We also performed $RP_{ITC\leq7}$ experiments using fully complementary promoter DNA to test whether the transcription bubble mismatch in the pmDNA influences the mean dwell times of the scrunched complexes. Again, we observed the same behaviours as described above for pmDNA data, and found very similar decay kinetics of the scrunched complexes (Fig. 6.11). These findings support our earlier conclusion that the mismatch in the pmDNA does not influence the abortive initiation behaviour of RNAP.

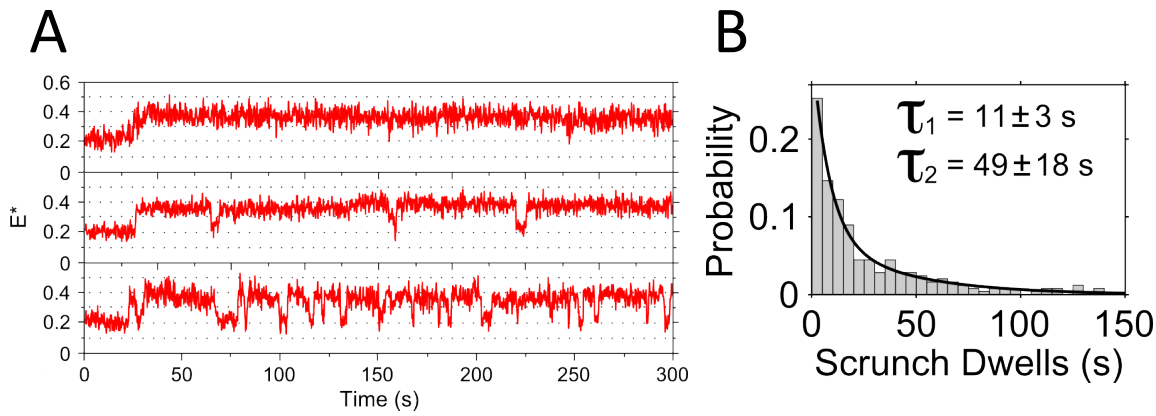

Figure 6.11. Abortive initiation behaviour is minimally affected by the bubble mismatch. (A) Example time trajectories of stable scrunched (top), and abortive cycling (middle and bottom) behaviours of immobilized RNAP-promoter DNA complexes formed using fully complementary DNA (dsDNA). (B) Dwell-time distribution for dsDNA abortive cycling data. The histogram shows 250 events from 91 molecules. Temporal resolution is 200 ms. The complex yield was low in these experiments compared with pmDNA experiments. A log-linear representations of the dwell time distribution is shown in Fig. C3 *Appendix C*.

6.3.7 σR3.2 in abortive initiation

In section 6.3.5 we hypothesized that there may be a barrier to the synthesis of RNAs beyond a length of ~5-nt. We attempted to identify a possible structural candidate for this barrier by generating structural models of the RNAP active centre during initial transcription (Fig. 6.12). Our structural models showed that the 5'-end of RNA molecules between 5-nt and 6-nt in length would be expected to clash with the σ^{70} region 3.2 loop (σR3.2 finger). Based on our model, we hypothesized that the σR3.2 finger forms a barrier to the extension of RNAs beyond 5–6-nt in length. We argued that if this hypothesis was correct, deleting the σR3.2 finger would increase the yield of 7-mer

RNAs in $RP_{ITC\leq7}$ experiments, and would eliminate the pauses en route to E*=0.45. We tested our hypothesis by performing $RP_{ITC\leq7}$ experiments using a mutant RNAP lacking the σR3.2 finger (residues 513-519; Δ3.2). We compared the behaviour of wt and Δ3.2 complexes, and found three major differences between the two complexes.

Firstly, Δ3.2 $RP_{ITC\leq7}$ complexes were found to reach E*=0.45 more often than wt $RP_{ITC\leq7}$ complexes (36% for mutant complexes compared with 12% for wt complexes). This can be seen when we compare the heat maps for the two complexes (Fig. 6.13A), which show that Δ3.2 complexes are able to sample higher FRET states more readily than wt complexes. Secondly, the vast majority (90%) of Δ3.2 complexes that did reach E*=0.45, did so without pausing at E*=0.35; the comparable number for wt complexes was only 26%. Example time trajectories showing this behaviour can be found in Fig. 6.13B–C. We note that only 36% of molecules reached E*=0.45, the remainder reached lower FRET states (E*=0.3–0.35). Thirdly, there was a decrease in the proportion of molecules exhibiting stable scrunched events (>120 sec) from 45% for wt complexes to 15% for Δ3.2 complexes. This reduced stability of the scrunched state can be clearly seen in the width of the heat maps (Fig. 6.13A), which show a much larger spread of E* values after NTP addition for Δ3.2 complexes compared with wt complexes.

The reduced stability of the Δ3.2 mutant complex scrunched states was also evident from our analysis of scrunched state dwell times. We fit a Hidden Markov Model to our Δ3.2 mutant data, extracted the mean dwell times of the scrunched states, and compared the results to that of wt complexes (Fig. 6.13D). Our analysis revealed that in contrast to wt complexes (Fig. 6.10B), Δ3.2 complexes exhibited monoexponential decay kinetics and a mean dwell time of 10.3±1.9 s. In fact, only a very small proportion of molecules displayed long dwell times. These results suggest that the σR3.2 finger contributes to the stability of nascent RNAs at the RNA polymerase active site. Our conclusions are supported by IVT assays comparing the abortive profiles of wt and Δ3.2 complexes (Fig. 6.13E). It should also be noted that Fig. 6.13E supports the conclusion drawn from the time trajectories in Fig 6.10 that more 5-mer abortive RNAs are produced than short 3-mer abortive RNAs.

