4 events in 360 trajectories) and does not thus affect conclusions. To obtain a rough estimate of how much the detection efficiency contributes to the estimated RP_i lifetime, we pooled and binned all RP_i dwells on different promoters and corrected the bins for the missed RP_i dwells. The fit of corrected and uncorrected dwell time histograms to mono-exponential function estimated the RP_i lifetime as 0.15 ± 0.02 s (±SE) and 0.22 ± 0.4 s, respectively (Figure S5(N)). Taking into account the dwell length distribution and detection efficiency, we would expect to identify the RP_i state in ∼60% of trajectories if the $RP_C \rightarrow RP_O$ transition would always proceed via this intermediate. The fact that this number is significantly larger than the measured 14–20% occurrence of RP_i raises the intriguing possibility that the RP_O formation pathway on *lacCONS* promoter is branched, i.e., one path is a direct $RP_C \rightarrow RP_O$ transition while the other path involves the RP_i as an intermediate between these states.

To compare the mean FRET of the RP_i intermediate to that of the RP_C and RP_O complexes, we extracted FRET efficiency histograms from HMM-segmented trajectories for each of the three states (Figure 3(B)). The mean FRET values, obtained as the centres of the fit Gaussian distribution (Eq. (2)), were found as 0.19 6 ± 0.001 (±SE), 0.318 ± 0.002 and 0.448 ± 0.001 for the RP_C, RP_i and RP_O complex, respectively. The mean FRET values suggest that the average distance between the −15 and +15 labels in the RP_i state has become shorter than in the RP_C complex but remains still longer than in the mature RP_O complex.

To probe the structural origin of the RP_i state, we included in the reaction buffer the RNAP inhibitor myxopyronin B (Myx). Biochemical and structural studies using Myx[35] and structural studies using corallopyronin A (Cor),[9] a Myx analogue, have suggested that this class of RNAP inhibitors block the formation of RP_O complex by preventing the loading of template DNA strand into the active site cleft. We observed that the FRET distribution of the RNAP–promoter complexes preformed in the presence of Myx was described by two Gaussians with mean FRET values 0.231 ± 0.001 and 0.307 ± 0.010 (Figure 3(C), long dsLC2 promoter). The inspection of individual trajectories revealed three classes of molecules: the first and most abundant class involved RNAP–promoter–Myx complexes characterised by an $E^* \sim 0.3$ state ($N = 58$, 60% of all molecules, Figure S6(A)). Interestingly, a sub-fraction ($N = 19/58$) of these molecules stochastically sam-

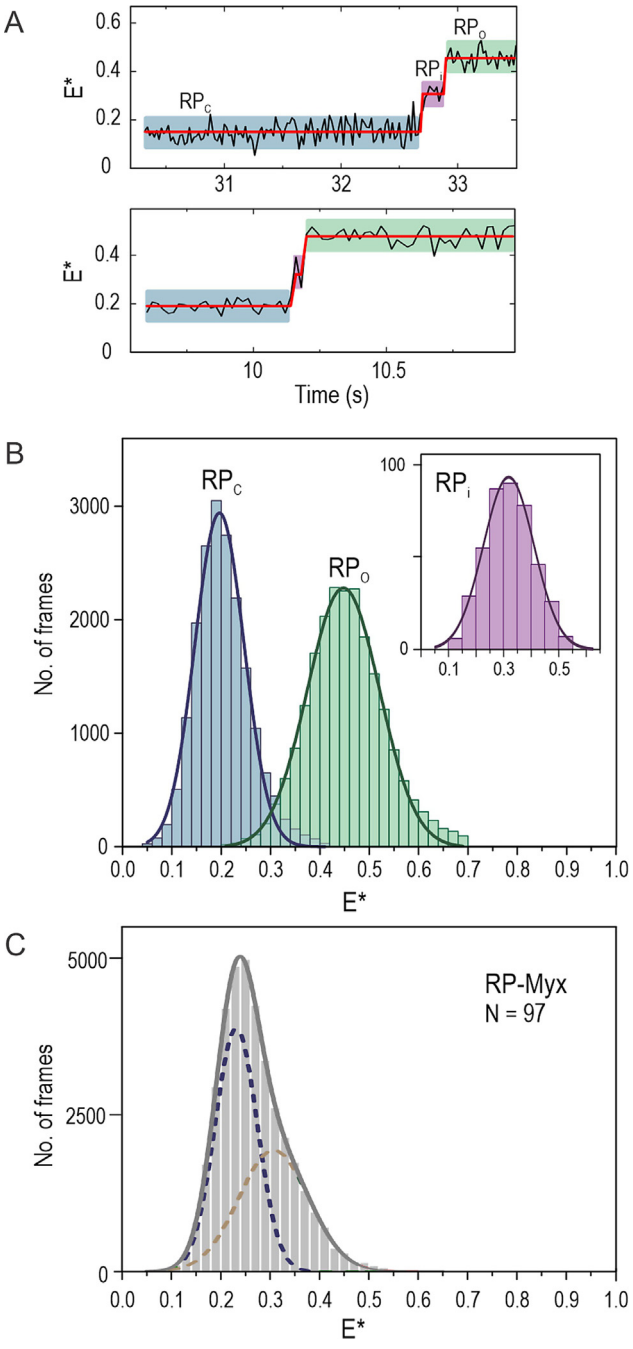

Figure 3. Intermediate on the RP_O complex formation pathway. (A) Example trajectories demonstrate the presence of an intermediate, RP_i complex, on the pathway from RP_C to RP_O complex. The trajectory was fit to 3-state HMM. (B) E^* histograms for the RP_C, RP_i and RP_O states were extracted from the HMM segmented trajectories. The E^* distributions were fit Eq. (2). Data from different promoter versions was pooled. (C) E^* histogram of the RNAP–promoter complexes formed in the presence of 100 μM Myx inhibitor. The complexes were imaged ~5 min after their initial formation on the coverslip surface. E^* distribution was fit to Eq. (2).

pled a very short-lived, i.e., typically 20–40 ms (1–2 frames), higher E^* state (Figure S6(B)). The second class of molecules involved potential non-productive complexes as indicated by a stable $E^* \sim 0.2$ value (N = 37, 38%, Figure S6(C)) and the third class involved RP_O complexes characterized by a long-lived $E^* \sim 0.45$ state (N = 2, 2%, Figure S6(D)). Preformed complexes on the long pre-melted LC2 promoter confirmed the bimodal FRET distribution with two sub-populations having mean E^* values of 0.224 ± 0.002 and 0.290 ± 0.045 (Figure S6(E)).

Consistent with the above equilibrium FRET values, 32% (N = 21) of real-time promoter

binding trajectories in the presence of Myx inhibitor demonstrated the formation of initial RP_C complex ($E^* \sim 0.2$) and its subsequent isomerisation to $E^* \sim 0.3$ state (Figure S6(F)) while the remaining 68% ($N = 45$) of the nascent RNAP–promoter complexes maintained the $E^* \sim 0.2$ state for the entire duration of the trajectory (Figure S6(G)). The increasing trend in observed FRET values as the RNAP–promoter complexes react towards RP_O is consistent with structural modelling data; the distance between the −15 and +15 labels decreases from 98 Å in the RP_C complex, to 87 Å in the Cor-stabilised RNAP-promoter intermediate, and further to 66 Å in the RP_O complex (Figure S6 (H)).

Transcription bubble opening leads to static and dynamic RP_O complexes

We next analysed the transcription bubble behaviour immediately *after* initial RP_O complex formation; our observation span for these measurements was 1.3–22 s (median = 4 s, $N = 119$). HMM analysis of the trajectories revealed two RP_O complex sub-populations: a "static" (or "stable") sub-population (73% of the nascent RP_O complexes; exact 95% binomial CI: 64–81%), where the transcription bubble remained open for the entire duration of the trajectory (reflected by an $E^* \sim 0.45$ state; Figure 4(A)); and a dynamic sub-population (27%, CI: 19–36%), where the transcription bubble fluctuates rapidly between the open state ($E^* \sim 0.45$) and state(s) characterised by lower FRET (Figure 4(B)). The dynamic RP_O complexes do not appear to convert to a static RP_O within our observation span, suggesting that the dynamic RP_O–like complex is not an on-pathway intermediate which eventually isomerises to form the stable RP_O complex. This conclusion, which invokes the formation of static and dynamic RP_O complexes on parallel pathways, is also supported by the presence of a similar distribution of static and dynamic RP_O complexes in samples of preformed, heparin-challenged RNAP–promoter complexes (see next section of the manuscript and Ref. 33). Notably, both the 2-state (i.e., showing no RP_i intermediate; Figure 2(B)) and 3-state (Figure 3(A)) bubble-opening mechanisms produced both static and dynamic RP_O complexes (Figure 4(C)), suggesting that the presence of the RP_i intermediate does not dictate the subsequent stability of the bubble.

