trajectory data reveals that ~40% of molecules adopt a stable E*~0.35 state compared with ~10% for wt complexes. These results indicate that the partially open bubble conformation is more readily adopted by ΔJaw complexes compared with wt complexes. The fact that the size of the partially open population can be modulated by changes to the RNAP structure, lends support to our hypothesis that the E*~0.35 state is a true independent structural intermediate.

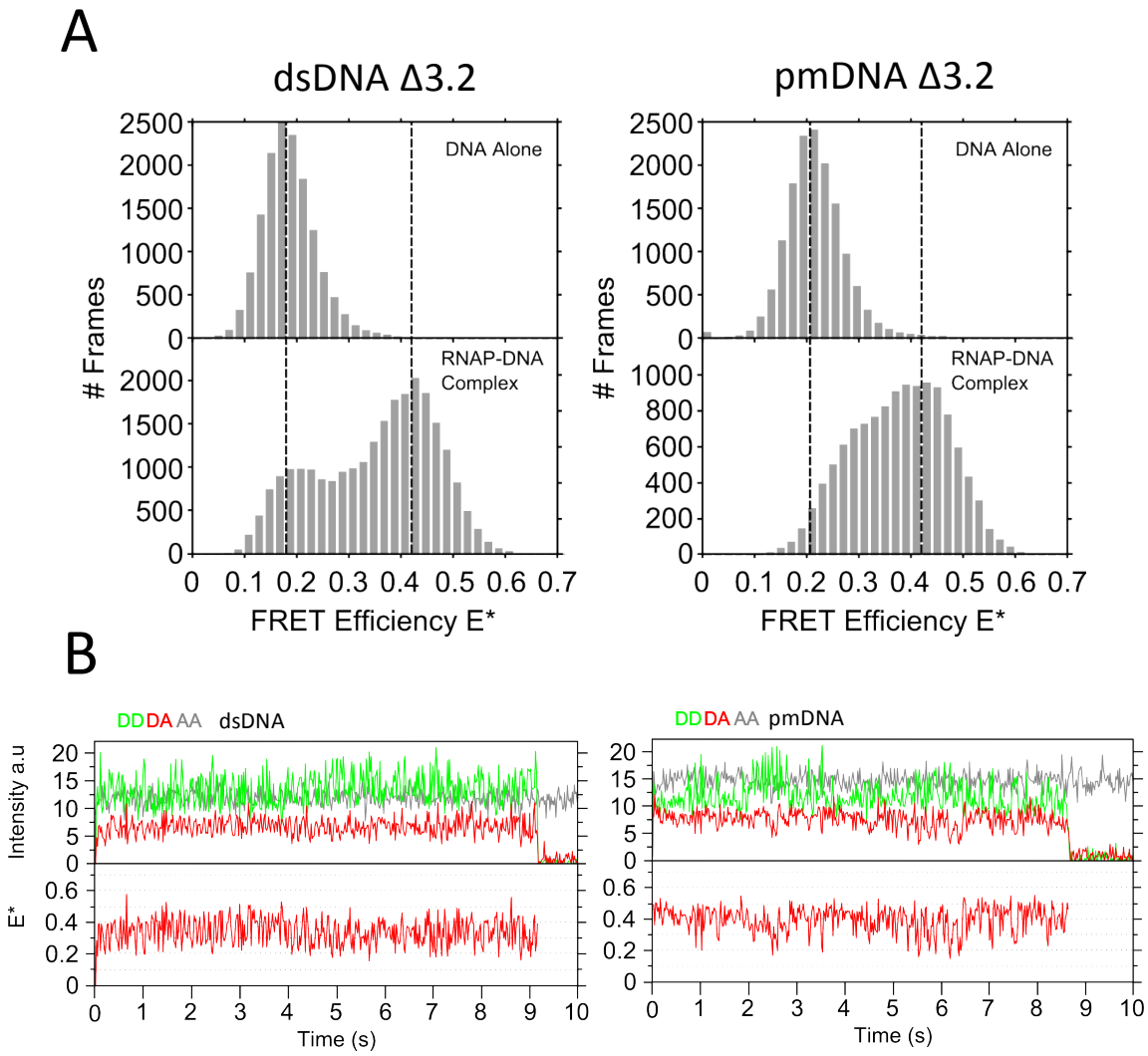

Figure 5.13. σR3.2 influences bubble dynamics. (A) Stacked E* histograms of DNA alone (top) and of RNAP-promoter DNA complexes (bottom). All histograms show data for 100 molecules. The left hand figure shows dsDNA data, the right hand figure shows pmDNA data. Vertical dashed lines indicate the approximate peak positions of the closed- and open-bubble populations. (B) Example time trajectories of Δ3.2 mutant complexes exhibiting dynamic transitions between open and closed bubble conformations. The left hand figure shows dsDNA data, the right hand figure shows pmDNA data. Temporal resolution is 20 ms.

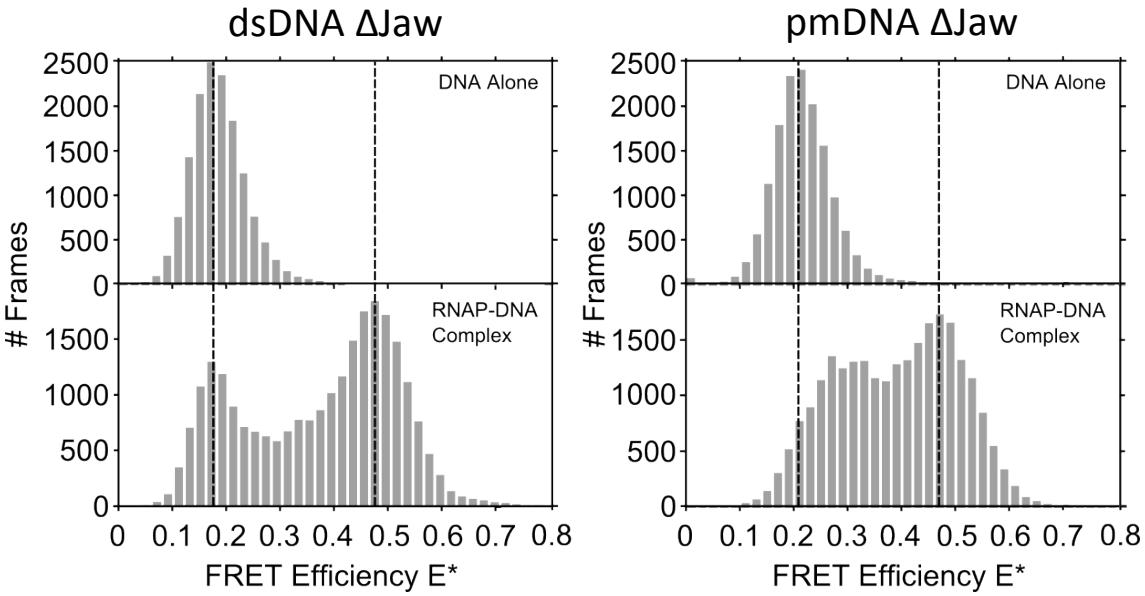

Figure 5.14. RNAP β' Jaw influences the transcription bubble conformation. Stacked E* histograms of DNA alone (top) and of RNAP-DNA complexes (bottom). The left hand figure was acquired using dsDNA, the right hand figure was acquired using pmDNA. Vertical dashed lines indicate the approximate peak positions of the closed- and open-bubble populations. Temporal resolution is 20 ms[8]. All histograms show data for 100 molecules.

5.3.5 Kinetic analysis of transcription bubble dynamics

To gain a more quantitative comparison of the differences in bubble dynamics between wt, Δ3.2, and ΔJaw RNAP-DNA complexes, we quantified the bubble opening and closing rates by applying Hidden Markov Modelling (HMM) to our data to create dwell time distributions of the closed and open bubble states. The distributions were fitted with exponential functions (see *Materials and Methods*) to extract the bubble opening and closing rates (Fig. 5.15). For RNAP-promoter dsDNA complexes we found that the closed bubble state exhibited a mean dwell time of ~150 ms. By contrast, the bubble closing transition was found to follow biexponential decay kinetics, and the open bubble state exhibited mean dwell times of ~220 ms and ~1 s. The biexponential fit suggests that there exist two subpopulations of complexes with different bubble closing rates. However, it is unclear what the biological basis for this difference is. In future studies it

[8] Example time trajectories are not shown because they shown no difference compared with wt RNAP time trajectories. Example time trajectories of dynamic ΔJaw RNAP-DNA complexes can be found in Fig. 5.15C.

will be interesting to perform a robust statistical comparison of the kinetic parameters of individual molecules to test whether individual molecules within the overall population of molecules behave differently i.e. exhibit different bubble dynamics kinetic parameters (see Chapter 8 for further discussion on future studies). For all conditions tested, the bubble mismatch was found to slightly increase the mean dwell-time of the closed bubble conformation, and to have no major effect on the open bubble dwell-times. This may reflect the fact that the closed bubble conformation is kinetically more unstable in RNAP-dsDNA complexes compared with RNAP-pmDNA complexes. It is also possible that the mismatch decreases the energetic barrier to bubble opening and produces a situation in which complexes are less likely to adopt a stable closed bubble conformation and instead exhibit dynamic behaviour. Our HMM analysis showed that deleting the jaw subdomain slightly decreases the average closed bubble dwell-time (Fig. 5.15D), and reduces the size of the slow exponential contribution to the open bubble dwell time histogram (Fig. 5.15E). This suggests that the stability of the open bubble conformation has been reduced, although it should be noted that the effect is marginal. The effect of deleting the σR3.2 finger was much more apparent; Δ3.2 dwell histograms in Fig. 5.15D–E showed that the bubble opening and closing rates increase considerably for Δ3.2 complexes compared with wt complexes. This finding supports the hypothesis that the σR3.2 finger influences the position of the template strand at the active centre (26, 72). In addition, the presence of the dinucleotide, ApA, which stabilises RP_O (6), was found to have little effect on the kinetics of the transcription bubble dynamics (Fig. 5.16).