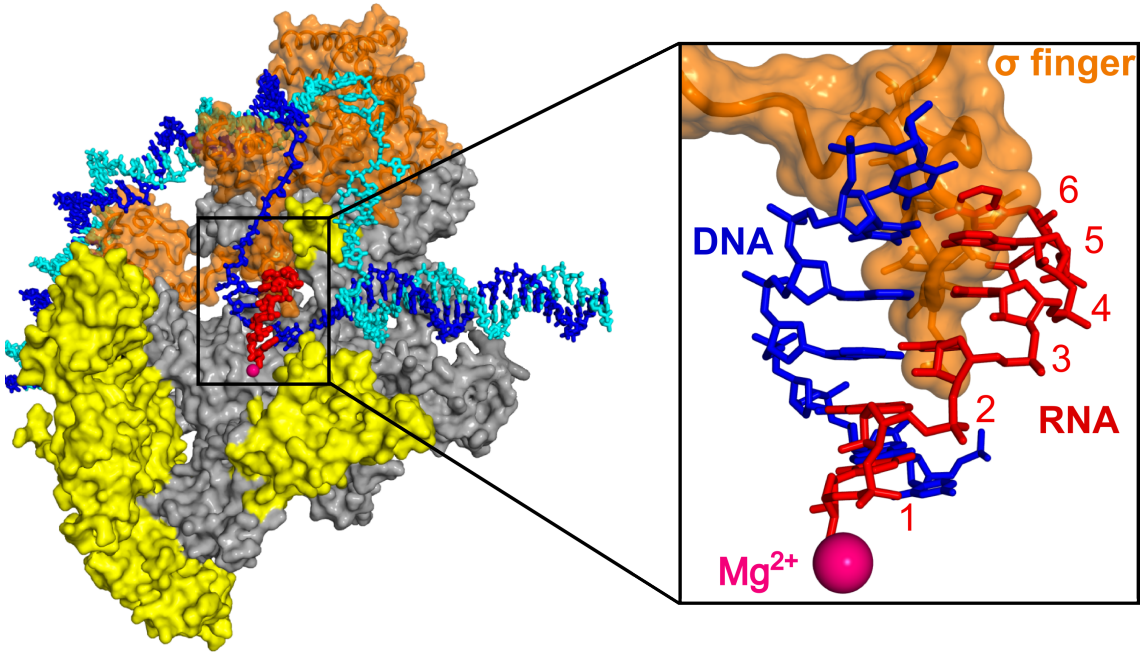

Figure 6.12. Structural model of the *E. coli* active centre during initial transcription. Cut-away Structural model showing a clash between the nascent 6-mer RNA (red stick representation) and the σR3.2 finger side chains (orange). The model is coloured as in Fig. 6.3.

6.3.8 Promoter escape observed in real-time

The transition from initiation to elongation is known as promoter escape. During this process, RNAP loses contacts with the promoter elements and begins translocating forwards without scrunching DNA. As the polymerase progresses along the DNA, the downstream edge of the transcription bubble moves with it while the upstream edge is re-annealed. In our last set of experiments, we attempted to observe this process directly using our smFRET assay.

In these experiments, we immobilized pre-formed RP_O and added all four NTPs to the observation chamber while recording data in order to monitor promoter escape in real-time. It should be noted that one major difference between this and previous experiments was that here we used dsDNA instead of pmDNA. We made this change because pmDNA has a mismatch in the transcription bubble region, which means that the upstream edge of the transcription bubble does not re-anneal during escape.