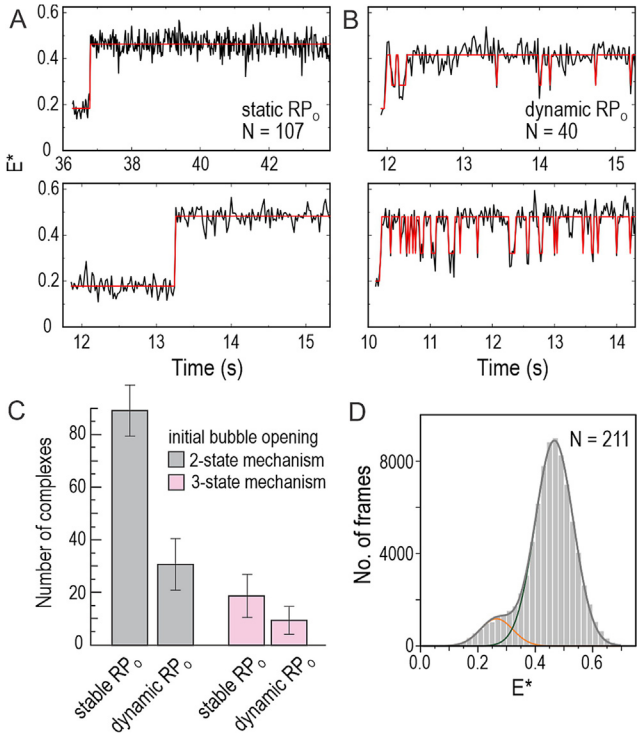

Figure 4. Parallel formation of static and dynamic RP_O's. (A) Example trajectories demonstrating the formation of static RP_O. (B) Example trajectories demonstrating the formation of dynamic RP_O. The static population represents 73% ($N = 107$) of the nascent RP_O complexes, whereas the dynamic population represents 27% of the complexes ($N = 40$). (C) The abundance of static and dynamic RP_O's is shown separately following the initial bubble opening either via 2-state (grey bars) or 3-state (pink) mechanisms. Exact 95% binomial confidence intervals are shown. (D) E^* histogram of dynamic RP_O's. The complexes were imaged ~5 min after their initial formation on the coverslip surface ($N = 211$ molecules). The E^* distributions were fit using Eq. (2).

To evaluate the bubble conformations accessed by the dynamic RP_O complexes, we analysed the E^* distribution of these complexes, which was fit well by two Gaussian functions with mean E^* values of 0.265 ± 0.004 and 0.467 ± 0.001 (Figure 4(D)). In contrast, the E^* histogram of the static RP_O showed only a single distribution centred at E^* of 0.448 ± 0.001 (Figure 3(B)). These values suggest that the dynamic RP_O complexes do not sample either the RP_C ($E^* \sim 0.20$) or the on-pathway intermediate RP_i ($E^* \sim 0.32$) states. Instead, it is more likely that the dynamic RP_O complexes sample an off-pathway state, which is characterised by $E^* \sim 0.27$ and which we coin as RP_{ISO}.

The role of RNAP–promoter upstream interactions in the RP_O pathway selection

We next evaluated how RNAP αCTD–promoter upstream sequence interactions contribute to the relative formation of stable and dynamic RP_O complexes and the kinetic parameters of the transcription bubble dynamics (Figure S7(A)). To this end, we prepared RP_O complexes at 37 °C using either a short or a long dsLC2, challenged them with heparin, i.e., a DNA competitor, and immobilised them on the coverslip surface for smFRET analysis. In this protocol, RP_O complexes also form on the short dsLC2 promoter fragments, allowing direct comparison to the long dsLC2 promoter. HMM-based classification of the trajectories indicated that the dynamic RP_O complexes were 1.7-fold more prevalent (25 ± 3.6% vs 15 ± 2.7% of all complexes; mean and SD of three independent experiments) on the short promoter compared to the long promoter (Figure S7(B)). A two-sample *T*-test ($p = 0.035$) confirmed that the observed difference in the relative number of dynamic complexes on the two promoters is statistically significant. Further, the observation span for the measurements was 2.0–25 s (median 8.0 s, $N = 348$) on the long promoter and 1.4–25 s (median 6.9 s, $N = 435$) on the short promoter, suggesting that the higher prevalence of dynamic RP_O complexes on the short promoter is not explained simply by longer trajectories that are expected to accumulate more state transitions.

As an additional control, we estimated the identification accuracy of dynamic/static RP_O complexes in simulated trajectories, which had state lifetimes and FRET noise levels similar to the experimental data. HMM analysis results indicate that the detection efficiency of dynamic RP_O complexes is ∼95% when the trajectory length is ≥4 s (a length observed for 84% of experimental trajectories) and remains at 84% when the trajectory length is ≥3 s (a length observed for 92% of experimental trajectories) (Figure S7(C)). These numbers indicate that the trajectory length variation in the experiment does not significantly bias the static/dynamic RP_O estimations. Noteworthy, the analysis of simulated trajectories did not produce any false positive dynamic RP_O's.

Kinetic analysis indicated that the lifetimes the RP_{ISO} (85–90 ms, Figure S7(D, F)) and RP_O (560–660 ms, Figure S7(E, G)) states were similar on both promoters (Figure S7(H)). Collectively, our results suggest that the αCTD–promoter interactions steer the RP_O pathway selection towards the static RP_O complex; however, the effect is moderate, and significant number of dynamic RP_O complexes form also on the long promoter. The similarity in the timescales of transcription bubble dynamics on the short and long promoters indicates that the bubble isomerisation rates are independent of the status of the αCTD–promoter upstream sequence interactions.

The role of RNAP cleft loops in the RP_O stabilisation

To interrogate protein structural elements contributing to the RP_O stability, we deleted the β gate loop (ΔGL), β' rudder loop (ΔRL) or β' lid loop (ΔLL) from the RNAP and determined the effects of these deletions on the transcription bubble dynamics using preformed RP_O complexes. Structurally, GL mediates the RNAP β-pincer interaction with the RNAP β'-clamp, thus forming a barrier for the DNA entry and exit from the RNAP cleft (Figure 5(A))[9,10]; RL locates between the tDNA and ntDNA strands in the RP_O; and LL locates adjacent to the RL and is able to interact with the tDNA around base −6 in the RP_O.