5.4 Discussion

5.4.1 The transcription bubble exhibits static heterogeneity

By specifically immobilizing pre-formed RNAP-promoter DNA complexes on a surface and using TIRF microscopy to monitor individual molecules for extended observation periods, we were able to identify considerable transcription bubble inter-molecular

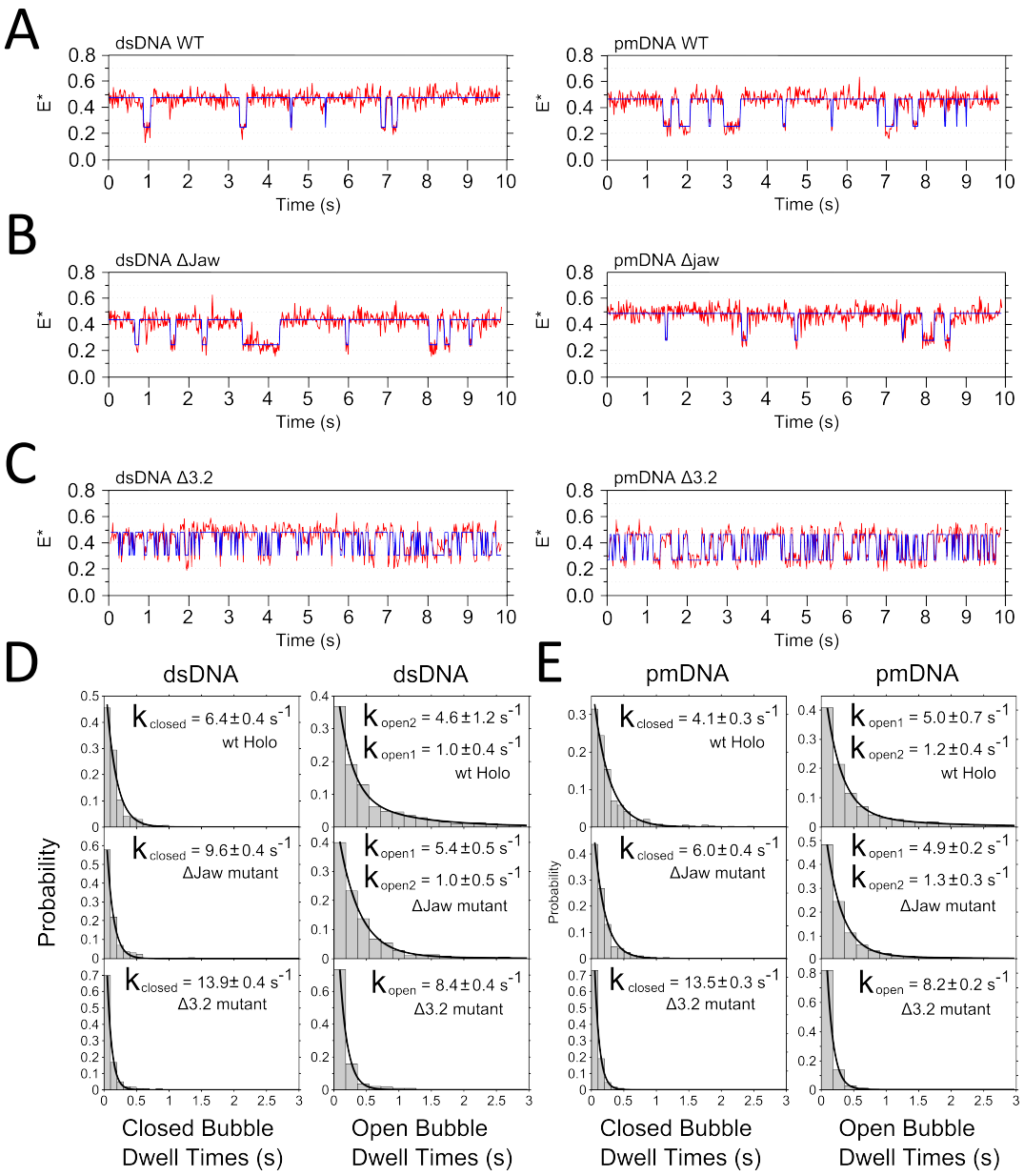

Figure 5.15. Kinetics of transcription bubble dynamics. Example time trajectories of (A) wt RNAP-DNA complexes, (B) ΔJaw mutant RNAP-DNA complexes, and (C) Δ3.2 mutant RNAP-DNA complexes displaying bubble opening and closing transitions. The data has been fit to a Hidden Markov Model (blue line). The left hand figures show dsDNA data, the right hand figures show pmDNA data. (D) Stacked histograms of closed and open bubble conformation dwells for dsDNA. The bubble opening rates are shown inset. 852 events from 80 molecules are shown in the top panels, 692 events from 71 molecules are shown in the middle panels, 924 events from 85 molecules are shown in the bottom panels. (E) Stacked histograms of closed and open bubble conformation dwells for pmDNA. The bubble closing rates are shown inset. 1165 events from 87 molecules are shown in the top panels, 2001 events from 101 molecules are shown in the middle panels, 2082 events from 105 molecules are shown in the bottom panels. For figures D and E, wt data is shown in the top panel, ΔJaw mutant data in the middle panel, Δ 3.2 mutant data in the bottom panel, and all data are fit to exponential functions (see

Materials and Methods). Temporal resolution is 20 ms. Log-linear plots of dwell time distributions are given in Fig. C1 in Appendix C.

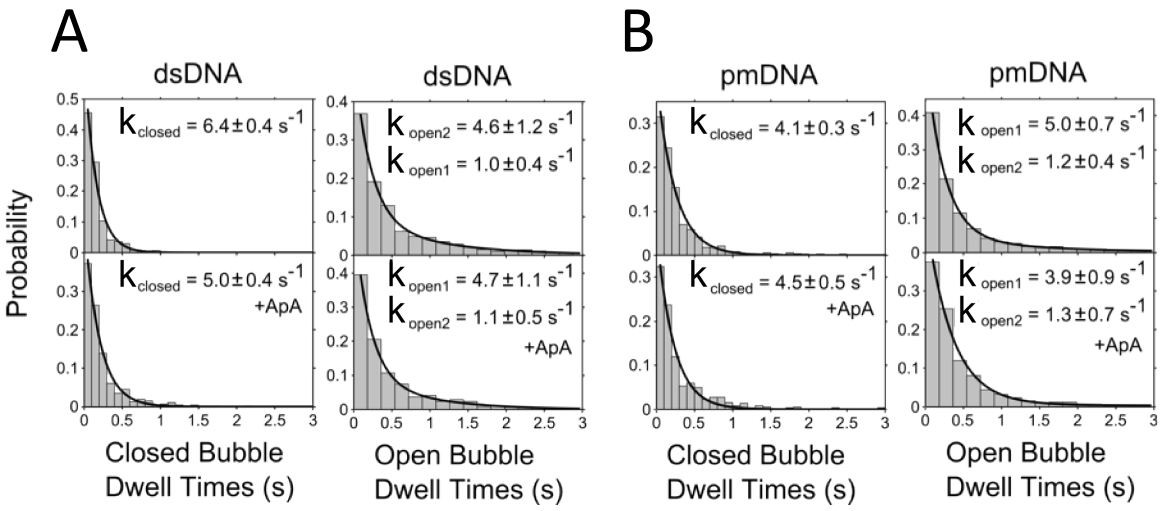

Figure 5.16. ApA has little effect on transcription bubble dynamics. (A) Stacked histograms of closed bubble conformation dwells for dsDNA and pmDNA. The data is fit to single exponential functions. The bubble opening rates are shown inset. 852 events from 80 molecules are shown in the top panels, 793 events from 78 molecules are shown in the bottom panels. (B) Stacked histograms of open bubble conformation dwells for dsDNA and pmDNA. The data is fit to double exponential functions. The bubble closing rates are shown inset. wt holoenzyme data was used to generate all figures. 1165 events from 87 molecules are shown in the top panels, 1111 events from 94 molecules are shown in the bottom panels. Bottom panel shows data for wt holoenzyme in the presence of 2 mM ApA. Log-linear plots of the distributions are given in Fig. C2 in Appendix C.

heterogeneity. Our results revealed three RNAP-promoter DNA subpopulations: one in which the transcription bubble is closed, another in which the bubble is open, and a third in which the bubble is neither fully open nor fully closed (Fig. 5.17). Our data also showed a subpopulation of complexes that undergo large-scale conformational changes between states. Experiments assessing targeted RNAP structural element deletions provided insight into the function of specific subdomains in transcription bubble opening.

Our results are broadly in agreement with previous studies showing that varying conditions such as temperature can trap distinct structural intermediates in the RP_O formation pathway (76, 77, 193). The predominant consensus that emerged from previous work is a general three-step RP_O formation mechanism via intermediates whose exact structural details vary with the promoter. We want to note two

intermediates of particular interest. First, we note the RP_i intermediate reported in (77) which had three key characteristics: 1) increased abundance below 22°C, 2) heparin resistance, and 3) the bubble in a closed conformation. We highlight that the E*~0.20 subpopulation we identify shares similar characteristics to RP_i. Second, we note the intermediate O_l identified at the *lac*UV5 promoter in (194, 195). The degree of promoter unwinding in this intermediate was found to differ from that of the O_h complex, which was the more abundant species at 37°C and presumably has a fully open transcription bubble. We speculate that the O_l intermediate may correspond to our E*~0.35 species, which we propose has a partially open or nucleated transcription bubble. We refer to the E*=0.35 state as RP_{i2}. To our knowledge, our work is the first to identify three distinct heparin-resistant RNAP-Promoter DNA complexes existing as a heterogeneous mixture at the *lac* promoter.

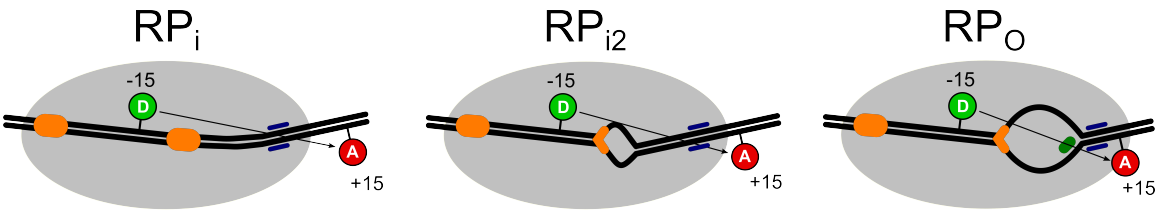

Figure 5.17. Schematic representation of transcription bubble conformations inferred from our studies. Leftmost figure shows an RNAP-DNA complex with the bubble in a closed conformation. We have assigned this species as RP_i given its similar features to RP_i in (77). The middle figure is assigned RP_{i2} and has the transcription bubble partially open or nucleated. The rightmost figure is assigned RP_O and has the transcription bubble in a fully open conformation. It should be noted that the relative enzymatic activities of the three complexes is still an open question. We therefore assign the right-most figure to RP_O because it has an open transcription bubble conformation, not because it is considered to be the only catalytically competent complex of the three[9]. We speculate that the downstream DNA is loaded in the binding cleft (blue lines) in all three states[10].

[9] It is important to make this distinction as one way in which RP_O has been defined in the past is by its 'catalytic competence'.

[10] We make this assumption because: 1) the three states are resistant to heparin, and 2) because the clamp adopts a closed conformation in immobilized RNAP-DNA complexes formed using the same promoter DNA (see chapter 7). Further investigations are needed to test the validity of this assumption.

5.4.2 The transcription bubble exhibits dynamic heterogeneity

We observed molecules that undergo large-scale conformational transitions in a manner similar to the mitochondrial and eukaryotic transcription initiation machinery (176, 177). Surprisingly, our results showed that bubble transitions persist and, moreover, increase when we form complexes using pre-melted DNA. Previous studies showed that the presence of fluorophore labels did not affect the transcription reaction (175, 188). These combined observations lead us to speculate that the dynamic behaviour is not an artifact due to deficient transcription bubble opening or some other factor, but rather is an intrinsic feature of RNAP-promoter DNA complexes that may play a functional role in initiation.

The transcription bubble structural dynamics has implications in addressing a long-standing question, which is, at what step does promoter unwinding and loading into the DNA binding cleft occur? Two models currently exist: the bend-load-open model and open-bend-load model (see chapter 2) (66, 68). In the former, full DNA unwinding occurs after the duplex DNA has entered the DNA binding cleft and is promoted by interactions between the clamp and downstream DNA (76). In the latter, DNA unwinding must occur before DNA enters the binding cleft (74). Our work shows that the transcription bubble readily undergoes conformational fluctuations once a stable complex has formed. Furthermore, TIRF experiments investigating the RP_O clamp conformation show that the clamp remains closed even when the bubble is expected to undergo opening and closing transitions (see chapter 7). This indicates that bubble opening and closing is independent of the clamp conformation, and presumably occurs when promoter DNA is located inside the DNA binding cleft and the complex is resistant to heparin. Our observations therefore support a model in which DNA melting does not need to occur prior to loading into the binding cleft. We note however, that our experiments do not monitor the clamp and bubble conformation from the onset of initial binding of RNAP to promoter DNA; it is possible that the exact sequence of events is different in this situation.