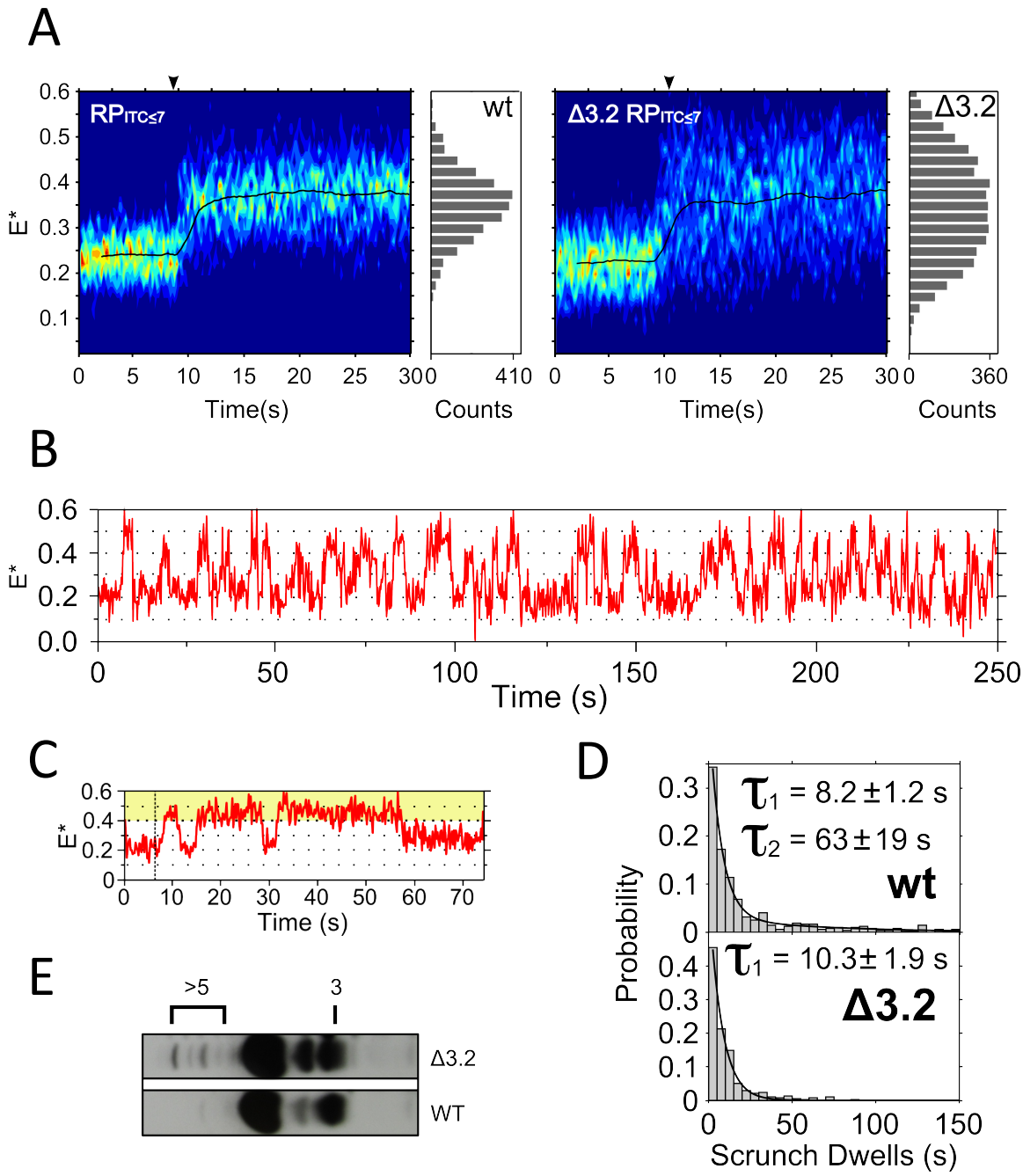

Figure 6.13. σR3.2 in abortive initiation. (A) Heat maps for wt and Δ3.2 complexes in $RP_{ITC\leq7}$ experiments. Histograms showing E* values sampled after NTP addition are shown as side projections. Blue to red colours represent an increasing number of events. (B) Example time trajectory showing abortive cycling behaviour where E*=0.45 is sampled without pauses. (C) Example time trajectory showing slower cycling behaviour than that shown in Fig. 6.13B. FRET values >0.4 are highlighted with a yellow background. This example time trajectory shows no pauses. (D) Dwell time distributions of wt (top; 445 events from 112 molecules) and Δ3.2 (bottom; 392 events from 95 molecules) scrunched states with mean dwell times shown inset. (E) *In vitro* transcription reactions showing the effect of deleting the σR3.2 finger on the abortive products synthesised at the lacCONS+2 promoter. Experiments were performed in the presence of [α^{32}P]UTP, ApA, U, and G ($RP_{ITC\leq7}$ conditions). The [α^{32}P]-RNA products were

separated on a 20% PAGE-7M urea gel. The gel was run from left to right. Log-linear representations of the distributions are shown in Fig. C3 *Appendix C*.

We reasoned that this loss of interactions might interfere with the escape process and therefore opted to use fully complementary DNA[17]. Our hypothesis that the mismatch might interfere with promoter escape is supported by IVT run-off reactions that appear to show a reduced efficiency in escape for the pmDNA compared with dsDNA (Fig. 6.6B).

Time trajectories from these experiments showed FRET changes that were unique to promoter escape experiments (Fig. 6.14A–B). These events exhibited the following basic features: 1) a non-instantaneous increase in FRET to a maximum value of 0.6–0.8; 2) in 50% of cases, this FRET change was accompanied by a pause at E*~0.35 with a mean dwell of 14.6±1.4 s (Fig. 6.14B–C). The other 50% of molecules showed no pause; 3) once the maximum FRET value was reached, we observed an almost instantaneous decrease in FRET back to the baseline FRET state; 4) after the baseline FRET value was restored, we saw no subsequent FRET events.

We interpret these FRET changes as follows: the FRET signal increases because of scrunching of the downstream DNA into the protein during the initial stages of *de novo* RNA synthesis. Scrunching reaches a maximum level when the stress accumulated is sufficient to drive promoter escape. At this point the polymerase loses promoter contacts and translocates forwards. The forward movement is most likely abrupt and driven by the release of strain in the maximally scrunched complex, hence the almost instantaneous return to the FRET baseline. No activity is observed afterwards because the polymerase has cleared the promoter. The acceptor dye most likely clashes with the active centre after the synthesis of ~12-mer RNAs. This is not an issue for our investigations since we expect promoter escape to occur when a ~9–10-mer RNA has been synthesised.

Having assigned our observed FRET changes to promoter escape, we were then able to proceed with further analysis of our time trajectory data. First we addressed the question of whether all promoter escape events are preceded by abortive cycling. Our

[17] As expected, the yield of active complexes on the surface was low.

time trajectory data showed that 35% of all promoter escape events occur after some form of abortive cycling i.e. the majority of molecules do not go through abortive cycling before escaping the promoter. Interestingly, of those molecules that did undergo abortive cycling prior to escape, 80% exhibited only one or two abortive cycles. The remaining molecules showed 3–4 abortive cycles prior to escape. However, it should be noted that since our assay lacks the sensitivity to reliably resolve individual rapid abortive cycles of RNAs smaller than 5-nt in length, it is possible, and indeed likely, that we miss events in which 2–4-mer RNAs are abortively synthesised and released.

Although the position at which promoter escape occurs is known to vary depending on the promoter sequence (84, 109), no clear relationship between the two has been shown (109). Since the maximum FRET value (E^*_{escape}) that is reached in our assay most likely represents the escape position, we decided to investigate whether E^*_{escape} shows variability between molecules. Our time trajectory data showed that E^*_{escape} varied between E*=0.6 and E*=0.8. To minimise the likelihood that this apparent inter-molecular heterogeneity was due to baseline shifts influencing E^*_{escape} (167), we calculated the difference between the baseline FRET value and the E^*_{escape} value (ΔE; Fig. 6.15). Our results showed a broad range of differences between the baseline and maximum FRET state, with two clear peaks centred at 0.41 and 0.58. This suggests that there is some plasticity regarding the exact position at which promoter escape occurs. However, further work is needed to identify the factors that influence ΔE in order to reliably state that promoter escape occurs at >1 position at the *lac* promoter.