We used 2-state HMM of FRET trajectories to classify RP_O complexes into static (i.e., no bubble dynamics) and dynamic (Figure 5(B)) and found that all deletions shifted the balance of RP_O formation towards the direction of dynamic RP_O complexes. The effects of ΔGL and ΔLL were moderate, as they increased the fraction of dynamic RP_O complexes from the wild-type (WT) RNAP level by 1.3–1.4-fold, i.e., from 26% (exact 95% binomial CI: 22–31%) in WT to 34% (CI: 29–40%) in ΔLL and 36% (CI: 30–43%) in ΔGL (Figure 5(C)). However, the ΔRL effect was much stronger, 2.9–fold, making most RP_O complexes, i.e., 77% (CI: 72–83%), dynamic. The addition of NTPs to ΔRL-RNAP–promoter complexes depopulated the RP_O state as indicated by the substantial decrease in $E^* \sim 0.45$ state probability and corollary increase in ∼0.2 E^* state (Figure 5(D)). Because the NTP-response was similar in the case of WT-RNAP (Figure S2(C)), which in contrast to ΔRL-RNAP forms mostly static RP_O's, both static and dynamic RP_O's appear capable to initiate RNA synthesis. Kinetic analysis of the dynamic RP_O's indicated that ΔRL and ΔLL shortened the lifetime of the open bubble state by 2.8–fold, whereas ΔGL had no effect (Figure 5(E, F)). In contrast, all three mutations increased the

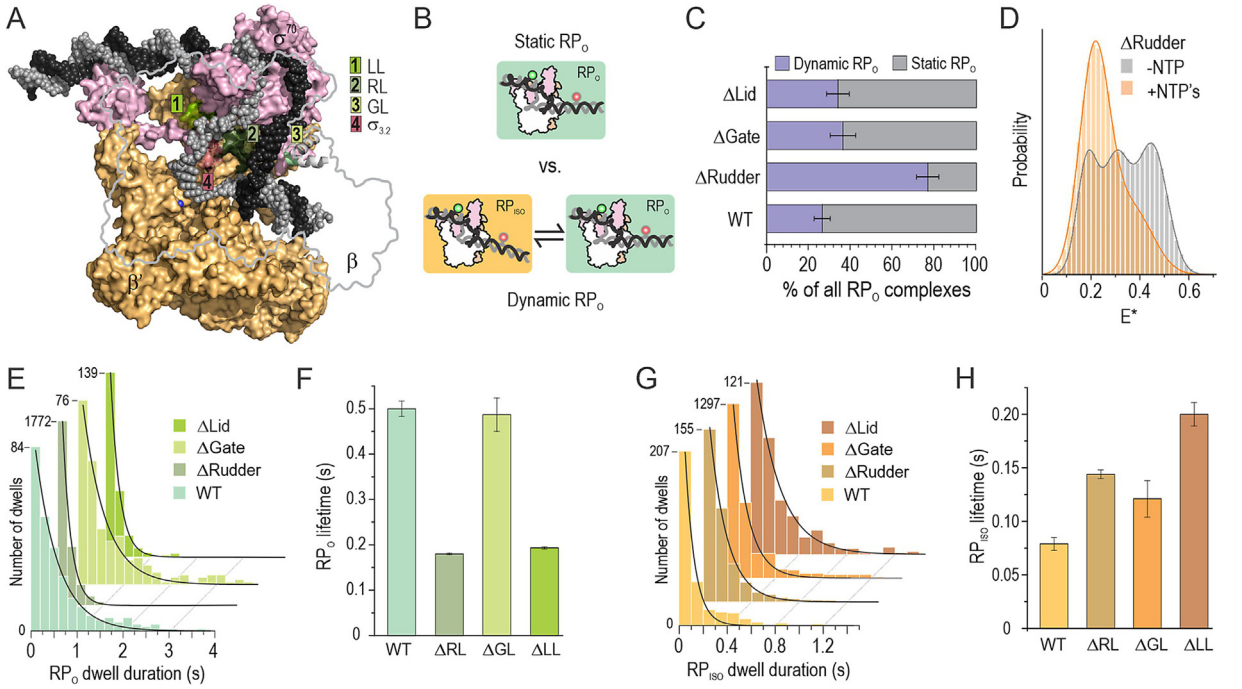

Figure 5. Effect of RNAP cleft loop deletions on the reaction pathway branching and transcription bubble kinetics. (A) The position of deleted lid loop (LL), rudder loop (RL) and gate loop (GL) are shown in the cryo-EM model of *E. coli* RP_O (PDB: 6psw, (10)). α, β and ω RNAP subunits and TraR transcription factor are omitted for clarity. ntDNA and tDNA are shown in dark and light grey, respectively. Blue sphere is the active site Mg^{2+} ion. (B) The RP_O complexes were classified as static or dynamic based on the 2-state HMM fit of the FRET trajectories. (C) The relative amounts of static and dynamic RP_O. Error bar: exact 95% binomial CI. WT, $N = 212$ molecules; ΔRL, $N = 206$; ΔGL, $N = 136$; ΔLL, $N = 115$. (D) The E^* histogram of ΔRL RNAP–promoter complexes on the coverslip surface before (grey, $N = 126$ molecules) and 2–5 min after the addition of 1 mM NTPs (orange, $N = 127$). (E) Dwell time distributions of the RP_O state within the dynamic RP_O population were fit to mono-exponential equation. (F) The lifetime of RP_O state was obtained from the dwell distributions in panel E. Error bars are 1 SE extracted from the fit. (G) Dwell time distributions of the RP_{ISO} state within the dynamic RP_O population were fit to mono-exponential decay equation. (H) The lifetime of RP_{ISO} state was obtained from the dwell distributions in panel G. Error bars are 1 SE extracted from the fit.

lifetime of the RP_{iso} state. Specifically, the RP_{iso} lifetime increased by 2.5-, 1.8- and 1.5-fold in the case of ΔLL, ΔRL and ΔGL mutant RNAPs, respectively (Figure 5(G, H)). Notably, the median observation span for these measurements was 4.26 s, 4.84 s, 5.20 s and 6.00 s in the case of WT, ΔLL, ΔGL and ΔRL, respectively. The variation in the observation span, however, does not significantly contribute to the classification of the molecules to the static and dynamic RP_O classes or kinetic parameters because, even in the combination of shortest trajectories (median 4.26 s) and the most stable state, i.e., wild-type RP_O complex (lifetime 0.5 s), the probability that $RP_O \rightarrow RP_{ISO}$ transition does not take place within the trajectory is extremely small (0.0002). Collectively, our data indicate that the GL, RL and LL domains in the RNAP contribute to the RP_O pathway branching (by changing the balance between static and dynamic RP_O complexes) and affect the rates of the transcription bubble conformation changes.

Discussion

In this work, we establish the ability to look at the earliest stages of transcription initiation in real-time and at the single-molecule level. This unique capability bypasses the need to synchronise complexes and offers unprecedented access to co-existing reaction pathways and transient intermediates; as a result, we gained new mechanistic insight about the paths and intermediates used by RNA polymerase to form RP_O complexes on a *lac* promoter derivative. Our work also provides further insight on the dynamics and heterogeneity of open complexes and their determinants.

Open complex formation may proceed via more than one path. Our data indicate that the $RP_C \rightarrow RP_O$ isomerisation step involves **mechanistic branching**. In most cases (~60% of molecules), the isomerisation occurs in one step without observable intermediate(s) within our 20-ms temporal resolution, suggesting that bubble

initiation was followed by rapid bubble progression, downstream DNA loading to the RNAP cleft, and template DNA loading to the active site cleft. However, the remaining molecular trajectories (~40%) have an **intermediate state** (**RP_i**) between the RP_C and RP_O. We employed simulated trajectories to exclude the possibility that all experimental trajectories without an identified RP_i are simply due to the failure to detect this state. Our data support the presence of parallel paths for RP_O formation, or at least, the presence of RNAP populations with different propensity to form an open complex.

Possible structure of RP_i. We first considered that possibility that the intermediate may have structural similarities to an RNAP-promoter complex stabilised by an antibiotic targeting RNAP. Recent cryo-EM data of *Mycobacterium tuberculosis* RNAP showed that corallopyronin A (a Myx analogue) stabilises a partially melted transcription bubble (region −11/−4).[9] The same cryo-EM work showed that a similar conformation was present in the absence of Myx, raising the possibility that the observed conformation may represent an intermediate on the RP_O pathway.[9] However, our real-time trajectories using *E. coli* RNAP show an intermediate (RP_i) with a structural signature with significantly different FRET efficiency ($E^* \sim 0.3$) than that expected by the partially melted intermediate ($E_{i2} \sim 0.2$). Further, the presence of Myx does not prevent full opening of the promoter DNA, as sensed by fluorescence enhancement of a Cy3 probe introduced at position +2, both at the ensemble[35] and at the single-molecule level[36]; we note that such enhancement is expected only when the bubble opening has been complete.

Instead, the possibility we favour for the structure of the RP_i is an intermediate further down the promoter opening pathway (hence, more consistent with the E^* value of ~0.3 we observe for RP_i), where all the melting has been completed, but the template DNA has not been fully loaded to the active site cleft; such a structure is supported by results showing that Myx does not prevent full opening of DNA on two prototypical promoters (λP_R and *lacCONS*).[36] Regardless, at least for the *lac* promoter derivative used in this work, the intermediate is kinetically significant only in a subset of RP_O formation events.