The observation that Δ3.2 RNAP-promoter DNA complexes exhibit an increased rate of bubble dynamics supports proposals based on structural studies that the σR3.2 finger plays an important role in stabilising the template strand at the active site, and also raises the intriguing possibility that the altered opening and closing kinetics might influence later steps in initiation. Our results do not support a model in which the β' jaw plays an important role in influencing bubble dynamics; however, they do suggest that the jaw domain influences the degree of transcription bubble opening, which is in agreement with the results of an earlier study (41).

5.4.3 Possible functions of the static and dynamic heterogeneity

What is the functional significance of the RNAP-promoter DNA complex heterogeneity? One possibility is that the intermediate states identified in our work may influence gene expression by affecting later stages of transcription initiation. For example, different bubble conformations could influence promoter escape in a manner similar to that reported in (193) where an O_l intermediate was found to produce more abortive products and be less efficient at promoter escape than an O_u complex. In this scenario, if the transcription initiation pathway were to stall at either RP_i or RP_{i2}, full-length RNA synthesis and therefore gene expression could be affected. In fact, we consider it possible that one or both of these intermediates may correspond to moribund complexes reported in (100) and discussed in chapter 2. Additional work is required to test these possibilities.

What role might transcription bubble dynamics play in initiation? The fact that the initiating dinucleotide ApA – which is expected to stabilise the template strand in one particular register at the active site (175) – does not influence bubble opening and closing kinetics, suggests that the bubble structural transitions do not reflect attempts by RNAP to place the template strand at the 'correct' start site. This observation allows us to propose alternative possibilities. In (195), RNAP-Promoter DNA complexes were found to be more likely to escape the promoter if interactions between the polymerase and promoter DNA were weak. We suggest that if the dynamic molecules we observe

have weaker RNAP-promoter DNA interactions compared with molecules with stable open conformations, they may have an increased propensity to break promoter contacts and escape to elongation. This is an intriguing possibility, however, more work remains to be done in order to determine what roles the observed static and dynamic heterogeneity play in transcription initiation.

5.5 Conclusions

The work presented in this chapter has revealed new insight into the conformation of the transcription bubble in RNAP-promoter DNA complexes, and also into the role specific RNAP structural elements play in transcription bubble melting. Our investigations have led to the identification of possible intermediates in the RP_O formation pathway, which are surprisingly stable. In the future, we intend to use our assay to monitor transcription initiation from initial DNA binding through to promoter escape and thereby confirm whether the intermediate conformational states we observe (RP_i and RP_{i2}) are *bona fide* on-pathway RP_O formation intermediates. We also intend to investigate the possible functional roles played by the dynamic subpopulation of RNAP-promoter DNA complexes.

Our work raises a number of intriguing questions regarding the biological significance of structural heterogeneity in the transcription mechanism. We hope that the observations made in this chapter will play a role in directing future investigations into the mechanism of RP_O formation.

5.6 Contributions

I performed all work described in this chapter except for the following contributions made by others: Nicole Robb (Oxford University, UK) gel-purified DNA substrates. Kristofer Gryte (Oxford University, UK) provided useful discussions in the interpretation of results.

Chapter 6: Initial transcription monitored in real-time

6.1 Introduction

Once the catalytically competent open complex (RP_O) has been formed, it can begin synthesising RNA *de novo* via productive or abortive pathways. In the productive pathway, RNAP synthesises RNA via a scrunching mechanism until the nascent RNA reaches 9–11-nt in length, at which point the polymerase escapes the promoter and transitions to elongation. This pathway results in a full length RNA transcript being synthesised. In the abortive pathway, RNAP begins synthesising a short RNA to form an RNAP-promoter initial transcribing complex (RP_{ITC}), also via a scrunching mechanism, but is unable to escape the promoter. Instead, the polymerase releases the short RNA, reverts back to RP_O, and re-initiates RNA synthesis (see chapter 2 for a detailed discussion on abortive initiation). The *lac* promoter produces abortive products up to 9-nt in length; however, the initial transcribed sequence resembles the -10 promoter element and causes promoter-proximal pausing, which results in the production of long (>9-nt) non-abortive RNA products (196).

Abortive initiation has been the focus of intense investigations for a number of years. However, since heterogeneity and asynchrony are features of initial transcription (84, 100, 109), it has been challenging to investigate the fine details of this process using classical ensemble techniques. Consequently, a number of unknowns remain in this area of research, including: 1) the lifetimes of scrunched complexes, 2) the exact mechanism of abortive RNA synthesis and release, 3) whether abortive cycling is obligatory for promoter escape, 4) how many abortive cycles RNAP performs before escaping the promoter, and 5) the exact position at which promoter escape occurs.

Single-molecule methods can allow researchers to follow multiple reactions in parallel and in real-time, and are able to directly uncover heterogeneity without the need to synchronise reactions. Single-molecule tools are therefore well suited to investigate the mechanism of initial transcription.

In earlier work, we used smFRET combined with confocal spectroscopy to show that initial transcription proceeds via a scrunching mechanism (6). In this study, a number of donor-acceptor FRET pairs were used to monitor multiple distance changes during initial transcription. The results showed that during initial transcription, positions downstream of the transcription bubble are pulled in towards the polymerase while no movements occur in regions upstream of the bubble. These observations led to the conclusion that DNA from the expanding transcription bubble is accommodated within the polymerase via a scrunching mechanism (Fig. 6.1). The DNA scrunching mechanism was also demonstrated in a separate DNA nanomanipulation study (7). However, because the smFRET study was performed using diffusing molecules, it was not possible to obtain extended time trajectories of the initial transcription reaction.

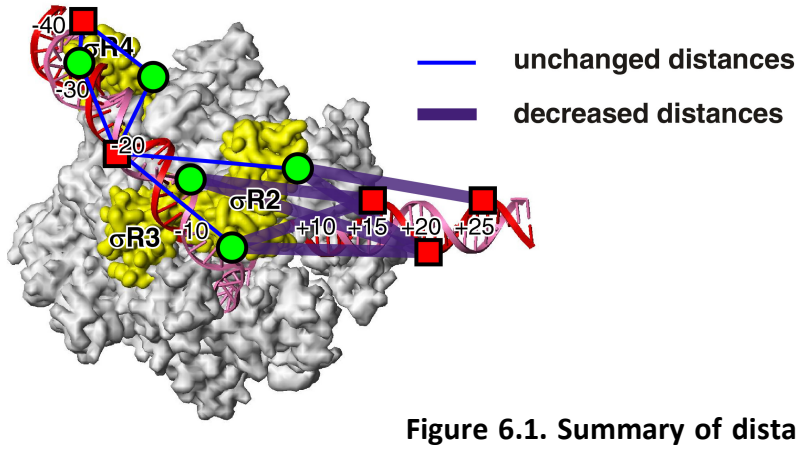

Figure 6.1. Summary of distance changes during initial transcription. Donor dye positions are shown as green spheres, acceptor dye positions are shown as red squares. Lines connecting donor and acceptor dyes indicate distances measured. Figure from (6).

In this chapter, we use smFRET combined with TIRF microscopy to monitor *de novo* RNA synthesis in real-time with the aim of addressing some of the outstanding questions mentioned above. Our smFRET assay reports on the RNA synthesis reaction by monitoring DNA scrunching, which happens concomitantly with each nucleotide incorporation event. Our results reveal heterogeneous behaviour in abortive initiation, which may account for the moribund complexes reported in the literature (84, 100, 102). We observe abortive cycles directly for the first time, and are able to gain new insight

into the function of the σR3.2 finger in abortive initiation. In the final section of this chapter, we extend our smFRET assay to monitor the promoter escape process. Our results reveal the number of abortive cycles that precede promoter escape and the apparent heterogeneity in the promoter escape position.

6.2 Materials and Methods

6.2.1 DNA and protein preparation

Oligonucleotides labelled with Cy3B and ATTO647N were purchased from IBA Life Sciences (Germany), and gel-purified by Nicole Robb. The purified single-stranded DNA (ssDNA) were annealed as described in chapter 5. DNA sequences and labelling schemes used in this chapter are shown in Fig. 6.2 and 6.4. WT RNAP core from *E. coli* with a His-tag at the C-terminal end of the β' subunit was prepared as described in chapter 5. The location of the His-tag on the RNAP structure is shown in *Appendix A*. WT RNAP core without a His-tag was purchased from Epicentre, USA. WT σ^{70} and mutant σ^{70} lacking residues 513-519 (Δ3.2) were both purified by our collaborator, Dr. Konstantin Brodolin (CPBS, France) as described in (80, 181). WT and Δ3.2 holoenzymes were prepared as described in chapter 5.

6.2.2 RP_O formation

RP_O was formed as described in chapter 5. For rifampicin experiments, 250 nM rifampicin was incubated with RNAP for 30 mins at 33°C before DNA was added. Then, we followed the RP_O formation protocol described in chapter 5.

6.2.3 *In vitro* transcription (IVT) reactions

In vitro transcription reactions were performed by Alexandra Tomescu (University of Oxford, UK) using the methods described in (175, 188) with some minor modifications

designed to mimic our smFRET experimental conditions (see section 6.2.6). Briefly, RP_O was formed as described in section 5.2.2 using KG7 buffer consisting of 40 mM HEPES-NaOH (pH 7), 100 mM potassium glutamate, 10 mM $MgCl_2$, 100 µg/mL BSA, 1 mM DTT, and 5% glycerol. *In vitro* transcription reactions were initiated by mixing 1 µL of the RP_O mixture with 4 µL of KG7 buffer supplemented with 4 units of Rnasin (Promega, USA), 0.1 mg/mL heparin, and the relevant NTP mixture. Details of the NTP mixtures are given in the relevant figure captions. In all cases however, the non-radioactive NTPs were added to a final concentration of 80 µM, and ApA was added to a final concentration of 500 µM. The radioactive NTP was either [α^{32}P]GTP or [α^{32}P]UTP (10 µCi/µL, PerkinElmer) and was added at 0.3 µCi/µL (see relevant figure captions). Reactions were incubated for 10 min at 21°C, and stopped by the addition of 5 µL of loading dye (90% formamide, 10 mM EDTA, bromophenol blue, and xylene cyanol). Mixtures were incubated for 5 min at 95°C before being loaded on a 7-M urea, 20% polyacrylamide sequencing gel and visualized by autoradiography.

6.2.4 Antibody-immobilization of RP_O

RP_O was immobilized using a biotinylated penta-His antibody (Ab; from Qiagen) as described in chapter 5.

6.2.5 Instrumentation

Single-molecule TIRF experiments were performed on the same microscope as described in chapter 4.

6.2.6 Single-molecule experiments

Once RP_O was immobilized onto the Ab-coated surface, KG7 imaging buffer consisting of 40 mM HEPES-NaOH (pH 7), 100 mM potassium glutamate, 10 mM $MgCl_2$, 1 mm DTT, 100 µg/mL BSA, 5% glycerol, 2 mM UV-treated Trolox, and an oxygen

scavenging system (described in chapter 4) was added. For abortive initiation experiments, the imaging buffer was supplemented with 500 μM ApA dinucleotide (6, 182). Where indicated, the imaging buffer was supplemented with 500 nM rifampicin.