6.3.9 Abortive cycling in the presence of all four NTPs

A subpopulation of RNAP-promoter DNA complexes showed activity but did not escape the promoter when all four NTPs were added to the observation chamber. Many of these molecules exhibited similar behaviours to those described earlier for complexes trapped in $RP_{ITC\leq7}$ abortive cycles (Fig. 6.16A). Other molecules showed FRET changes that appeared to show the abortive synthesis of RNAs longer than 5–7-mer RNAs (Fig. 6.16B). This suggests that our smFRET assay is able to detect the synthesis of medium

(5–7-nt) and long (>7-nt) abortive RNA products. It will be interesting to perform similar experiments in the presence of Gre factors to see what effect these regulators have on the propensity for RNAP to escape the promoter.

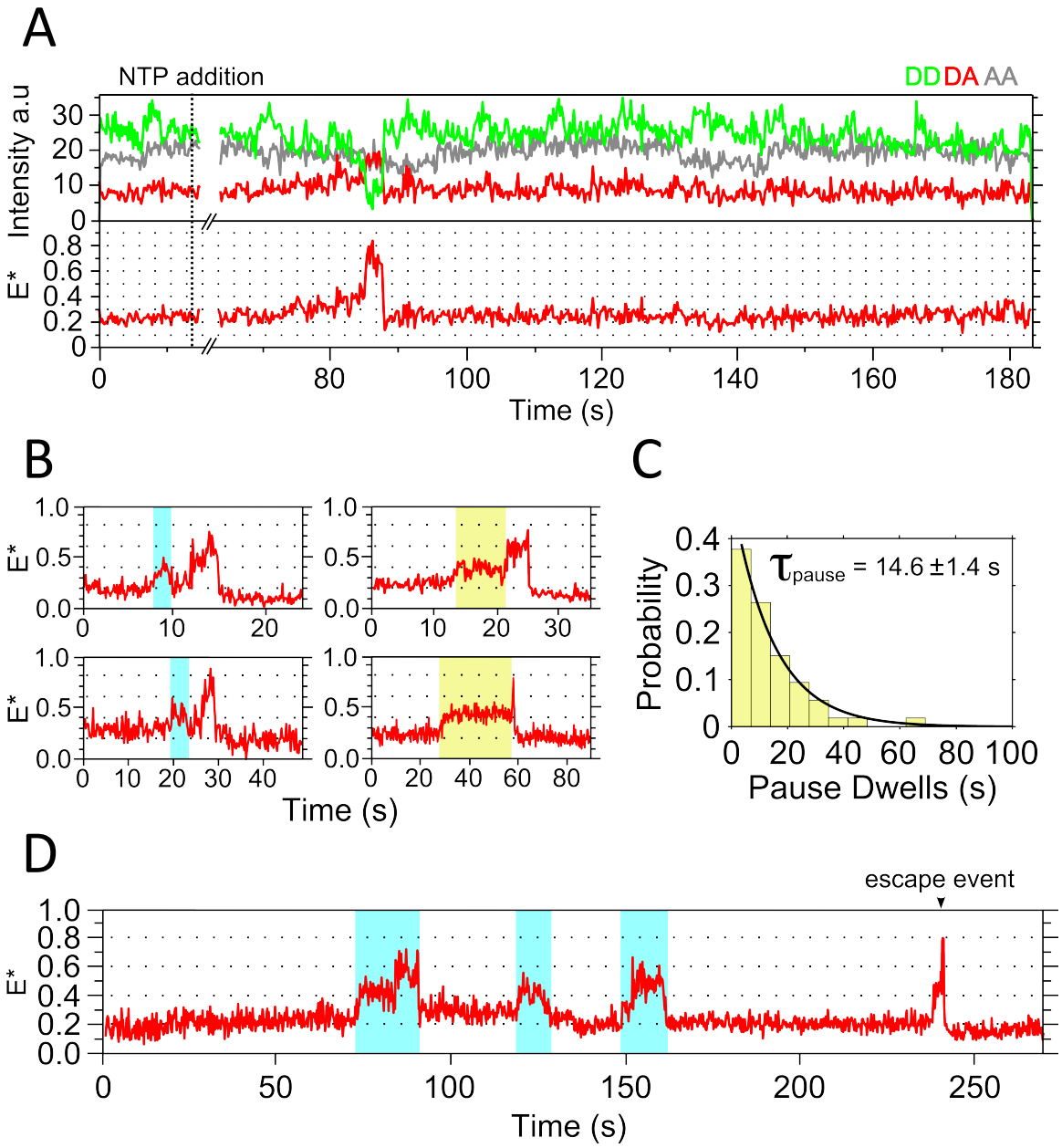

Figure 6.14. Promoter escape in real-time. (A) Example time trajectory showing FRET signal changes attributed to promoter escape (see main text). No activity is observed after the escape event. (B) Example time trajectories showing common escape behaviours. Left hand figures show examples in which escape is preceded by a clear single abortive cycle highlighted in blue. The right hand figures show examples of escape events preceded by a pause at E*~0.38 highlighted in yellow. (C) Dwell time distribution of the pauses shown in B with the mean dwell time shown inset. 130 events from 130 molecules are shown in the histogram. (D) Example time trajectory of a promoter

escape event preceded by three abortive cycles (highlighted with a blue background). Temporal resolution is 200 ms.

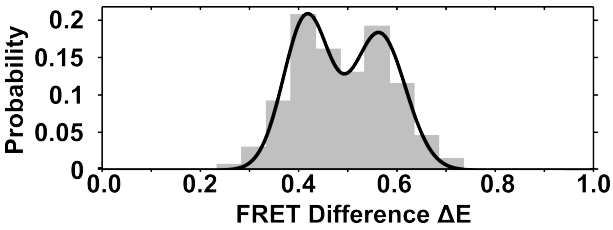

Figure 6.15. Variability in the promoter escape position. Histograms showing the ΔE values calculated for promoter escape events. The baseline and E^*_{escape} values were obtained from HMM fits to the data. 100 events are shown in the histogram from 100 molecules.