Possible sources for heterogeneity in open complex formation pathways. Since the **RP_i** is detectable only in a subset of trajectories, it is conceivable that, for those trajectories, a structural module of the RNAP delays downstream tDNA loading to the active site cleft. One candidate for such a structural module is the RNAP β′ **switch-2 segment**. Based on mutational analysis, structural studies, and observed Myx effects, it has been established that the switch-2 region can adopt two different conformations.[37] The conformation dominant in the absence of Myx is compatible with template loading to the active site; in contrast, the alternative conformation (which is stabilised by Myx and specific mutations in the switch-2) blocks template loading to the active site. If the switch-2 was at the moment of DNA bubble initiation in the blocking conformation in a fraction of RNAPs, the loading of template DNA to the active site cleft would be delayed by the necessary switch-2 refolding.

The RP_i heterogeneity may also result from alternative **promoter discriminator** (region −6/+1) conformations in different RNAP molecules. This hypothesis is supported by the previous finding that $G^{-6}G^{-5}G^{-4}$ and $C^{-6}C^{-5}C/T^{-4}$ motifs in the ntDNA stabilised *in crystallo* two distinct discriminator conformations and imposed in solution one base-pair difference in the predominant transcription start site.[38] GTG motif, which is found in our promoter, had transcription start site statistics halfway between the GGG and CCC/T motifs, consistent with the assumption that a promoter with this motif can readily adopt either of the two discriminator conformations.

RP_O complexes appear to have different stability immediately after their formation. A longstanding question in the transcription field is whether all RP_O on a given promoter are the same or differ in their structural and functional properties.[22–25] Our results show that indeed there is another layer of heterogeneity as indicated by the **differing stability of the transcription bubble,** even immediately after RP_O formation (as judged by the appearance of the $E^* \sim 0.45$ state). Most RP_O complexes can keep the bubble open for at least several seconds; however, a more dynamic RP_O subpopulation samples different bubble states in the millisecond timescale. The stable or dynamic RP_O mode was set before or during the first-time opening of the bubble and the modes did not interconvert within our observation span (~8 seconds); this observation rules out the possibility that the dynamic RP_O were indeed intermediates on the linear pathway leading to the stable RP_O complexes. This new insight aligns well with our previous observation of stable and dynamic RP_O molecules within the population of pre-formed RP_O complexes,[33] while allowing further mechanistic insight by linking the formation of stable and dynamic complexes on the existence of a branched RP_O pathway and the sampling of a short-lived off-pathway state (RP_{ISO}) by the dynamic RP_O's.

The analysis of **mutant RNAPs** suggest that the main difference between the dynamic and stable RP_O complexes arises from the RNAP interaction with the single-stranded template DNA in the active site cleft. The deletion of rudder loop, which presses against the template DNA positions −7 to −9, tripled the amount of dynamic RP_O (from 26% to 77%; Figure 5(C)) and decreased 3-fold the open bubble lifetime in the dynamic RP_O complexes (Figure 5(F)). The deletion of lid loop,

which interacts with the template DNA base −6, had a similar effect on the open bubble lifetime. We previously found that deletion of the σ^{70} 3.2 region (the σ "finger", which interacts with the template DNA strand from bases −3 to −6), destabilised the RP_O.[33] These interactions with the template DNA form late in the RP_O mechanism, i.e., when the bubble forms fully and the template DNA strand loads to the active site cleft. Our data suggest that these final interactions form by two alternative ways generating "tight" and "loose" template DNA binding modes: the tight binding mode gives rise to the stable RP_O complexes, and the loose binding mode gives rise to the dynamic RP_O complexes.

It will be interesting to investigate in future single-molecule studies whether such significant differences in RP_O stability have functional consequences, and whether they are related to reports of non-uniform RP_O function. Specifically, a subset of RP_O complexes (on many promoters) appear to be locked in an abortive initiation mode, in which they repetitively synthesise short RNA products (<12-mer, with the exact sequence depending on the specific promoter), whereas another RP_O subset escape the promoter efficiently and synthesise full-length RNA products.[22–25] The failure of promoter escape may also result from long-lived backtracking and arrest of initially transcribing RNAPs.[28,29] The RP_O's apparently locked in the abortive mode are also referred as "moribund" complexes, and they apparently have a role in transcription regulation in the cell.[27] Mechanistically, the dynamic RP_O could be candidates to form such moribund complexes, since unstable template DNA binding to the active site is likely to enhance the dissociation probability of short RNAs, leading to abortive initiation. Consistent with this reasoning, the Δ3.2 σ^{70} mutant (which increase RP_O dynamics substantially) released 4–7-mer RNAs more efficiently compared to the WT.[24] However, it has also recently been suggested that the intermediate RP_{i3} and not the stable RP_O is the productive initiation complex on the λP_R promoter.[39,40] Our observation of NTP-dependent depopulation of RP_O's formed either by wild-type RNAP (mostly static RP_O's) or ΔRL RNAP (mostly dynamic RP_O's) indicates that both forms of RP_O's can at least initiate RNA synthesis (Figure S2(C), Figure 5(D)).

Role of the promoter upstream interactions on the RP_O formation and stability. The **RNAP αCTDs** interact with the promoter upstream sequences either by specifically recognising the promoter UP element[14] or via sequence-independent interactions[14,19]; both interactions are important for RP_O formation. We found that the upstream part of the *lacCONS* promoter (from −40 to −89; Figure S2(A, B)), which does not contain a full UP element but is partially similar to the distal UP subsite,[41] facilitates transcription bubble melting in the context of fully double-stranded promoters. In fact, the short double-stranded promoter (which lacks αCTDs interactions) failed to form RP_O under our experimental conditions, which involve measurements at room temperature. On the other hand, if the requirement for the DNA melting nucleation was bypassed (e.g., by using a pre-melted promoter), the αCTD interactions with upstream sequence no longer increased the rate of RP_O formation (Figure 2(D, E)). This finding is consistent with previous biochemical studies showing that the αCTD interactions with the UP element enhance both the initial promoter binding and subsequent isomerisation to competitor-resistant conformation.[14,19]

We also found that the presence of upstream sequence interactions did not substantially change the ratio of initial bubble opening events that occur in one step or in two steps (i.e., via the RP_i), and did not significantly change the rates of transcription bubble dynamics in the pre-formed RP_O. However, the dynamic RP_O complexes formed more often on the short promoter in comparison to the long promoter (25% vs. 16%), suggesting that the αCTD–promoter interactions, instead of being fully decoupled from mechanistic steps occurring after the bubble nucleation, have a role in the late steps of RP_O pathway branching; the exact mechanism of such modulation is unclear, but it may involve the bending of the upstream sequence on the RNAP surface, as observed by Record and collegues,[17] and subsequent interactions that affect RNAP conformation dynamics in a way that it influences bubble dynamics. Our promoter sequence has only a partial similarity to the distal UP element subsite, predicting a non-ideal (if any) sequence-specific interaction with αCTDs. It has been shown that promoters with a full UP element drive stronger transcription activity than promoters with partial UP element or non-UP element sequence.[41] To generalise our findings, further studies on many different promoters are needed to establish the UP element and general sequence-dependence of the promoter upstream control over the formation of static vs dynamic RP_O's.

A working model for the RP_O formation mechanism on a bacterial consensus promoter. We summarise our key findings in the context of the RP_O formation mechanism on the *lacCONS* promoter in Figure 6.[2] The process begins with the RP_C complex formation as the RNAP holoenzyme binds to the promoter and establishes interactions with the −35 element, −10 element and upstream sequences. Interaction of αCTDs with upstream sequences stimulate RP_O formation by bending the upstream DNA around the RNAP[12–17] and coupling it energetically with bubble formation.