To form RP_{ITC} engaged in the abortive synthesis of RNA products up to N-nt in length ($RP_{ITC \leq N}$), an NTP reaction mixture was added to the observation chamber manually during data acquisition by pipetting 3 μL into the well. The NTP reaction mixture varied for different experiments; however, unless stated otherwise, the final concentration of NTPs in the observation chamber was 80 μM. For $RP_{ITC \leq 4}$ experiments the NTP reaction mixture consisted of the imaging buffer supplemented with UTP. For $RP_{ITC \leq 5}$ experiments, 3'-dGTP (TriLink BioTechnologies, USA) was added to the $RP_{ITC \leq 4}$ mixture. For $RP_{ITC \leq 7}$ experiments, GTP was added to the $RP_{ITC \leq 4}$ mixture.

For promoter escape experiments, the NTP reaction mixture consisted of imaging buffer supplemented with all four NTPs. The final concentration of ATP in the observation chamber was 200 μM, while that of UTP, GTP and CTP was 100 μM. The CRIFF autofocusing system was used during data acquisition to ensure focal stability.

The temporal resolutions used for experiments are given in the relevant figure captions. For 20 ms temporal resolution experiments, excitation powers of 8 mW (532 nm laser) and 4 mW (635 nm laser) were used. For 200 ms temporal resolution experiments, excitation powers of 0.5 mW (532 nm laser) and 0.15 mW (635 nm laser) were used. All experiments were performed at 21°C.

6.2.7 Data analysis and visualisation

Fluorescence intensities were extracted from images and the uncorrected FRET efficiency (E*) calculated as described in chapter 4. We manually inspected intensity time trajectories and selected molecules for analysis according to the following criteria: 1) bleached in a single step. 2) Circular PSF across DD, DA, and AA channels. 3) No donor photoblinking (although fluctuations permitted). 4) No acceptor photoblinking although we allowed for different states (see chapter 4). 5) Fluorescence intensities within a

limited range. 6) No defocusing. 7) No nearby molecules as measured using a nearest-neighbour criterion. 50% of fluorescent spots passed all selection criteria.

E* histogram and time trajectory figures were prepared using the software package Seneca built by Kristofer Gryte. We manually inspected the intensity time trajectories of molecules in abortive initiation experiments and classified them into groups. The baseline state was assigned to RP_O and any increase in FRET that surpassed E*=0.3 was assigned to an initial transcribing complex with the transcription bubble in a scrunched conformation. We refer to these FRET states as scrunched FRET states. Time trajectories showing a stable scrunched state signal that lasted for >120 sec were classified as stably scrunched. Time trajectories showing cyclical increases and decreases in FRET between the baseline and the scrunched FRET states were classified as cycling events in which RNA is synthesised during the increase in FRET, and released as abortive RNA during the decrease in FRET. Molecules showing a stepped increase in FRET, where the step lasted for >4 frames were classified as pausing molecules. Molecules showing a decrease in FRET from a maximally scrunched FRET state (excursion state) to a scrunched state above the baseline that lasted for >4 frames were classified as backtracking molecules. Molecules that bleached within 120 sec of the addition of the NTP reaction mixture were excluded from analysis since they could not be classified as either stably scrunched or cycling. The maximum FRET value reached in an $RP_{ITC\leq N}$ experiment was assigned as the FRET value corresponding to RNA molecules N-nt in length (see *Results*).

For promoter escape experiments, we identified specific patterns of signal changes that we assigned to promoter escape events; these signal changes are described in *Results*. For both abortive initiation and promoter escape experiments, molecules exhibiting confounding photophysical fluctuations were excluded from analysis.

6.2.8 HMM Analysis of cycling smFRET time trajectories

HMM analysis was performed on time trajectories displaying dynamic FRET

fluctuations using custom-written MATLAB software described in chapter 5. For abortive initiation experiments, each time trajectory was fitted with 1 to 3 states and the best model (number of states) was selected automatically using maximum evidence criteria (184, 185). The FRET efficiency states above the baseline were classified as scrunched states. The dwell times of the scrunched states were extracted and used to generate dwell time distributions, which were fitted with an exponential decay curve as described in chapter 5. The mean scrunched state dwell times ($\tau_{scrunched}$) were extracted from the fits.

HMM analysis was also performed to extract the mean dwell times of the paused and excursion states. These states were categorized according to their FRET efficiency and the states that preceded and followed them. FRET states that were preceded by the baseline and were followed by a higher FRET state were assigned as paused states. FRET states that were preceded by the paused state and were followed by backtracked or baseline states were assigned as excursion states. The dwell times of the states were used to generate dwell time distributions, which were fitted with exponential decay curves as described in chapter 5. The mean paused state dwell times (τ_{pause}) and the mean excursion dwell times (τ_{excurs}) were extracted from the fits.

6.2.9 Accessible volume (AV) modelling

An *E. coli* RP_O model (see *Appendix A*) was used to model the positions of the donor and acceptor dyes during the initial transcription reaction[11]. The AV of the donor and acceptor dyes attached to the DNA via a flexible linker were calculated according to (197). Here, the sterically accessible positions of the dye were obtained from lattice calculations performed on a flexible linker (length=14 Å, width=4.5 Å) attached to the C5 atom of the dT base with the dye modelled as a sphere (radius 7 Å). The accessible volumes were approximated by a 3D Gaussian, the centre of which was approximated as

[11] It should be noted that there is evidence in the literature that the exact RP_O structure differs between promoters (66), therefore the AV modelling we perform is not a concrete expectation of distances but rather a guide in the design of our experiments.

the average position of the dyes. Visualisations and distance measurements between the donor and acceptor average dye positions were performed in Pymol.

For RP_O formed using the promoter DNAs shown in Fig 6.2, the donor dye was modelled attached to position -15, and the acceptor dye attached to position +15. For $RP_{ITC\leq7}$, the donor dye was modelled still attached to position -15 (6), while the acceptor dye was modelled attached to position +10 because DNA scrunching is expected to translocate the +15 acceptor dye upstream by 5 base pairs (bp) (7)[12].

For RP_O formed using the promoter DNA shown in Fig 6.4, the donor dye was modelled attached to position -15, and the acceptor dye attached to position +20. For each successive nucleotide incorporation step that results in DNA scrunching, we modelled the acceptor dye attached to a position 1-bp upstream of the previous attachment position. Therefore, modelling the acceptor dye attached to +19, +18, +17, +16, and +15 approximated the average positions of the acceptor dye in RP_{ITC3}, RP_{ITC4}, RP_{ITC5}, RP_{ITC6}, and RP_{ITC7}, respectively. The donor dye was modelled still attached to position -15. The inter-fluorophore distances calculated using this model are shown in Table 6.1. In all AV models, the donor dye was modelled attached to the non-template strand, and the acceptor dye was modelled attached to the template strand.

6.3 Results

6.3.1 Initial attempts to monitor RNA synthesis produces unexpected results

First, we attempted to reproduce earlier confocal spectroscopy results published by our group showing that RNAP trapped in abortive cycles produces an increase in FRET due to DNA scrunching (Fig. 6.2C) (6). For these experiments, we used the labelled *lac* promoter dsDNA fragment used in chapter 5 and shown in Fig. 6.2B. Unexpectedly however, we observed a decrease in FRET during initial transcription (Fig. 6.2D). During our investigations into the origin of this discrepancy, we discovered that since the

[12] The logic used here was that i) the synthesis of an N-mer RNA is accompanied by the scrunching of N-2-bp of DNA (6, 7), and ii) scrunching occurs along the helical path of DNA (198).

publication of (6), the length of the promoter DNA was increased in the downstream direction; originally, the dsDNA fragment spanned from -39 to +15 relative to the transcription start site, and was *end-labelled* at +15 (lacCONS+2(-39/+15); Fig. 6.2A). The current substrate DNA extends downstream to +25 and is *internally* labelled at +15 (lacCONS+2(-39/+25); Fig. 6.2B)[13]. We hypothesized that this change in the substrate DNA was responsible for the discrepancy in our results. To test whether this change in DNA length would be expected to affect the dye positions during initial transcription, we performed accessible volume (AV) modelling (see *Materials and Methods*) of the fluorophore dyes in RP_O and RP_{ITC7}.

The AV models showed that the inter-fluorophore distance was expected to decrease during scrunching with the previous lacCONS+2(-39/+15) DNA (Fig. 6.3A), and increase with the current lacCONS+2(-39/+25) DNA (Fig. 6.3B) during screw rotation of the downstream DNA (7, 198). This observation therefore provided an explanation for the results shown in Fig. 6.2C–D.

The AV models showed two further notable observations. First, with the lacCONS+2(-39/+25) DNA, the AV of the acceptor dye became highly restricted in RP_{ITC7} (Fig. 6.3B), which suggested that the acceptor dye might come into close contact with the protein during abortive initiation. Second, the inter-fluorophore distance change using lacCONS+2(-39/+25) was relatively small, and could have been increased if the acceptor were moved to an alternative attachment site. These two observations combined with the fact that RP_O formed using lacCONS+2(-39/+25) shows large-scale FRET fluctuations that could be mistaken for abortive cycling (see chapter 5) led us to explore alternative labelling schemes that would: 1) avoid clashes with the protein, 2) maximise the inter-fluorophore distance change during DNA scrunching, and 3) be blind to transcription bubble dynamics[14].

[13] The change was made to ensure that RNAP did not lose interactions with downstream DNA during DNA scrunching.

[14] This would simplify our data analysis because the only FRET changes we would be expected to see would be due to RNA synthesis and not the transcription bubble opening and closing.