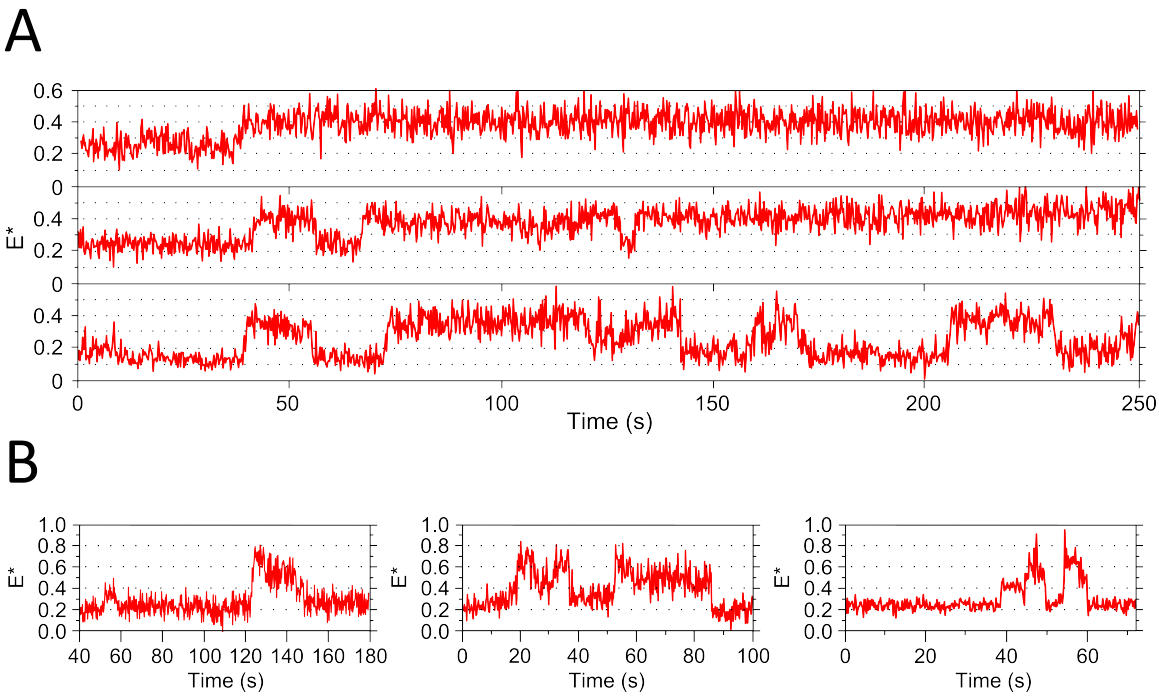

Figure 6.16. RNAP molecules unable to escape the promoter. (A) Example time trajectories of molecules showing similar behaviours in promoter escape experiments as that seen for earlier $RP_{ITC\leq7}$ experiments (see main text). (B) Example time trajectories of molecules showing abortive behaviour not seen previously. These examples show abortive cycles that appear to produce longer RNA products than those seen in $RP_{ITC\leq7}$ experiments. Temporal resolution is 200 ms.

6.4 Discussion

6.4.1 Designing a DNA substrate to monitor initial transcription in real-time

In this chapter, we monitored abortive initiation and promoter escape in real-time. We achieved this by first optimising our labelling strategy in light of unexpected

results when using a *lac* promoter DNA fragment labelled with a donor dye at -15 and an acceptor dye at +15 (lacCONS+2(-39/+25)[18]). In contrast to an earlier study carried out with a shorter DNA substrate labelled at the same positions (lacCONS+2(-39/+15)), we observed a decrease in FRET when RP_O was provided with NTPs to form $RP_{ITC\leq7}$ (Fig. 6.2). To identify possible structural explanations for this discrepancy in results, we modelled the accessible volumes (AV) of the fluorescent dyes attached to the current lacCONS+2(-39/+25) DNA and to the earlier lacCONS+2(-39/+15) DNA (Fig. 6.3). Our models showed that when the acceptor dye is internally labelled at position +15 (lacCONS+2(-39/+25)), DNA scrunching causes the dye to rotate around the DNA helical axis and move away from the donor (Fig. 6.3B). By contrast, when the acceptor dye is attached to the downstream blunt end of the substrate DNA (lacCONS+2(-39/+15)), a slight decrease in the inter-fluorophore distance is expected. This would provide an explanation for why lacCONS+2(-39/+15) showed an increase in FRET and lacCONS+2(-39/+25) showed a decrease in FRET in Fig. 6.2.

In light of these observations, we decided to explore alternative labelling schemes by modelling the acceptor dye at a number of possible attachment sites (Fig. 6.4). Our models showed that the +20 position was expected to give the largest FRET change during initial transcription (Table 6.1). In addition, this labelling site was expected to mask FRET changes due to transcription bubble dynamics (see chapter 5) that might have complicated our analysis of abortive cycling. We tested the predictions made by our *in silico* RNAP-promoter DNA models by performing experiments using a *lac* promoter DNA fragment labelled at position -15 with a donor dye and at position +20 with an acceptor dye. As expected, RP_O showed no difference in FRET compared with free DNA (Fig. 6.5B). Upon the formation of $RP_{ITC\leq7}$ using this labelling scheme, we saw an increase in FRET, which agreed reasonably well with the expected FRET change based on our structural models (Fig. 6.5B). We therefore proceeded with experiments designed to observe RNA synthesis in real-time. We note that the fact that we see an increase in FRET with the acceptor dye at +20 and a decrease in FRET with the acceptor dye at +15

[18] The numbers in parentheses indicate the limits of the substrate DNA. This DNA ranges from -39 to +25 relative to the transcription start site.

supports earlier reports showing that DNA scrunching is accompanied by rotational motion of downstream DNA (200).