The initial nucleated bubble expands via two different mechanistic paths: in the first path (most common for our *lac* promoter derivative), the

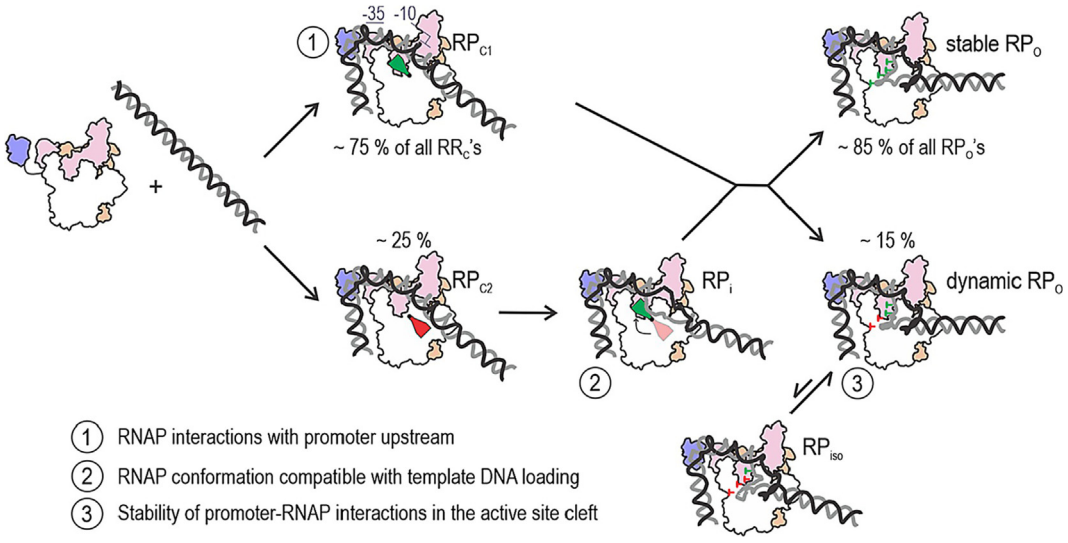

Figure 6. A working model for the RP_O complex formation mechanism on the *lacCONS* promoter. The reaction pathway from the promoter binding to the RP_O complex has heterogeneity in two separate steps. The first step is hypothesised to depend on a mobile RNAP element, which can be either in an active or inactive conformation (green and red flaps, respectively). The inactive conformation blocks the loading of the tDNA strand into the active site cleft, resulting the formation of intermediate RP_i. The isomerisation of the mobile element to the active conformation clears the block and allows progress from RP_i to RP_O. The second branching is related to the stability of the RNAP–template DNA interaction in the active site cleft. In ~15% of the RP_O complexes, the interaction is weak, allowing continuous dynamic movement of the template DNA and thus the downstream DNA. Because stable and dynamics RP_O complexes formed both from RP_{C1} and RP_i, we assume that these pathways merge before the next branching step leading to the stable and dynamics RP_O's. Green vs red pins depict tight and loose interactions between the tDNA and the RNAP, respectively. The numeration (1, 2 and 3) indicates the key steps in the mechanism that may define the rate and efficiency of RP_O complex formation.

RNAP melts the entire bubble and loads the template DNA strand to the active site cleft in one apparent step without detectable intermediates; the second path, however, involves a short-lived intermediate, RP_i, which features incomplete template loading to the active site cleft. We hypothesise that the intermediate appears when a mobile element of the RNAP, e.g., the switch-2 module, is initially in a conformation incompatible with template loading to the active site cleft. Template DNA loading to the active site cleft leads to the tight-binding and loose-binding states, which do not readily interconvert. Because stable and dynamics RP_O complexes formed with similar probability both directly from RP_C and via RP_i, we assume that these pathways merge before the branching to the stable and dynamics RP_O's takes place at the template DNA loading step (Figure 6). The tight template DNA binding mode keeps transcription bubble open whereas the loose-binding features dynamic movement of the template DNA. Template DNA interactions with the RNAP rudder loop and σ finger are part of the key determinants of tight binding mode. Ongoing studies in our group aim to decipher the promoter-sequence dependence of the RP_O formation mechanism and functional significance of the RP_O heterogeneity.

Materials and methods

Promoter preparation

Labelled and unlabelled DNA oligos to make *lacCONS* promoter, also known as *lacCONS+2*, constructs were purchased from IBA Lifesciences (Germany) (Figure S1(A–C)). Short *lacCONS* promoters (−39/+25) were reconstituted by annealing PAGE purified labelled template and non-template DNA oligos at 1 μM in annealing buffer [10 mM Tris-HCl (pH 8.0), 50 mM NaCl, 0.1 mM EDTA]. The annealing program consisted of initial denaturation (93 °C, 3 min) followed by step-wise cooling to 4 °C (each step: 1.2 °C, 30 s). DNA strands for the long *lacCONS* promoters (−89/+25) were constructed using a previously described protocol.[29]

Protein preparation

Escherichia coli core RNAPs were expressed in *E. coli* and purified as previously described.[42] The wild-type RNAP was expressed from plasmid pVS10 (T7p-α-β-β′_His_6-T7p-ω).[43] ΔRudder loop RNAP (T7p-α-β-β′[ΔN309-K325]_TEV_His_{10}-T7p-ω), Δlid loop RNAP (T7p-α-β-β′[ΔP251-S263 → G G]_TEV_His_{10}-T7p-ω) and Δgate loop RNAP (T7p-α-His_6_β[ΔR368-P376 → GG]-β′-ω) were

expressed from pMT041, pHM001 and pTG011, respectively.[44] Wild-type *E. coli* σ^{70} was purified as previously described.[45] Holoenzymes were assembled by incubating 0.5 μM RNAP with 1.5 μM σ^{70} for 15 min at 30 °C in RNAP storage buffer [20 mM Tris-HCl (pH 8.0), 150 mM NaCl, 50% (vol/vol) glycerol, 0.1 mM EDTA, 0.1 mM dithiothreitol (DTT)].

Microscope coverslip preparation

Borosilicate glass coverslips (1.5 MenzelGläzer, Germany) were heated to 500 °C in oven for 1 h to reduce background fluorescence. The coverslips were then rinsed with HPLC-grade acetone and immerged into 1% (v/v) Vectabond (product code #SP-1800, Vector Labs, CA, USA) in acetone for 10 min to functionalise the glass surface with amino groups. Coverslips were then rinsed with acetone followed by deionized water before drying them under a stream of nitrogen gas. A silicone gasket (103280, Grace Bio-Labs, OR, USA) with four reaction wells was placed in the middle of the coverslip. The coverslip surface was then simultaneously passivated by pegylation against unspecific protein/DNA binding and biotinylated to provide attachment points for specific protein immobilisation. Each well on the coverslip was thus filled with 20 μl of 180 mg/ml methoxy-PEG (5 kDa)-SVA (Laysan Bio, AL, USA) and 4.4 mg/ml biotin-PEG (5 kDa)-SC (Laysan Bio, AL, USA) in 50 mM MOPS-KOH buffer (pH 7.5), incubated for ~ 3 h at room temperature and finally the wells were thoroughly rinsed with phosphate-buffered saline (PBS; Sigma Aldrich, UK). The coverslips remained functional for at least two weeks when stored at 4 °C in plastic pipette tip box containing a layer of deionised water at the bottom. During the storage the coverslip wells were filled with PBS.

Single-molecule experiments

On the day of microscopy experiment, the pegylated coverslips were incubated for ~10 min with 0.5 mg/ml of Neutravidin (31050, ThermoFisher Scientific, UK) in 0.5 × PBS and subsequently rinsed with 1 × PBS. The coverslips were then incubated for ~10 min with 4 μg/ml of Penta·His biotin conjugate antibody (34440, Qiagen, UK) in reaction buffer [40 mM HEPES buffer (pH 7.3, BP299100, Fisher Scientific, UK), 100 mM potassium glutamate, 10 mM $MgCl_2$, 1 mM DTT, 1 mM cysteamine hydrochloride, 5% glycerol (vol/vol), 0.2 mg/ml bovine serum albumin] and subsequently rinsed with the reaction buffer.