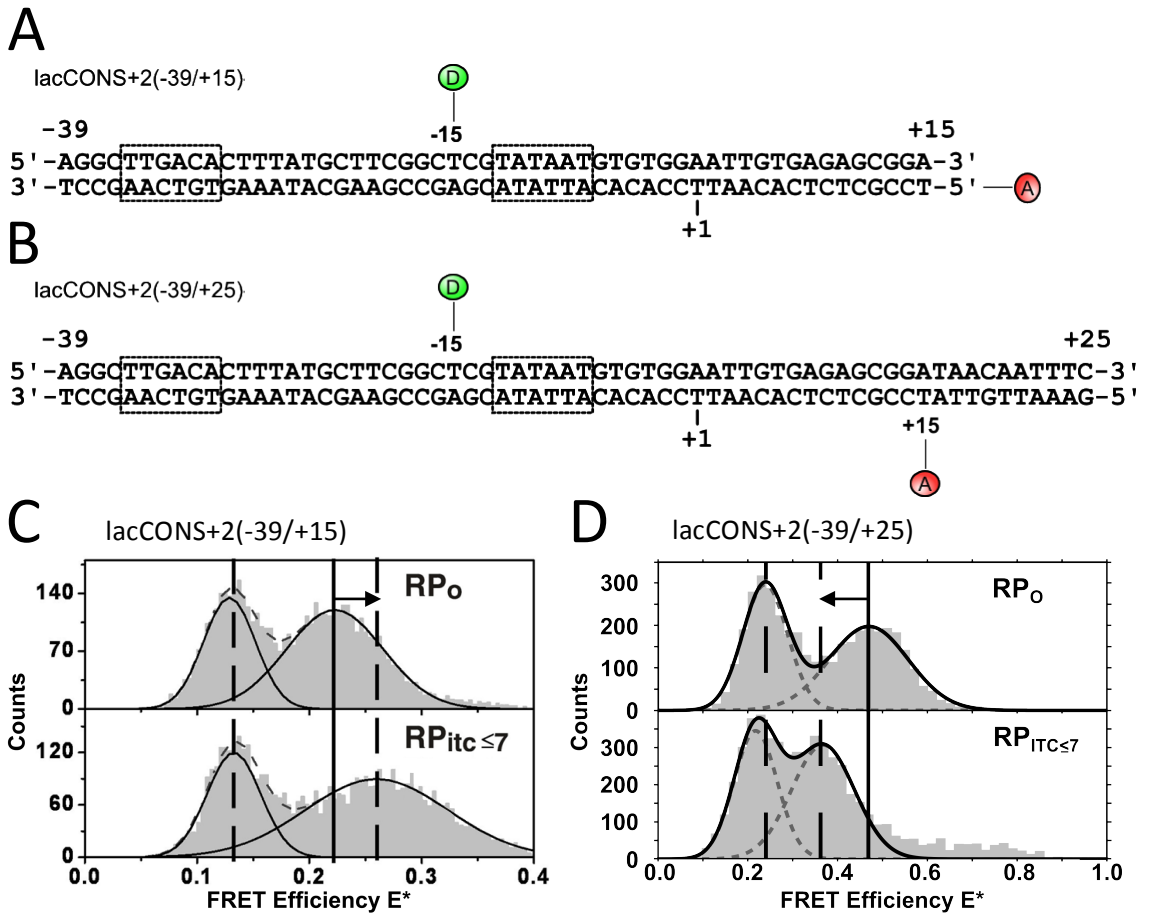

Figure 6.2. FRET changes during initial transcription. (A) Promoter DNA fragment used in (6) labelled with Cy3B (donor) and Alexa647 (acceptor). (B) Promoter DNA fragment used in chapter 5 labelled with Cy3B (donor) and ATTO647N (acceptor). The black boxes indicate the -10 and -35 elements. The transcription start site is labelled, as are the fluorophore labelling sites. (C) Stacked E* histograms of RP_O (top) and $RP_{ITC\leq7}$ (bottom) using the substrate DNA shown in A. Figure adapted from (6). (D) Stacked E* histograms of RP_O (top) and $RP_{ITC\leq7}$ (bottom) using the substrate DNA shown in B. Experiments were performed using confocal spectroscopy as described in (6). 4321 molecules contribute to the top histogram; 4651 molecules contribute to the bottom histogram.

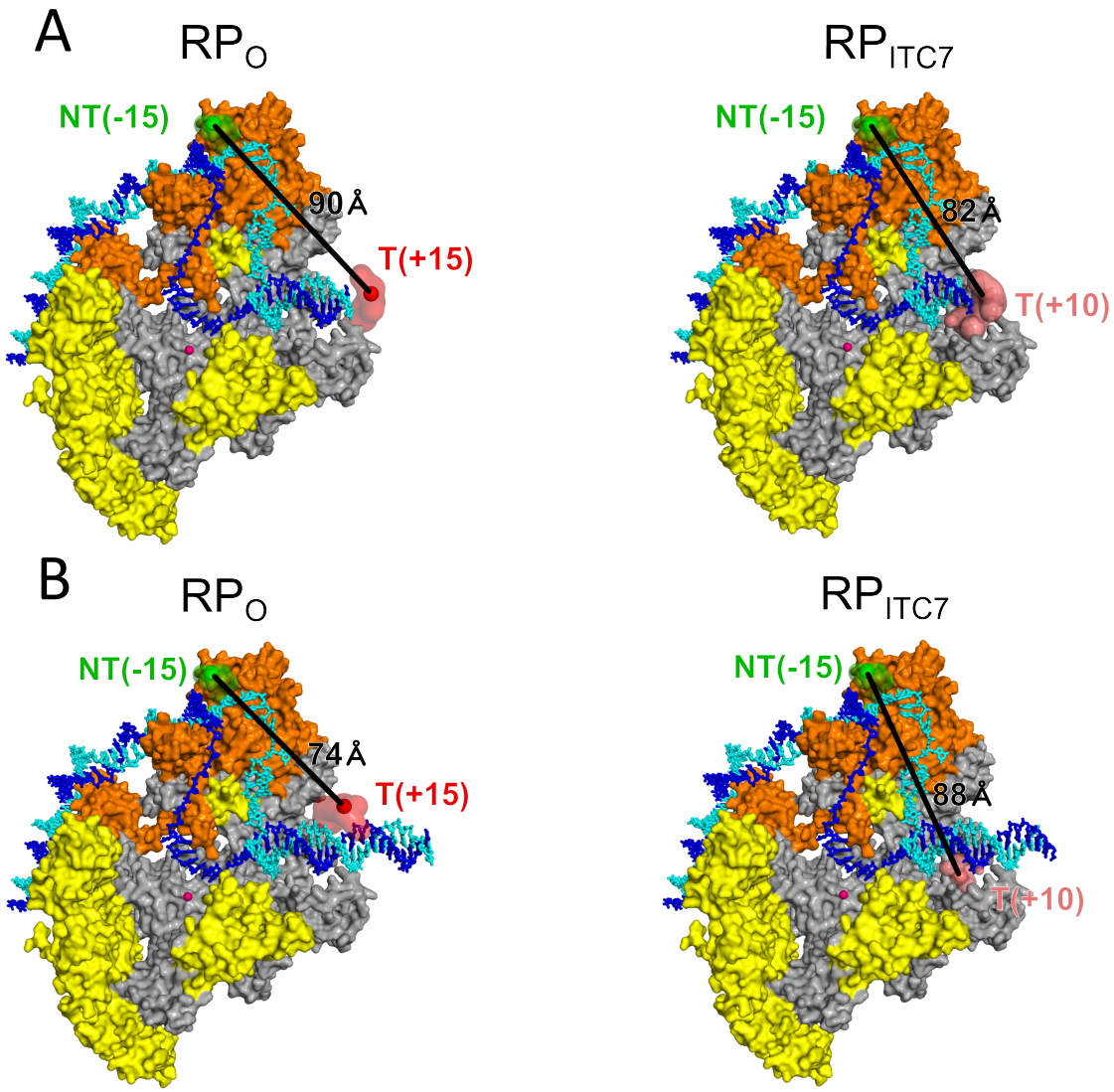

Figure 6.3. AV models of RNAP during initial transcription. (A) Cut-away structural models of *E. coli* RNAP-promoter DNA complexes (see *Appendix A*) showing the accessible volumes of the donor and acceptor dyes attached to lacCONS+2(-39/+15) (Fig. 6.2A). The acceptor dye is attached to +15 in RP_O. In RP_{ITC7}, it is assumed that the acceptor dye translocates with the downstream DNA in a corkscrew fashion during scrunching, and therefore the dye at +15 appears to be attached to +10 (see *Materials and Methods*). (B) Structural models showing the donor and acceptor dyes attached to lacCONS+2(-39/+25) (Fig. 6.2B). The mean positions of the donor and acceptor dyes are shown as spheres. The dye accessible volumes are shown as partially transparent surfaces. The donor dye is labelled in green in RP_O and RP_{ITC7}. The acceptor dye is labelled in red in RP_O and in pink in RP_{ITC7}. Inter-fluorophore distances are marked and labelled. The β subunit is omitted for clarity. σ^{70} is shown in orange, RNAP protein is shown in grey except for regions that protrude in front of the cut-away plane, which are yellow. The template strand is dark blue; the non-template strand is light blue.

6.3.2 An alternative labelling scheme

AV models of the acceptor dye attached to a number of possible labelling sites revealed that the +20 position was expected to give the largest FRET change during the initial stages of *de novo* RNA synthesis (Table 6.1: movement of dye from +20 to +15 leads to the most sensitive FRET increase of the possible dye attachment positions. FRET increases by 0.32), and was not expected to clash with the protein during abortive initiation (Fig. 6.4A). Furthermore, the AV modelling data indicated that RP_O would be expected to give very little FRET using this labelling scheme, and therefore would not report on transcription bubble dynamics. We tested the predictions made by the AV models by performing smFRET experiments using promoter DNA labelled at the -15 (donor) and +20 (acceptor) positions (Fig. 6.4C).

Table 6.1. Inter-fluorophore distances and expected FRET changes based on accessible volume modelling. The accurate FRET values were calculated using R_0=67 Å (139). The donor was modelled at -15.

Acceptor position in model (n)	Inter-fluorophore distance (Å) with donor at position -15	Expected accurate FRET (E)	Expected FRET change from position n to (n+1)
+25	98.5	0.09	NA
+24	100.4	0.08	-0.01
+23	102.6	0.07	-0.01
+22	105.5	0.06	-0.01
+21	108.6	0.05	-0.01
+20	107	0.05	0.00
+19	102.3	0.07	0.02
+18	92.4	0.10	0.03
+17	82.2	0.22	0.12
+16	76.3	0.31	0.09
+15	73	0.37	0.06

RP_O complexes were formed *in vitro* and immobilized on a PEG-passivated glass surface via a His-tag on RNAP using an Ab-immobilization technique (see *Materials and Methods*). The immobilized complexes were imaged in the presence of the initiating dinucleotide, ApA. ApA binds to the RP_O active site and produces RP_{ITC2}, which is primed for RNA synthesis but does not scrunch DNA (6, 7). Control experiments in which

immobilized DNA alone was imaged were also performed. Both experiments generated E* histograms with a single population centred at E*=0.23 (Fig. 6.5B). Furthermore, individual time trajectories of DNA and RP_O showed no transitions to FRET states above the baseline state (Fig. 6.5C). This confirmed that the new labelling scheme masks transcription bubble dynamics. I*n vitro* transcription (IVT) reactions confirmed that the presence of the dyes did not lead to major changes in the abortive products synthesised at this promoter (Fig. 6.6A)[15].

Next, we added a reaction mixture consisting of ApA, UTP, and GTP to immobilized RP_O to form $RP_{ITC\leq7}$ (Fig. 6.5A). The E* histogram from 6 fields-of-view is shown in Fig. 6.5B. As predicted by our AV models, we observed an additional population at a higher FRET value (centred at E*=0.35), which we attribute to complexes in which the DNA is in a scrunched conformation (scrunched complexes). Because experiments were performed at 21°C, a temperature at which RNAP-promoter DNA complexes are destabilised (66), we decided to include a mismatch in the transcription bubble region in order to increase the lifetime of active RP_O molecules attached to the surface (189) (sequence is shown in Fig. 6.4D). As expected, experiments performed using the pre-melted DNA (pmDNA) produced a greater number of initiation events per field-of-view compared with fully complementary DNA (dsDNA). Importantly, we saw no major difference in behaviour between the two DNAs using our smFRET assay (see later for a quantitative comparison), which was also reflected in IVT reaction gels, which showed similar abortive products produced under $RP_{ITC\leq7}$ conditions for both dsDNA and pmDNA (Fig. 6.6B). We therefore decided to perform the majority of our real-time abortive initiation experiments using pmDNA in order to increase the yield of reactions on the surface. It is worth noting however that the pmDNA did appear less efficient in promoter escape compared to dsDNA, which has implications for the design of promoter escape experiments (see section 6.3.8).

[15] The dyes did however lead to an increase in the production of longer non-abortive products, presumably because the acceptor dye clashes with the protein active site after escaping the promoter.