6.4.2 RNA synthesis observed in real-time

We observed an increase in FRET upon the addition of NTPs to immobilized RP_O that we attribute to DNA scrunching during RNA synthesis (Fig. 6.7). In order to make structural interpretations of the different behaviours observed, we calibrated our FRET signal by performing a series of experiments in which we controlled the maximum length of RNA that could be synthesised by varying the NTP availability. Using this approach, we were able to assign RP_{ITC4}, RP_{ITC5}, and RP_{ITC7} states to E*=0.30, E*=0.35, and E*=0.45, respectively (Fig. 6.8).

In our real-time experiments we found that although the majority of immobilized active complexes began synthesising RNA within ~3 sec of the NTP addition point, synchrony was quickly lost and followed by molecules displaying a variety of behaviours. We classified these behaviours into a stably scrunched group and a cycling group. Stably scrunched molecules adopted scrunched conformations that lasted for more than 120 sec, while cycling molecules exhibited clear RNA synthesis and release cycles during our observation periods (Fig. 6.10). We note that the stable scrunched complexes we observed may correspond to a stable transcription complex previously reported in the literature (201).

Our results also showed that complexes in a single field-of-view synthesised RNAs of different lengths. This inter-molecular heterogeneity was seen with both stably scrunched and cycling molecules. Intriguingly, cycling molecules often exhibited persistent behaviours in terms of the length of RNAs synthesised i.e. a single immobilized complex appeared to iteratively synthesise RNAs of the same length. Having said this, cycling molecules did show events in which they appear to synthesise RNAs of a different length to those they most frequently synthesised. Examples of these types of events are marked with asterisks in Fig 6.10A. However, it should be noted that in general, our noise did not permit us to easily distinguish between noise fluctuations and

the synthesis of RNAs shorter than 5-nt in length. Increasing the laser power is an option for future studies looking specifically at the cycling kinetics of very short (3–4-nt long) RNAs; however, we expect these experiments to be technically challenging as the fluorophore bleaching rates will no doubt increase at higher excitation powers.

We note that similar behaviours were observed in abortive initiation experiments using an alternative labelling scheme in which the σ^{70} subunit was labelled with a donor dye and the DNA +20 position was labelled with an acceptor dye (see *Appendix B*).

6.4.3 Extracting abortive cycling kinetics

We performed dwell-time analysis of the scrunched conformation in $RP_{ITC\leq7}$ experiments by applying a Hidden Markov Model to all $RP_{ITC\leq7}$ cycling data. Our results showed that the scrunched conformation follows biexponential decay kinetics with mean dwell times of 10±1.8 s and 39±12 s (Fig. 6.10B). Control experiments testing whether the mismatch in the bubble region or the immobilization strategy were responsible for the observed heterogeneity showed the same overall behaviour as that seen with Ab-immobilized RNAP-promoter pmDNA complexes (Fig. 6.11). Interestingly, we observed rare instances (<1% of actively transcribing complexes) where molecules appeared to display changes in behaviour during our observation period (see *Appendix B*). This observation leads us to speculate that the inter-molecular heterogeneity observed in our experiments may be due to reversible conformational differences between complexes that may underpin a mode of transcriptional regulation. If, for instance, regulatory molecules or different promoter sequences influence the proportions of molecules exhibiting certain behaviours, the propensity to produce full-length RNA products might be altered. This is highly speculative at the moment; however, future studies investigating whether regulatory factors cause these types of changes should now be relatively straightforward to perform.

6.4.4 RNAP pauses and backtracks in initial transcription

In addition to the stably scrunched and cycling behaviours described above, we also observed more complex behaviours for $RP_{ITC\leq7}$ experiments. Firstly, the majority of molecules in these experiments did not reach E*=0.45, which was the FRET value assigned to the RP_{ITC7} state. Instead, most molecules reached E*=0.35. This observation led us to hypothesize that under $RP_{ITC\leq7}$ conditions, RNAP rarely produces 7-mer RNAs. This was subsequently confirmed by IVT reaction gel studies (Fig. 6.6). Secondly, we observed clear pausing events at E*=0.35 en route to E*=0.45, which were quickly followed by either backtracking or RNA release events (Fig. 6.9A–B). Less than 5% of all molecules that reached E*=0.45 showed dwell-times that lasted longer than 120 sec at this FRET state. These observations indicate that the E*=0.35 pause state is relatively stable compared with the E*=0.45 state. Based on these observations we propose that there is a barrier that stalls the polymerase at the RP_{ITC5} state and impedes its progression forwards (section 6.4.5 discusses the nature of this barrier).

Our observations have implication for the possible RNA release mechanism during abortive initiation. The backtracked state we observe is expected to have the 3'-end of the RNA peeled away from the DNA-RNA hybrid and inserted into the NTP-entry channel i.e. it is an intermediate in the RNA release process. The fact that we observe such an intermediate provides support for a model in which abortive RNA release occurs through a backtracking mechanism via the NTP-entry channel (54).

6.4.5 σR3.2 as a barrier to RNA extension

σR3.2 forms a highly conserved flexible loop referred to as the σR3.2 finger that interacts with the template strand at the RNAP active site and partially occupies the RNA exit path (16, 26, 72). The tip of the loop has negatively charged residues and is known to participate in initial transcription (17, 60); however the function of σR3.2 is not yet fully understood.

As mentioned in section 6.4.4, our single-molecule data showed that RNAP under

$RP_{ITC\leq7}$ conditions rarely produces 7-mer RNA products. This led us to hypothesise the existence of a barrier that impedes RNAP progression beyond the RP_{ITC5} state. In our attempt to identify possible candidates for this barrier, we produced structural models of *E. coli* initial transcribing complexes, which indicated that a physical clash would be expected to occur between the σR3.2 finger and the 5'-end of nascent RNAs 5–6-nt in length (Fig. 6.12). A recent crystal structure of a *T. thermophilus* initial transcribing complex has also shown a clash occurring between RNAs of similar lengths and the σR3.2 finger (26), thus confirming the validity of our structural model. We hypothesized that the σR3.2 finger constitutes this barrier to RNA extension and tested this hypothesis by performing $RP_{ITC\leq7}$ experiments using RNAP lacking the σ3.2 finger (Δ3.2).