To analyse the RP_O complex formation in real-time at 22 °C the anti-His-tag-antibody coated coverslip was incubated ~10 min with 1 nM label-free holoenzyme in the reaction buffer, rinsed thoroughly with the reaction buffer and mounted on the microscope. 25 μl of imaging buffer [i.e., reaction buffer supplemented with 2 mM UV-treated Trolox, 1% (w/v) glucose, 0.4 μg/ml catalase (10106810001, Roche Diagnostics, Germany), 1 μg/ml glucose oxidase (G2133, Sigma Aldrich, UK)] was replaced to the imaged well. Data recorder was started to take an 80 s movie. 1 μl of 4 nM promoter in the reaction buffer was gently pipetted to the well at ~8 s time-point. The formation of RNAP–promoter complexes was evident by the appearance of bright co-localised spots on the Cy3B and ATTO647N fluorescence channels. In some experiments these surface-formed RNAP–promoter complexes were further imaged after exchanging fresh imaging buffer to the well and finding non-bleached field-of-view. The age of RNAP–promoter complexes at the moment of recording these 20 s post-binding movies was ~3–7 min. In some control experiments, we monitored the initial RNA synthesis activity of RNAP by including 1 mM NTPs (ATP, GTP, CTP and UTP) in the imaging buffer.

To analyse transcription bubble dynamics in pre-assembled RP_O complexes, 2 nM holoenzyme was incubated with 5 nM promoter in reaction buffer for 15 min at 37 °C. 100 μg/ml sodium heparin (H4784, Sigma, UK) was added to disrupt non-specific RNAP–promoter complexes and ~1.3 μl of the mixture was transferred to anti-His-tag-antibody coated coverslip well containing 25 μl reaction buffer. The RP_O complex immobilisation at 22 °C was let to continue until ~50 molecules were detected on the field-of-view. The well was then rinsed with reaction buffer and finally filled with 25 μl imaging buffer. Data was recorded as 20 s movies from about ten field-of-view per well at 22 °C.

Single RNAP–promoter complexes were imaged using objective-based single-molecule TIRF microscope previously described.[46] The donor (Cy3B) and acceptor (ATTO647N) fluorophores in the promoter were excited using 532 nm and the 642 nm lasers in alternating laser excitation (ALEX) mode, respectively.[47] The emission of donor and acceptor fluorophores was separated from each other and from the excitation light, using dichroic mirrors and optical filters, and recorded side-by-side on an electron multiplying charge-coupled device camera (iXon 897, Andor Technologies, Northern Ireland). The frame time of the recordings was 20 ms with 10 ms ALEX excitation by each laser. The measured laser power before the dichroic mirror was set to ~4 mW and ~1 mW for the 532 nm and 642 nm laser, respectively.

Single-molecule data analysis

To extract the intensities of co-localised donor and acceptor fluorophores, the recorded movies were processed with custom-made TwoTone TIRF-FRET analysis software (46; see also

https://groups.physics.ox.ac.uk/genemachines/group/Main.Software.html). If the processed movies had fluorescent complexes on the surface already at the beginning of the movie, i.e., post-binding and pre-formed RP_O complex movies (see above), the following Twotone-ALEX parameters were applied to select only complexes containing a single ATTO647N acceptor dye and a single Cy3B donor dye: channel filter as DexDem&&AexAem (colocalisation of the donor dye signal upon donor laser excitation, the acceptor dye signal upon acceptor laser excitation), a width limit between the donor and the acceptor between 1 and 2 pixel, a nearest-neighbour limit of 6 pixels, and signal averaging from the frames 2–40. In the case of real-time RP_O complex formation movies, the nearest-neighbour limit was turned off and the time-window for the search of the surface-bound fluorescent molecules was set with the signal averaging setting (i.e. typically frames ~1000–3000) to the part of the movie in which most promoter binding events took place. The trajectories selected by the TwoTone-ALEX analysis were manually sorted by eliminating all traces that displayed extensive fluorophore blinking, multi-step photobleaching indicating more than one donor or acceptor dye in the same diffraction limited intensity spot, or other aberrant behaviour.

The apparent FRET efficiency (E^*) was calculated using Eq. (1) where I_{DD} and I_{DA} are the emission intensities of the donor (Cy3B) and acceptor (ATTO647N) dyes upon donor excitation (532 nm), respectively.[48]

$$E^* = \frac{I_{DA}}{I_{DA} + I_{DD}} \quad (1)$$

The trajectories were analysed using a modified version of the hidden Markov model ebFRET software.[49,29] The trajectories from the pre-formed RP_O or post-binding movies were fit to 2-state HMM model followed by noise filtering by requiring an accepted dwell time to satisfy the criteria that the step (i.e., change in E^*) is separated from the subsequent step by more than 3-fold the Allan deviation.[50,29] The trajectories were then classified into dynamic or static populations depending whether they displayed >2 or ≤2 accepted E^* transitions, respectively. The dwell times were extracted from the dynamic trajectories to compile a dwell time distribution. The dwells with undefined length, i.e. the first and last dwell, were discarded at this point.

The trajectories from the real-time RP_O formation movies were analysed separately for the first transcription bubble opening event, i.e., the $RP_C \rightarrow RP_O$ transition, and transcription bubble dynamics after the RP_O formation. The latter analysis was identical to the case of pre-formed RP_O complexes with the exception that the $RP_C \rightarrow RP_O$ transition at the beginning of the trajectory was trimmed away before HMM. In contrast, the analysis of the $RP_C \rightarrow RP_O$ transition in the trajectories was performed after trimming away possible bubble dynamics subsequent to the first transcription bubble opening event. We fit the first bubble opening trajectories using 2-state HMM, i.e., $RP_C \rightarrow RP_O$ mechanism, and 3-state HMM, i.e., $RP_C \rightarrow RP_i \rightarrow RP_O$ mechanism. The initial fits were filtered by requiring true state transitions to be at least 2-fold the Allan deviation.[50,29] Selection of the more complex 3-state model for the trajectory also required that both the HMM lower bound value and Aikake information criteria, calculated as previously described,[51] favoured this model. The dwell times in the RP_C and RP_i states were compiled to separate dwell time distributions. The lifetime of the RP_C state was determined by fitting the dwell time distribution to the mono-exponential decay function using Origin software (OriginLab Corporation, MA, USA).

We validated the above analysis procedure for its accuracy to detect the RP_i state and dynamics RP_O's. To this end, we used DeepFRET software[52] to simulate FRET trajectories for each state RP_C, RP_i and RP_O using FRET efficiency and FRET noise levels extracted from the experimental trajectories. Specifically, the FRET efficiency was 0.196 (noise 0.05), 0.318 (0.06) and 0.448 (0.05) for the simulated RP_C, RP_i and RP_O state, respectively. The noise of the corresponding experimental FRET data was 20–28% smaller, i.e., 0.036, 0.047 and 0.040 for the RP_C, RP_i and RP_O state, respectively. The complete trajectories for a $RP_C \rightarrow RP_i \rightarrow RP_O$ mechanism, were then assembled from the state-specific simulation by using a custom Python script and mono-exponential distribution of state dwell lengths (as in experimental data). The trajectories to determine the detection efficiency of dynamic RP_O complexes, i.e., complexes showing $RP_{ISO} \leftrightarrow RP_O$ dynamics, were simulated by DeepFRET using FRET efficiency setting 0.279 (noise 0.06) and 0.462 (noise 0.06) for the RP_{ISO} and RP_O state, respectively. The mean lifetime of the RP_{ISO} and RP_O state complex was set as 0.085 s and 0.56 s, respectively. The trajectory length was 2–7 s (100–350 frames).

The histograms of E^* values were fit in Origin software to one or two Gaussian distributions using Eq. (2) with n fixed as 1 or 2, respectively. The fit parameters E_c^*, w and A are the centre, width and area of the Gaussian distributions, respectively.

$$y = \sum_0^n \frac{A}{w\sqrt{\pi/2}} e^{-2\frac{(E^* - E_c^*)^2}{w^2}} \quad (2)$$

CRediT authorship contribution statement. **Anssi M. Malinen:** Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Funding acquisition. **Jacob Bakermans:** Software, Investigation, Writing – review & editing. **Emil Aalto-Setälä:** Software, Investigation.