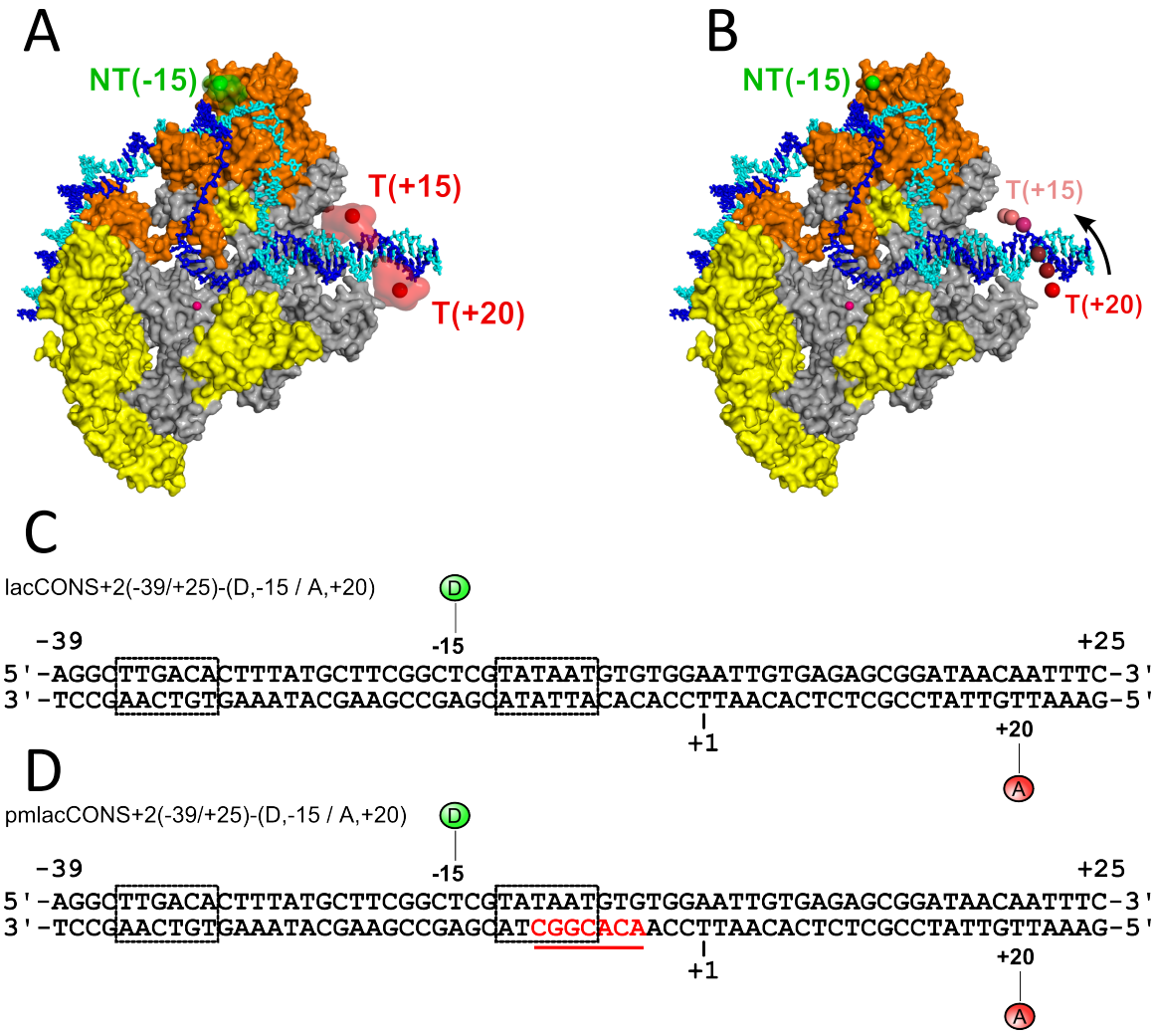

Figure 6.4. An alternative labelling scheme. (A) AV model of the donor dye attached to -15 in RP_O and RP_{ITC7}, and the acceptor dye attached to +20 in RP_O and +15 in RP_{ITC7}. The accessible volumes and mean positions of the dyes are shown. In both positions the acceptor dye does not clash with protein. (B) AV model of RNAP during initial transcription. The acceptor dye is attached at position +20 and moves towards the donor during initial transcription. The final position reached in the model is +15, which corresponds to RP_{ITC7}. (C) Schematic representation of the new promoter DNA labelling scheme used for experiments. (D) As C except with a mismatch in the bubble region (shown in red and underlined). The DNA schematics in C and D are coloured and labelled as in Fig. 6.2. The donor is Cy3B and the acceptor is ATTO647N.

6.3.3 RNA synthesis observed in real-time

Next, we decided to monitor initial transcription in real-time by adding the NTP reaction mixture to the observation chamber while we recorded data. First, we monitored the behaviour of immobilized complexes when we added the $RP_{ITC\leq7}$ reaction

mixture (see *Materials and Methods*). Our results showed that ~60% of immobilized complexes exhibited an increase in FRET upon NTP addition. Furthermore, the increase in FRET was relatively well synchronised across the field-of-view with >95% of active molecules displaying the increase in FRET within 3 sec of the NTP addition point.

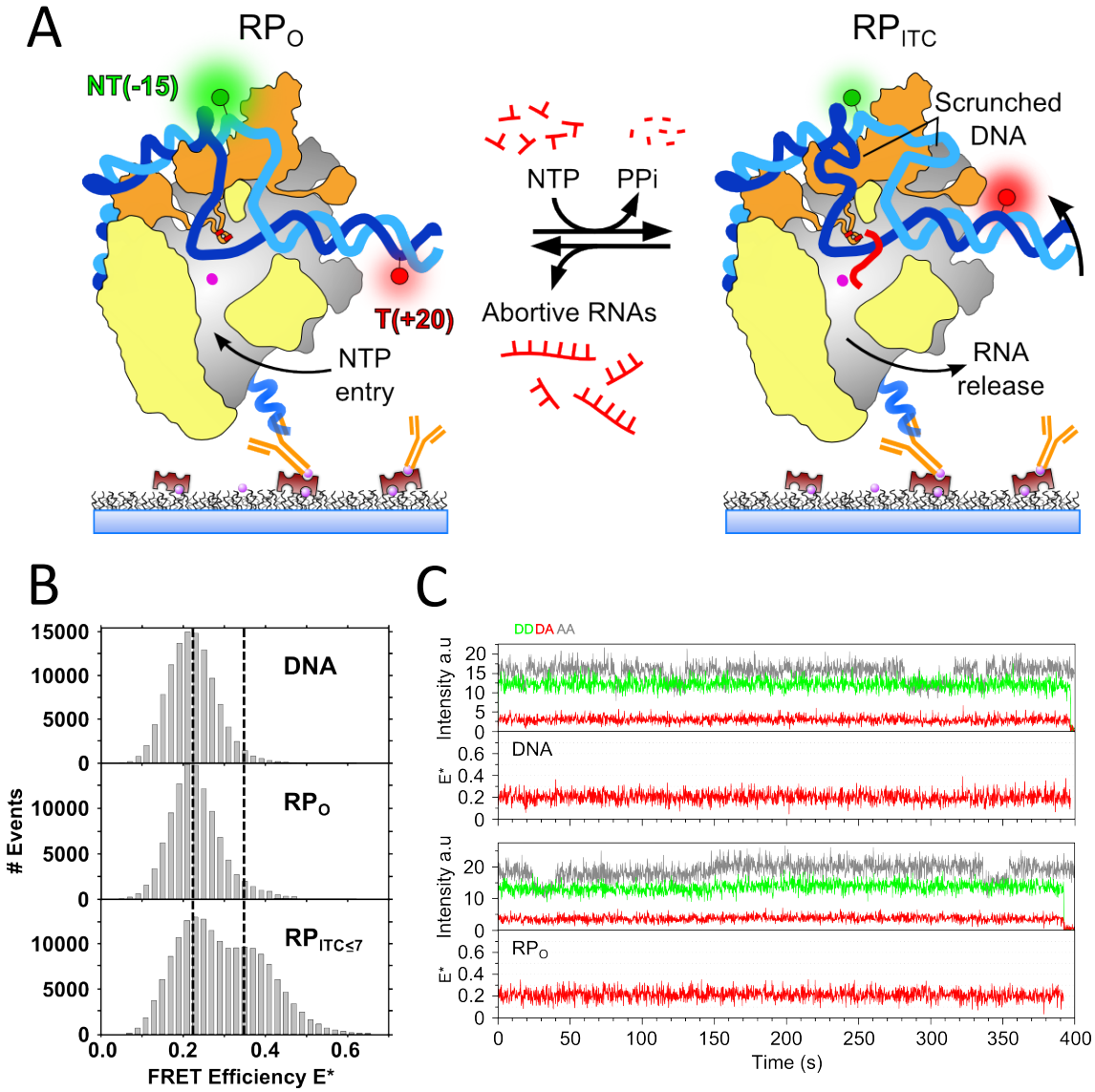

Figure 6.5. smFRET experiments using the *lac* promoter labelled at positions -15 and +20. (A) Schematic of RNA synthesis at the surface. The penta-His antibody binds to the RNAP His-tag and anchors the RNAP to the surface (see chapter 5). The donor and acceptor dyes are labelled and shown in green and red, respectively. RNA is shown as a red strand at the active site in RP_{ITC}. Scrunched DNA is labelled in RP_{ITC}, as is the movement of the acceptor dye. The protein and§ DNA are coloured as in Fig. 6.3 and 6.4. (B) Stacked E* histograms of surface-immobilized DNA (top); RPo + ApA (middle); and $RP_{ITC\leq7}$ (bottom). Each histogram was constructed using data from six fields-of-view. The NTP reaction mixture was added before we commenced data acquisition. Temporal

resolution is 20 ms. (C) Example time trajectories of DNA (top) and of RPo + ApA (bottom). FRET remained constant at E*~0.23 for many minutes. ATTO647N intensity state changes can be seen in the AA channel. Temporal resolution is 200 ms[16]. 600 molecules (250 frames for each molecule) contributed to each E* histogram.

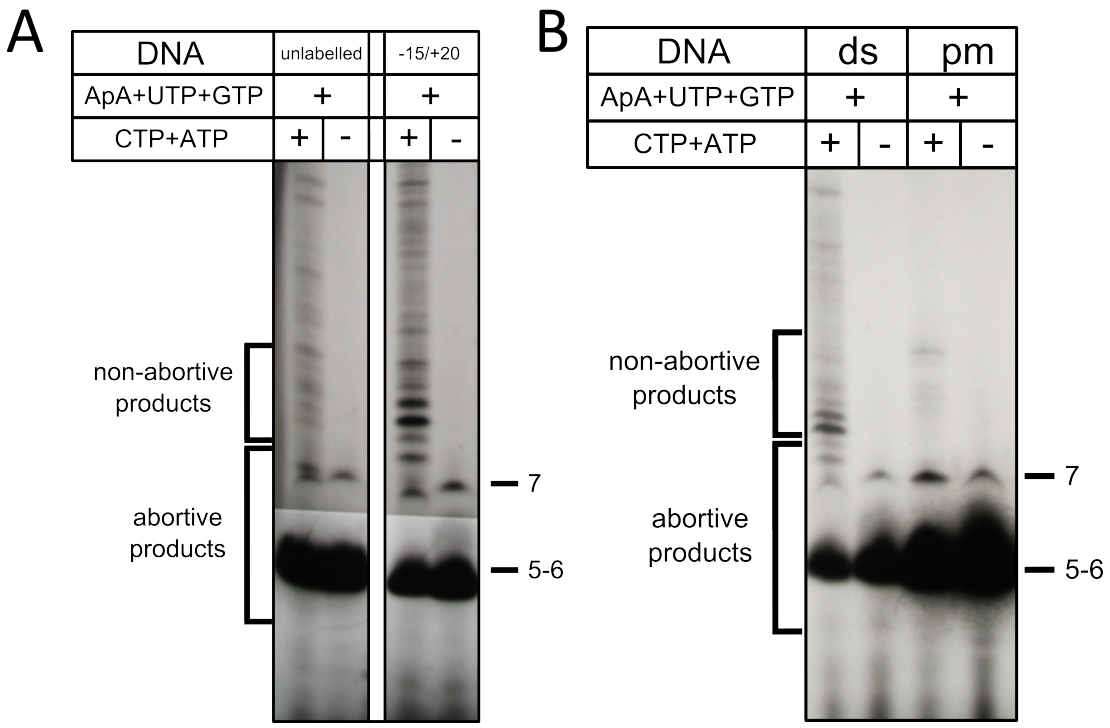

Figure 6.6. *In vitro* transcription reactions. [α^{32}P]GTP was used in both A and B. The initial transcribed sequence is 5'-AAUUGUG-3'. (A) Effect of the donor and acceptor dyes on the activity of RNAP at the lacCONS+2 promoter. The line/break in the middle of the gel results from the overlay of two X-ray films (B) Effect of the transcription bubble mismatch on the abortive RNA products synthesised using a fluorescently labelled lacCONS+2 promoter. For both figures, transcription was performed in the presence of ApA, U, G, C, and A (run-off conditions) or ApA, U, and G ($RP_{ITC\leq7}$ conditions) as indicated. The [α^{32}P]-RNA products were separated on a 20% PAGE-7M urea gel. The gel was run from top to bottom. For both gels, abortive and non-abortive RNA products are labelled. Non-abortive RNA products include RNAs that result from promoter-proximal pausing (196). The increase in non-abortive products for labelled DNA compared with unlabelled DNA in Fig. 6.6A most likely results from clashes between the acceptor dye and the protein after the polymerase has escaped the promoter.