Our Δ3.2 data showed an increased tendency for mutant complexes to reach E*=0.45, and a reduced tendency for mutant complexes to pause at E*=0.35 compared to wt complexes. This suggests that RNAP pausing at E*=0.35 is largely due to the σR3.2 finger blocking RNA extension. This conclusion is supported by IVT reaction gel-studies that show an increased production of RNAs >5-nt in length for Δ3.2 complexes relative to wt complexes (Fig. 6.13).

Interestingly, our time trajectory data also revealed that the fraction of stable scrunched complexes was smaller for Δ3.2 complexes relative to wt complexes i.e. RNA release was more likely for mutant complexes. We quantitatively assessed the stability of the scrunched state of Δ3.2 complexes by extracting the dwell times of the scrunched conformations (Fig. 6.13). Our results showed that, in contrast to the wt polymerase, the Δ3.2 mutant scrunched states exhibited monoexponential decay kinetics due to a decrease in the number of synthesis events with long scrunched dwell times. These findings suggest that the σR3.2 finger contributes to the stability of the scrunched conformation. However, it should be noted that the stable scrunched population was not totally eliminated when the σR3.2 finger was deleted. This suggests that other factors also play a role.

Collectively, our results suggest that the σR3.2 finger is involved in stabilising short RNAs at the active site during initial transcription, and also acts as a barrier to the extension of RNAs longer 5-nt in length. These conclusions are supported by reports in

the literature (60, 181, 202). One possible explanation for these apparently contradicting functions is that in order for the σR3.2 finger to stabilise short RNAs at the active site it has to be physically present there, and must be removed at some stage in the reaction for RNAP to continue progressing forwards. The displacement of the σR3.2 finger therefore competes with RNA extension and impedes progression of RNAP beyond the RP_{ITC5} state. Another possibility is that by creating a barrier to DNA scrunching, σR3.2 is able to act as a spring and accumulate stress that may drive the transition from initiation to elongation. This possibility is supported by studies showing that deleting the σR3.2 finger hinders promoter escape (60). These two scenarios are not mutually exclusive.

The exact mechanism by which the σR3.2 finger stabilises short RNAs at the active site remains unclear. One possibility is that once a 5–6-nt long nascent RNA is synthesised, it forms direct interactions with the σR3.2 finger that stabilise it at the active site. Recent structural studies have confirmed that close contacts are indeed made between the σR3.2 finger and the 5'-end of the nascent RNA (26); however, they do not reveal any specific contacts made between the two elements. The fact that sequence comparisons show that the σR3.2 finger sequence is not as highly conserved across species as the negative charge is (60), suggests that non-specific electrostatic interactions may play a role here. A separate possibility is that changes in transcription bubble dynamics and/or stability upon deleting the σR3.2 finger negatively influence the stability of the nascent RNA at the active site. Indeed, in the previous chapter we showed that deleting the σR3.2 finger increases the kinetics of transcription opening and closing, and we hypothesized that this change in kinetics may affect later steps in the transcription initiation pathway. In this scenario, interactions between the template strand and the σR3.2 finger would be responsible for stabilising short nascent RNAs at the active site.

We highlight that the importance of the role played by the σR3.2 finger is evident in the fact that its function is well conserved from bacteria through to eukaryotes, where the TFIIB transcription factor B-finger is believed to play similar roles as the σR3.2 finger in bacteria (117).

6.4.6 Pausing and backtracking: an additional explanation

As an RNA molecule is synthesised during transcription initiation, there is a characteristic probability for RNA release with each nucleotide addition cycle (84). This probability is partly governed by the underlying sequence of the promoter elements and the initial transcribed sequence (84, 109). This fact, combined with the following two observations: 1) Δ3.2 complexes show only minor improvements in the synthesis of RNAs >5-nt in length (Fig. 6.13), and 2) pausing is not completely eliminated in Δ3.2 complexes, suggests that other factors such as the sequence of the promoter DNA fragment may also explain why RNAP stalls at RP_{ITC5}. A recent *in vivo* and single-molecule study has identified a consensus pause sequence (GGCATAATTGTGGCCG) that displays remarkable similarities to our initial transcribed sequence (5'-AATTGTG-3') (131). We speculate that our initial transcribed sequence may also act as a pause sequence that impedes RNAP from progressing beyond the RP_{ITC5} state, and may have some regulatory role in initiation. Future work will investigate this hypothesis through promoter DNA mutational studies.

6.4.7 Direct observations of promoter escape

In our final set of experiments we sought to observe promoter escape in real-time by monitoring the behaviour of immobilized complexes upon the addition of all four NTPs to the observation chamber. Our results revealed a pattern of FRET signal changes that were unique to promoter escape experiments and could be explained by conformational changes that would be expected to occur during promoter escape (Fig. 6.14). We observed an increase in FRET due to scrunching of downstream DNA as RNA was synthesised; this continued until a maximum FRET value (E^*_{escape}) was reached that ranged between E*=0.6 and E*=0.8. Once E^*_{escape} was reached, we observed a rapid decrease in FRET to the baseline level. We believe that the maximum FRET value is reached when the initial transcribing complex is maximally scrunched, and therefore highly stressed. We propose that the accumulated stress in the maximally scrunched

complex drives the rapid release of promoter contacts and propels the polymerase forwards while dissipating the accumulated energy stored in the scrunched complex. Interestingly, 50% of escaping events exhibited a pause at E*=0.35–0.4 prior to reaching E^*_{escape}. Although the function of the pause is not yet understood, it is possible that it may result from the clash between RNA and the σR3.2 finger. We speculate that this clash may lead to the accumulation of strain in the σR3.2 finger that drives promoter escape. In this model, the σR3.2 finger would act like a spring that is ultimately displaced by the growing nascent RNA chain; this displacement would weaken σ^{70}-promoter DNA contacts (particularly σR2 contacts with the -10 element) and drive escape. This latter scenario is supported by a recent report suggesting that the σR3.2 finger is the main driver of promoter escape while strain accumulated in the scrunched DNA plays only a minor role (116).