Martin Blessing: Resources, Writing – review & editing. **David L.V. Bauer:** Resources, Writing – review & editing, Funding acquisition. **Olena Parilova:** Investigation, Funding acquisition. **Georgiy A. Belogurov:** Resources, Writing – review & editing. **David Dulin:** Conceptualization, Formal analysis, Writing – review & editing. **Achillefs N. Kapanidis:** Conceptualization, Methodology, Writing – review & editing, Funding acquisition, Supervision.

Acknowledgements

This research was funded in whole, or in part, by the Wellcome Trust [110164/Z/15/Z; FC011104]. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission. We thank Richard E. Ebright for providing myxopyronin B. We are grateful to Matti Turtola, Thadée Grocholski and Henri Malmi for constructing plasmids. We thank Abhishek Mazumder for insightful discussions and critical reading of the manuscript. We thank the Cell Imaging and Cytometry core at Turku Bioscience Centre, which is supported by Biocenter Finland, for the access to the microscopes.

Data availability

All our time-trace data and software used for their analysis will be available from the authors upon reasonable request.

Funding

This work was supported by Academy of Finland [grant numbers 307775, 314100, 335377 to A.M. M]; Instrumentarium Science Foundation [grant to A.M.M.]; Finnish Cultural Foundation [grants to A. M.M. and O.P.]; the Francis Crick Institute which receives its core funding from Cancer Research UK (FC011104), the UK Medical Research Council (FC011104), and the Wellcome Trust (FC011104) [to D.L.V.B.]; Wellcome Trust grant [grant number 110164/Z/15/Z to A.N.K.]; and UK Biotechnology and Biological Sciences Research Council [grant number BB/H01795X/1 to A.N.K].

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmb.2021.167383.

Received 7 October 2021;
Accepted 25 November 2021;
Available online 1 December 2021

Keywords:
transcription initiation;
molecular mechanism;
reaction pathway;
σ factor;
total internal reflection fluorescence microscopy

Abbreviations:
ALEX, Alternating laser excitation; αCTD, C-terminal domain of RNA polymerase α-subunit; CI, confidence interval; Cor, corallopyronin A; cryo-EM, cryo-electron microscopy; E^*, apparent FRET efficiency; FRET, fluorescence energy transfer; GL, gate loop in the RNAP β subunit; HMM, Hidden Markov modelling; LL, lid loop in the RNAP β′ subunit; Myx, myxopyronin B; RL, rudder loop in the RNAP β′ subunit; RNAP, RNA polymerase; RP_C, RNAP–promoter closed complex; RP_i, RNAP–promoter intermediate complex; RP_O, RNAP–promoter open complex; *lacCONS*, consensus *lac* promoter; dsLC2, double-stranded consensus *lac* promoter; pmLC2, pre-melted consensus *lac* promoter; ntDNA, non-template DNA; tDNA, template DNA

References

1. Feklístov, A., Sharon, B.D., Darst, S.A., Gross, C.A., (2014). Bacterial Sigma Factors: A Historical, Structural, and Genomic Perspective. *Annu. Rev. Microbiol.* **68**, 357–376.
2. Ruff, E.F., Record, M.T., Artsimovitch, I., (2015). Initial events in bacterial transcription initiation. *Biomolecules* **5**, 1035–1062.
3. Campbell, E.A., Korzheva, N., Mustaev, A., Murakami, K., Nair, S., Goldfarb, A., Darst, S.A., (2001). Structural mechanism for rifampicin inhibition of bacterial RNA polymerase. *Cell* **104**, 901–912.
4. Feklistov, A., Mekler, V., Jiang, Q., Westblade, L.F., Irschik, H., Jansen, R., Mustaev, A., Darst, S.A., et al., (2008). Rifamycins do not function by allosteric modulation of binding of Mg^{2+} to the RNA polymerase active center. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 14820–14825.
5. Lin, W., Das, K., Degen, D., Mazumder, A., Duchi, D., Wang, D., Ebright, Y.W., Ebright, R.Y., et al., (2018). Structural Basis of Transcription Inhibition by Fidaxomicin (Lipiarmycin A3). *Mol. Cell* **70**, 60–71.e15.
6. Ho, M.X., Hudson, B.P., Das, K., Arnold, E., Ebright, R.H., (2009). Structures of RNA polymerase- antibiotic complexes. *Curr. Opin. Struct. Biol.* **19**, 715–723.
7. Zuo, Y., Steitz, T.A., (2015). Crystal Structures of the *E. coli* Transcription Initiation Complexes with a Complete Bubble. *Mol. Cell* **58**, 534–540.

8. Bae, B., Feklistov, A., Lass-Napiorkowska, A., Landick, R., Darst, S.A., (2015). Structure of a bacterial RNA polymerase holoenzyme open promoter complex. *Elife*, e08504. Article 4.
9. Boyaci, H., Chen, J., Jansen, R., Darst, S.A., Campbell, E. A., (2019). Structures of an RNA polymerase promoter melting intermediate elucidate DNA unwinding. *Nature* **565**, 382–385.
10. Chen, J., Chiu, C., Gopalkrishnan, S., Chen, A.Y., Olinares, P.D.B., Saecker, R.M., Winkelman, J.T., Maloney, M.F., et al., (2020). Stepwise Promoter Melting by Bacterial RNA Polymerase. *Mol. Cell* **78**, 275–288.e6.
11. Mazumder, A., Kapanidis, A.N., (2019). Recent Advances in Understanding σ70-Dependent Transcription Initiation Mechanisms. *J. Mol. Biol.* **431**, 3947–3959.
12. Craig, M.L., Tsodikov, O.V., McQuade, K.L., Schlax, P.E., Capp, M.W., Saecker, R.M., Record, M.T., (1998). DNA footprints of the two kinetically significant intermediates in formation of an RNA polymerase-promoter open complex: Evidence that interactions with start site and downstream DNA induce sequential conformational changes in polymerase and DNA. *J. Mol. Biol.* **283**, 741–756.
13. Sclavi, B., Zaychikov, E., Rogozina, A., Walther, F., Buckle, M., Heumann, H., (2005). Real-time characterization of intermediates in the pathway to open complex formation by *Escherichia coli* RNA polymerase at the T7A1 promoter. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 4706–4711.
14. Davis, C.A., Capp, M.W., Record, M.T., Saecker, R.M., (2005). The effects of upstream DNA on open complex formation by *Escherichia coli* RNA polymerase. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 285–290.
15. Rogozina, A., Zaychikov, E., Buckle, M., Heumann, H., Sclavi, B., (2009). DNA melting by RNA polymerase at the T7A1 promoter precedes the rate-limiting step at 37 °C and results in the accumulation of an off-pathway intermediate. *Nucleic Acids Res.* **37**, 5390–5404.
16. Mangiarotti, L., Cellai, S., Ross, W., Bustamante, C., Rivetti, C., (2009). Sequence-Dependent Upstream DNA–RNA Polymerase Interactions in the Open Complex with λPR and λPRM Promoters and Implications for the Mechanism of Promoter Interference. *J. Mol. Biol.* **385**, 748–760.
17. Sreenivasan, R., Heitkamp, S., Chhabra, M., Saecker, R., Lingeman, E., Poulos, M., McCaslin, D., Capp, M.W., et al., (2016). Fluorescence Resonance Energy Transfer Characterization of DNA Wrapping in Closed and Open Escherichia coli RNA Polymerase−λPR Promoter Complexes. *Biochemistry* **55**, 2174–2186.
18. Estrem, S.T., Gaal, T., Ross, W., Gourse, R.L., (1998). Identification of an UP element consensus sequence for bacterial promoters. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 9761–9766.
19. Ross, W., Gourse, R.L., (2005). Sequence-independent upstream DNA-alphaCTD interactions strongly stimulate *Escherichia coli* RNA polymerase-*lacUV5* promoter association. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 291–296.
20. Cook, V.M., DeHaseth, P.L., (2007). Strand opening-deficient Escherichia coli RNA polymerase facilitates investigation of closed complexes with promoter DNA: Effects of DNA sequence and temperature. *J. Biol. Chem.* **282**, 21319–21326.
21. Feklistov, A., Darst, S.A., (2011). Structural Basis for Promoter −10 Element Recognition by the Bacterial RNA Polymerase σ Subunit. *Cell* **147**, 1257–1269.
22. Kubori, T., Shimamoto, N., (1996). A Branched Pathway in the Early Stage of Transcription by *Escherichia coli* RNA Polymerase. *J. Mol. Biol.* **256**, 449–457.
23. Ko, J., Heyduk, T., (2014). Kinetics of promoter escape by bacterial RNA polymerase: effects of promoter contacts and transcription bubble collapse. *Biochem. J.* **463**, 135–144.
24. Duchi, D., Bauer, D.L.V., Fernandez, L., Evans, G., Robb, N., Hwang, L.C., Gryte, K., Tomescu, A., et al., (2016). RNA Polymerase Pausing during Initial Transcription. *Mol. Cell* **63**, 939–950.
25. Henderson, K.L., Felth, L.C., Molzahn, C.M., Shkel, I., Wang, S., Chhabra, M., Ruff, E.F., Bieter, L., et al., (2017). Mechanism of transcription initiation and promoter escape by *E. coli* RNA polymerase. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E3032–E3040.
26. Hsu, L.M., (2002). Promoter clearance and escape in prokaryotes. *Biochim. Biophys. Acta - Gene Struct. Expr.* **1577**, 191–207.
27. Susa, M., Kubori, T., Shimamoto, N., (2006). A pathway branching in transcription initiation in *Escherichia coli*. *Mol. Microbiol.* **59**, 1807–1817.
28. Lerner, E., Chung, S.Y., Allen, B.L., Wang, S., Lee, J., Lu, S.W., Grimaud, L.W., Ingargiola, A., et al., (2016). Backtracked and paused transcription initiation intermediate of *Escherichia coli* RNA polymerase. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E6562–E6571.
29. Dulin, D., Bauer, D.L.V., Malinen, A.M., Bakermans, J.J. W., Kaller, M., Morichaud, Z., Petushkov, I., Depken, M., et al., (2018). Pausing controls branching between productive and non-productive pathways during initial transcription in bacteria. *Nature Commun.* **9** Article 1478.
30. Sohn, B.K., Basu, U., Lee, S.W., Cho, H., Shen, J., Deshpande, A., Johnson, L.C., Das, K., et al., (2020). The dynamic landscape of transcription initiation in yeast mitochondria. *Nature Commun.* **11**, 1–12.
31. Kapanidis, A.N., Margeat, E., Ho, S.O., Kortkhonjia, E., Weiss, S., Ebright, R.H., (2006). Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science* **314**, 1144–1147.
32. Robb, N.C., Cordes, T., Hwang, L.C., Gryte, K., Duchi, D., Craggs, T.D., Santoso, Y., Weiss, S., et al., (2013). 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.
33. Duchi, D., Gryte, K., Robb, N.C., Morichaud, Z., Sheppard, C., Brodolin, K., Wigneshweraraj, S., Kapanidis, A.N., (2018). Conformational heterogeneity and bubble dynamics in single bacterial transcription initiation complexes. *Nucleic Acids Res.* **46**, 677–688.
34. Clopper, C.J., Pearson, E.S., (1934). The use of confidence or fiducial limits illustrated in the case of binomial. *Biometrika* **26**, 404–413.
35. Feklistov, A., Bae, B., Hauver, J., Lass-Napiorkowska, A., Kalesse, M., Glaus, F., Altmann, K.-H., Heyduk, T., et al., (2017). RNA polymerase motions during promoter melting. *Science* **356**, 863–866.
36. Mazumder, A., Ebright, R.H., Kapanidis, A.N., (2021). Transcription initiation at a consensus bacterial promoter