Once the reaction was initiated however, molecules exhibited heterogeneous behaviour. Some molecules adopted stable scrunched conformations that lasted for many seconds, and often beyond the fluorescent lifetimes of the fluorophores (Fig.

[16] Experiments were performed at temporal resolutions of up to 10 ms. No fluctuations were observed at all temporal resolutions tested.

6.7A), while others returned to the FRET baseline after some dwell in the scrunched state (Fig. 6.7B). The majority of active molecules reached E*=0.35 smoothly i.e. without stepping (Fig. 6.7A). However, some molecules (10%) showed complex behaviours such as pausing at E*=0.35 before reaching a final value of E*=0.45 (Fig. 6.7B, left), which was then followed by a partial decrease in FRET to a scrunched state above the baseline (Fig. 6.7B, right).

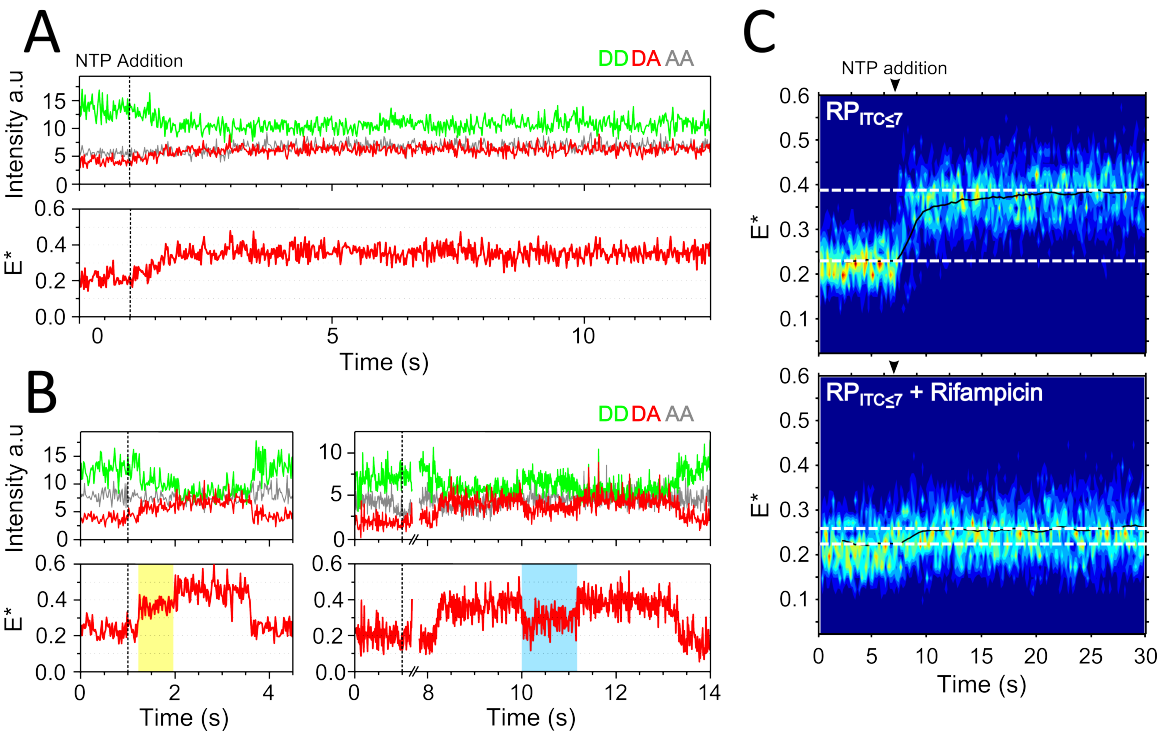

Figure 6.7. RNA synthesis observed in real-time. All figures are from $RP_{ITC\leq7}$ experiments (A) Example time trajectory showing a stable increase in FRET to E*=0.35 upon the addition of the $RP_{ITC\leq7}$ NTP reaction mixture. The NTP addition point is shown with a dashed line. (B) Example time trajectories showing two commonly observed behaviours. The left hand figure shows an increase in FRET to E*=0.45 after a pause at E*=0.35 highlighted with a yellow background. The right hand figure shows a backtracking event (highlighted with a blue background) that is followed by a return to the maximally scrunched state. In both figures the FRET signal returns to the baseline to complete an abortive cycle. (C) Heat maps of the E* distributions of FRET time trajectories showing activity upon the addition of the $RP_{ITC\leq7}$ NTP reaction mixture in the absence (top) and presence (bottom) of rifampicin. The NTP addition point is highlighted with the arrowhead. 45 molecules from a single field-of-view were used to produce the top figure, 37 molecules from a single field-of-view were used to produce the bottom figure. Blue to red colours represent an increasing number of events.

By performing ALEX, we were able to probe the acceptor fluorescence directly and thereby eliminate the possibility that the heterogeneity we observed was a result of known intensity fluctuations of ATTO647N (183). Furthermore, control experiments in which we added the $RP_{ITC\leq7}$ NTP reaction mixture to the well in the presence of rifampicin – an inhibitor that blocks the synthesis of RNA products >3-nt in length (6, 7, 199) – drastically reduced the FRET change to only ~0.04 (Fig. 6.7C). We therefore concluded that the FRET changes we observed were due to RNA synthesis. The fact that we observed a small FRET increase for $RP_{ITC\leq7}$ experiments in the presence of rifampicin led us to hypothesize that the magnitude of the FRET increase was related to the length of the RNA synthesised. We decided to test this hypothesis by investigating whether there is a relationship between the magnitude of the FRET increase and the length of the nascent RNA synthesised.

6.3.4 Calibrating our FRET signal

We performed a calibration of our FRET signal by comparing the magnitude of the FRET increase in $RP_{ITC\leq4}$, $RP_{ITC\leq5}$, and $RP_{ITC\leq7}$ experiments (Fig. 6.8). Our results showed that the average FRET change (ΔE) for active molecules across a field-of-view increased in size as follows:

$$\Delta E(RP_{ITC\leq4}) < \Delta E(RP_{ITC\leq5}) < \Delta E(RP_{ITC\leq7}).$$

This observation confirmed that ΔE was correlated with the length of the RNA synthesised, which enabled us to assign E* values to specific RNA lengths. The heat maps shown in Fig. 6.8A provided us with a guide as to what these values might be for 4-mer, 5-mer, and 7-mer RNAs; however, the heat maps show an average of the behaviours across a field-of-view and therefore could not be used to assign E* values in a reliable manner. Instead, we inspected individual time trajectories to identify the maximum E* value (E^*_{max}) reached in each experiment (Fig. 6.8B). Our logic was that E^*_{max} would correspond to the longest RNA synthesised under specific experimental conditions. So for example, in $RP_{ITC\leq4}$ experiments E^*_{max} would correspond to a 4-mer

RNA, while in $RP_{ITC\leq5}$ experiments, E^*_{max} would correspond to a 5-mer RNA, etc. We found that for $RP_{ITC\leq4}$ E^*_{max}=0.30, for $RP_{ITC\leq5}$ E^*_{max}=0.35, and for $RP_{ITC\leq7}$ E^*_{max}=0.45. We therefore concluded that E*=0.30 corresponds to the synthesis of a 4-mer RNA, E*=0.35 corresponds to the synthesis of a 5-mer RNA, and E*=0.45 corresponds to the synthesis of a 7-mer RNA. We did not calculate accurate FRET values for our data since we were not interested in retrieving accurate distances; relative FRET changes were sufficient for us to make our conclusions. Consequently, the predicted accurate FRET values (Table 6.1) and the apparent FRET values we assigned to RNA lengths were invariably going to be different. Having said this, the predicted and observed ΔE values showed reasonably good agreement. Any differences are likely due to the fact that our predictions are based on an RP_O structural model that may deviate slightly from the true RP_O structure.

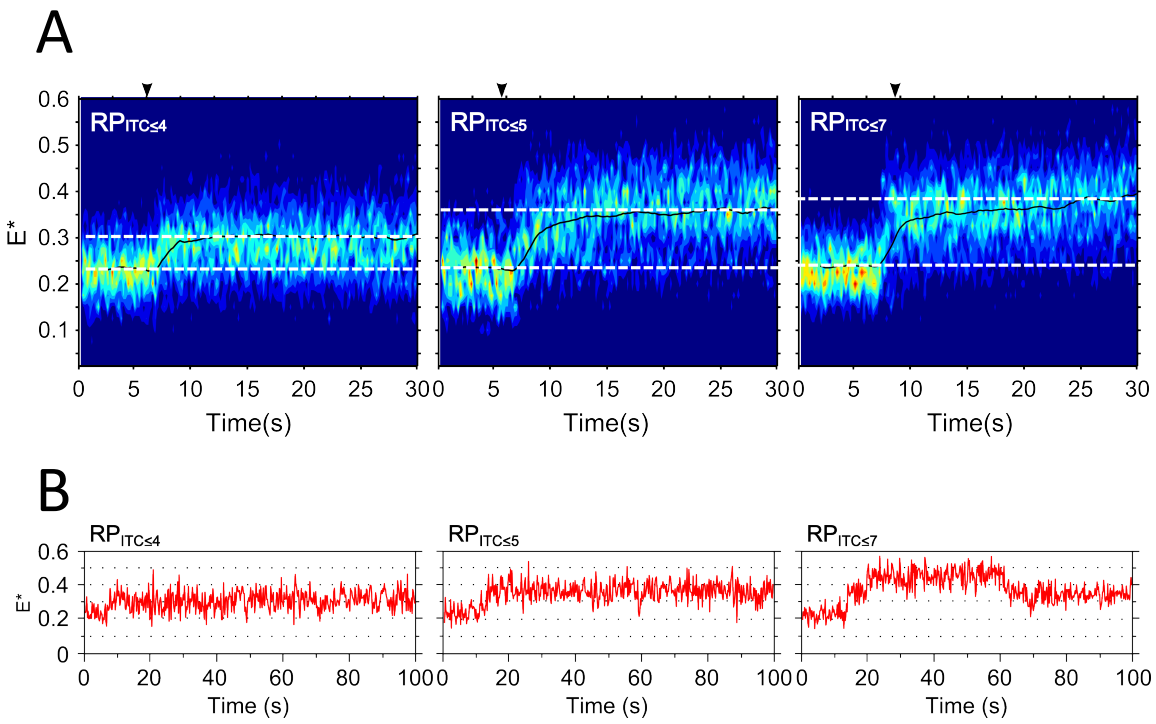

Figure 6.8. FRET calibration. (A) Heat maps of the E* distributions of FRET time trajectories showing activity upon the addition of the $RP_{ITC\leq4}$, $RP_{ITC\leq5}$, and $RP_{ITC\leq7}$ NTP reaction mixtures. The NTP addition point is shown at the top with the arrowhead. 45, 53, and 51 molecules were used to produce the $RP_{ITC\leq4}$, $RP_{ITC\leq5}$, and $RP_{ITC\leq7}$ heat maps, respectively. Blue to red colours represent an increasing number of events. (B) Example time trajectories showing the largest FRET change observed in $RP_{ITC\leq4}$, $RP_{ITC\leq5}$, and $RP_{ITC\leq7}$ experiments. The $RP_{ITC\leq7}$ example time trajectory also shows pausing and backtracking events. Temporal resolution is 200 ms.