proceeds via a 'bind-unwind-load-and-lock' mechanism. *eLife* **10**, e70090. https://doi.org/10.7554/eLife.70090.
37. Belogurov, G.A., Vassylyeva, M.N., Sevostyanova, A., Appleman, J.R., Xiang, A.X., Lira, R., Webber, S.E., Klyuyev, S., et al., (2009). Transcription inactivation through local refolding of the RNA polymerase structure. *Nature* **457**, 332–335.
38. Winkelman, J.T., Vvedenskaya, I.O., Zhang, Y., Zhang, Y., Bird, J.G., Taylor, D.M., Gourse, R.L., Ebright, R.H., et al., (2016). Multiplexed protein-DNA cross-linking: Scrunching in transcription start site selection. *Science* **351**, 1090–1093.
39. Plaskon, D.M., Henderson, K.L., Felth, L.C., Molzahn, C. M., Evensen, C., Dyke, S., Shkel, I.A., Record, M.T., (2021). Temperature effects on RNA polymerase initiation kinetics reveal which open complex initiates and that bubble collapse is stepwise. *Proc. Natl. Acad. Sci. U. S. A.* **118** Article 2021941118.
40. Saecker, R.M., Chen, J., Chiu, C.E., Malone, B., Sotiris, J., Ebrahim, M., Yen, L.Y., Eng, E.T., Darst, S.A., (2021). Structural origins of Escherichia coli RNA polymerase open promoter complex stability. *Proc. Natl. Acad. Sci. U. S. A.* **118** Article 2112877118.
41. Estrem, S.T., Ross, W., Gaal, T., Chen, Z.W., Niu, W., Ebright, R.H., Gourse, R.L., (1999). Bacterial promoter architecture: subsite structure of UP elements and interactions with the carboxy-terminal domain of the RNA polymerase alpha subunit. *Genes Dev.* **13**, 2134–2147.
42. Svetlov, V., Artsimovitch, I., (2015). Purification of bacterial RNA polymerase: Tools and protocols. In: Artsimovitch, I., Santangelo, T.J. (Eds.), *Bacterial Transcriptional Control: Methods and Protocols*. Springer, New York, pp. 13–29.
43. Belogurov, G.A., Vassylyeva, M.N., Svetlov, V., Klyuyev, S., Grishin, N.V., Vassylyev, D.G.G., Artsimovitch, I., (2007). Structural Basis for Converting a General Transcription Factor into an Operon-Specific Virulence Regulator. *Mol. Cell* **26**, 117–129.
44. Turtola, M., Belogurov, G.A., (2016). NusG inhibits RNA polymerase backtracking by stabilizing the minimal transcription bubble. *Elife* **5** Article e18096.
45. Kulbachinskiy, A., Mustaev, A., (2006). Region 3.2 of the σ 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.
46. Holden, S.J., Uphoff, S., Hohlbein, J., Yadin, D., Le Reste, L., Britton, O.J., Kapanidis, A.N., (2010). Defining the limits of single-molecule FRET resolution in TIRF microscopy. *Biophys. J.* **99**, 3102–3111.
47. Kapanidis, A.N., Lee, N.K., Laurence, T.A., Doose, S., Margeat, E., Weiss, S., (2004). Fluorescence-aided molecule sorting: Analysis of structure and interactions by alternating-laser excitation of single molecules. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 8936–8941.
48. Ha, T., Enderle, T., Ogletree, D.F., Chemla, D.S., Selvin, P. R., Weiss, S., (1996). Probing the interaction between two single molecules: Fluorescence resonance energy transfer between a single donor and a single acceptor. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 6264–6268.
49. Van De Meent, J.W., Bronson, J.E., Wiggins, C.H., Gonzalez, R.L., (2014). Empirical Bayes methods enable advanced population-level analyses of single-molecule FRET experiments. *Biophys. J.* **106**, 1327–1337.
50. Dulin, D., Cui, T.J., Cnossen, J., Docter, M.W., Lipfert, J., Dekker, N.H., (2015). High Spatiotemporal-Resolution Magnetic Tweezers: Calibration and Applications for DNA Dynamics. *Biophys. J.* **109**, 2113–2125.
51. McKinney, S.A., Joo, C., Ha, T., (2006). Analysis of single-molecule FRET trajectories using hidden Markov modeling. *Biophys. J.* **91**, 1941–1951.
52. Thomsen, J., Sletfjerding, M.B., Jensen, S.B., Stella, S., Paul, B., Malle, M.G., Montoya, G., Petersen, T.C., Hatzakis, N.S., (2020). DeepFRET, a software for rapid and automated single-molecule FRET data classification using deep learning. *eLife* **9** Article e60404.