6.3.5 Structural interpretations of complex $RP_{ITC\leq7}$ behaviour

Having completed our empirical assignment of the FRET states, we then made structural interpretations of some of the more complicated behaviours we observed. First we addressed the pausing behaviour seen for $RP_{ITC\leq7}$ experiments (see section 6.3.3). In these experiments we found that ~10% of molecules reached E*=0.45 after pausing at E*=0.35 (Fig. 6.9A–B). Based on our FRET calibration, we interpret this as the scrunching of 3-bp of DNA to form RP_{ITC5}, at which point the polymerase pauses for an average duration (dwell-time) of ~24 sec (Fig. 6.9C), and then re-commences scrunching to form RP_{ITC7}. We found that once E*=0.45 was reached, it was only transiently sampled with an average dwell-time of 5.1 sec (Fig. 6.9C) before the polymerase either backtracked to E*=0.35 (Fig. 6.9A), or returned to the FRET baseline (Fig. 6.9B). We refer to this transient sampling as an excursion. One possible explanation for this behaviour is that there is an energetic barrier that impedes the smooth progression of RNAP from the RP_{ITC5} state to the $RP_{ITC6/7}$ states. Presumably this barrier destabilises RP_{ITC7}, which is why the mean dwell-time of the excursions to E*=0.45 are much shorter than the pause dwell times. It should be noted that there is a small amount of variability in the exact value of the excursion FRET states. This variability is very likely a result of baseline shifts in FRET, but may also be due to events in which RP_{ITC6} is formed. However, because our initial transcribed sequence does not permit us to assign a FRET state to RP_{ITC6}, we are unable to comment further on this variability.

The presence of an energetic barrier that impedes RNAP progression beyond the RP_{ITC5} state would also help explain why the $RP_{ITC\leq7}$ heat map in Fig. 6.8A appears to reach a plateau considerably below E*=0.45, which is the FRET value expected for RP_{ITC7} based on our empirical calibration. Presumably, the barrier at E*=0.35 reduces the ability of RNAP to synthesise 7-mer RNAs, which means most polymerases in $RP_{ITC\leq7}$ experiments synthesise only 5-mer RNAs. This interpretation is supported by IVT reactions gels, which clearly show that 7-mer RNAs are rarely produced under $RP_{ITC\leq7}$ conditions at our promoter (Fig. 6.6).

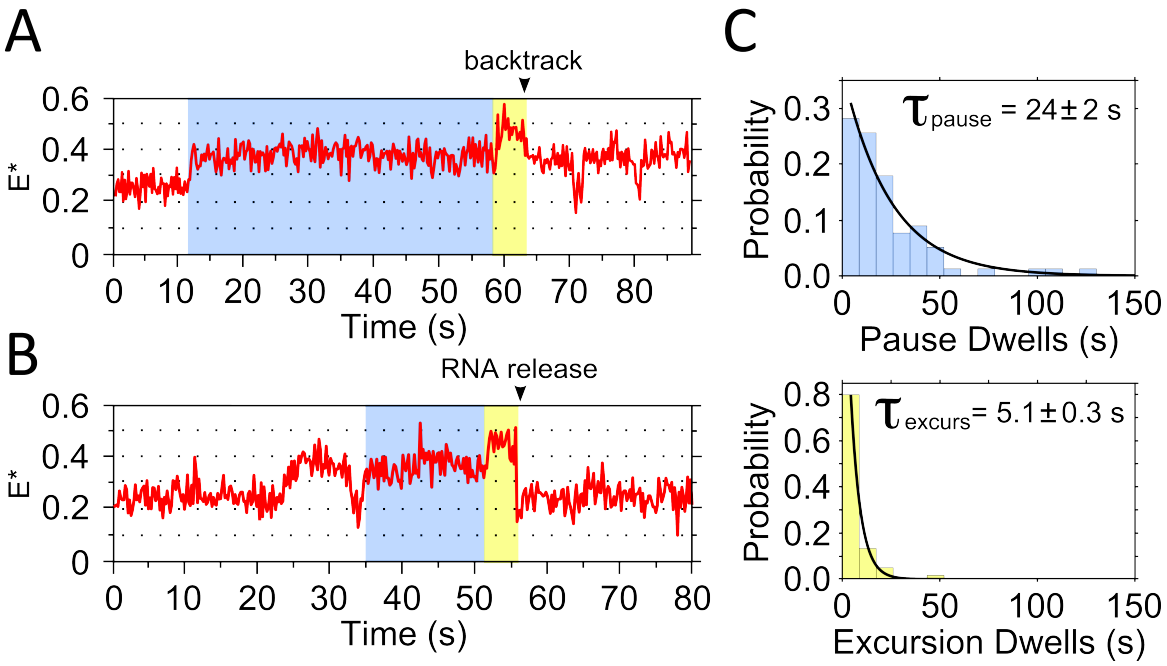

Figure 6.9. Complex $RP_{ITC\leq7}$ behaviours. (A) Example time trajectory showing pausing of the polymerase at E*=0.35 (blue background), followed by a transient excursion to E*=0.45 (yellow background). The excursion is followed by a backtracking event indicated with an arrowhead. (B) Example time trajectory showing a pause and excursion as in A except that in this example the excursion to E*=0.45 is followed by an RNA release event shown with an arrowhead. (C) Dwell time histograms and exponential fits of the pause dwell (top; 84 events from 84 molecules) and excursions dwell (bottom; 60 events from 60 molecules). The mean dwell-times are given inset. Temporal resolution is 200 ms.

6.3.6 Kinetics of abortive cycling

Next, we monitored $RP_{ITC\leq7}$ molecules for hundreds of seconds in order to observe multiple RNA synthesis and release cycles per polymerase (Fig. 6.10). The results of these experiments revealed two types of behaviour: stably scrunched behaviour and cycling behaviour. Molecules classified as stably scrunched (45%) adopted scrunched states that lasted for more than 120 sec (Fig. 6.10A, top). Cycling molecules (55%) adopted scrunched conformations that lasted for less than 120 sec and were followed by a return to the baseline FRET state and further rounds of RNA synthesis (Fig. 6.10A middle and bottom). Interestingly, more than doubling the NTP concentration to 200 µM produced only a small increase (5%) in the proportion of molecules showing stably scrunched behaviours.

In order to extract the mean dwell times of the scrunched states, we fit our time trajectory cycling data with a Hidden Markov Model (HMM) as described in *Materials and Methods* (Fig. 6.10B). Our analysis revealed that the scrunched state exhibits biexponential decay kinetics and mean dwell times of 10±1.8 sec (65% of all events) and 39±12 sec (35%). As expected from earlier gel studies (84), we found no strong correlation between the length of the RNA in the scrunched complex and the stability of that scrunched complex.

To rule out the possibility that the immobilization approach was responsible for the observed inter-molecular heterogeneity in abortive cycling we performed the same experiments described above, this time, using complexes immobilized via the promoter DNA instead of the protein. Our results revealed no obvious difference in behaviour.

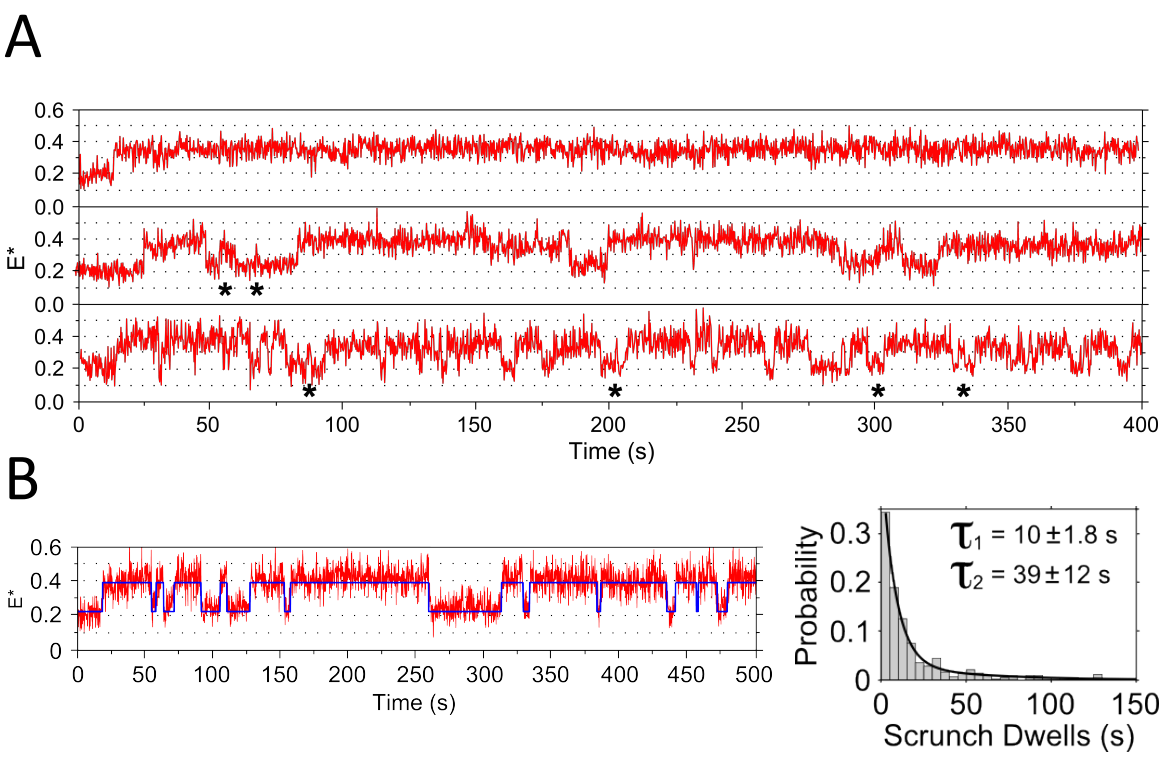

Figure 6.10. Heterogeneity and kinetics of abortive initiation. (A) Example time trajectories of transcribing complexes in $RP_{ITC\leq7}$ experiments. Examples of stable scrunched (top), and abortive cycling (middle and bottom) behaviours shown. FRET increase events to E*<0.4, which may represent the synthesis of short (<5-nt) abortive RNAs are marked with asterisks. (B) Example time trajectory of abortive cycling data fit to a HMM (blue line) is shown on the left. The distribution of scrunched state dwell times is shown on the right with the extracted mean dwell times inset. The histogram shows 445 events from 112 molecules. The temporal resolution is 200 ms. A log-linear representations of the dwell time distribution is shown in Fig. C3 *Appendix C*.