An alternative explanation is that the pause may provide a window for regulatory molecules to modulate the efficiency of promoter escape. In this scenario the pause might be caused by the initial transcribed sequence, which displays similarities to the consensus pause sequence identified in (131), and would provide a window of opportunity for regulatory molecules to modulate the behaviour of RNAP. Future work is needed to determine the exact cause and function of the pause associated with promoter escape events.

6.4.8 Abortive cycling does not appear to be a pre-requisite for promoter escape

Our smFRET assay allowed us to observe both abortive cycling and promoter escape in a single experiment, which in turn allowed us to address the question of whether or not abortive cycling is obligatory for promoter escape. Our results suggest that iterative synthesis and release of RNAs >4-nt in length is not an obligatory process that precedes promoter escape. However, our current level of noise makes it is technically difficult for us to eliminate the possibility that 2–4-mer RNAs are synthesised and released before the polymerase escapes the promoter. Having said this, there are clear examples in which the immobilized complex immediately escapes the promoter

upon the addition of NTPs, leaving very little time for any abortive cycles (Fig. 6.14B). This suggests that abortive cycling is not an obligatory intermediate for promoter escape.

35% of all escape events in our experiments do follow a period of abortive initiation. In the vast majority of cases, the polymerase performed 1–2 abortive cycles before escaping the promoter (Fig. 6.14B). In fact, less than 10% of all escape events occur after the polymerase has gone through ≥3 abortive cycles (Fig. 6.14D). It is unclear what the origin of these differences in behaviour is; however, it is tempting to speculate that it is due to different conformational states of the polymerase. Future studies will work to address the statistical and then functional significance of this heterogeneity.

6.4.9 RNAP-promoter DNA complexes can be trapped in abortive initiation

In addition to a subpopulation of molecules undergoing promoter escape, we also observed molecules that appeared to be trapped in abortive initiation, at least for the duration of our observation period (Fig. 6.16). The behaviour of these molecules was qualitatively the same as that seen in our earlier abortive initiation experiments; however, we also observed novel abortive cycling behaviours in which scrunched complexes reached values as high as E*=0.6. Presumably, these events represent the synthesis of RNAs that are longer than 7-nt in length.

The *lac* consensus promoter we work with forms strong contacts with RNAP holoenzyme that make it deficient in promoter escape (10). This leads us to speculate that it is the strong contacts formed between the polymerase and the promoter DNA that is responsible for the subpopulation of molecules seemingly trapped in abortive initiation. However, it is unclear why some RNAP molecules escape easily and others do not when all are bound to the same promoter. Future studies will focus on the staistical significance of the difference observed (see Chapter 8). We note, however, the possibility that conformational differences in the RNAP may influence the likelihood that it will escape the promoter. In this scenario, RNAP would adopt multiple states that have different efficiencies in promoter escape. Perhaps our abortive fraction of molecules

represents moribund complexes that may be responsive to regulatory factors (100). Future investigations will aim to address these and other questions raised by the work presented in this chapter.

6.5 Conclusion

In this chapter we demonstrated our ability to monitor the RNA synthesis reaction in real-time. Our work reveals the detailed DNA conformational changes that occur during initial transcription, and in the process, begins to address a number of questions that have remained unanswered for many years in the field of transcription initiation. Our work has shed new light into the mechanism of abortive initiation and promoter escape, and indicates that abortive initiation is caused by a combination of factors that extend beyond simply the strength of RNAP-promoter DNA interactions, or one specific protein structural element. Questions raised by the work presented in this chapter can now be tackled using the smFRET assay we have demonstrated here. It is our hope that the work presented in this chapter will drive future investigations in this field of research.

6.6 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. Laurent Fernandez (Oxford University, UK) assisted in performing experiments. Geraint Evans (Oxford University, UK) assisted with data visualisation and provided useful discussions on data analysis methods. Alexandra Tomescu (Oxford University, UK) performed all *in vitro* transcription reactions.

Chapter 7: RNAP clamp dynamics

7.1 Introduction

The RNAP β' clamp is an important mobile structural element that has been captured in different conformations using X-ray crystallography (23, 25, 49, 118). This apparent conformational plasticity led to initial proposals that clamp mobility plays an important functional role in transcription (121). However, considering the possibility that crystal packing forces might have influenced the β' clamp conformation, it was unclear what the functional relevance of the different conformations observed *in crystallo* were (see chapter 2 for a detailed discussion of the clamp).

Recently, a smFRET study that monitored the clamp conformation through fluorophores attached to the β' clamp and β pincer (Fig. 7.1), confirmed that the clamp is able to adopt different conformations during the transcription cycle in solution (27). This study showed that the clamp adopts multiple conformations in holoenzyme, and a closed conformation in both the open complex (RP_O) and in the elongation complex (RD_e). This led the authors to propose a model in which clamp closure stabilises RP_O, and allows RDe to transcribe a gene with high processivity. This study also showed that antibiotics that target the switch region – an area located at the base of the clamp – are able to modulate the clamp conformation. Based on this observation, the authors proposed that the mechanism of action of a number of switch-region-target inhibitors is to impede functionally important clamp conformational changes.

This smFRET study provided a wealth of new insight into the role played by the clamp in transcription; however, because experiments were performed using confocal spectroscopy of diffusing molecules, the observation windows were limited to ~1 ms. This study could therefore not tell whether the clamp undergoes dynamic opening and closing transitions at different stages of transcription, or is largely an immobile domain. In this chapter, we use an antibody (Ab) immobilization technique (see chapter 4) combined with single-molecule TIRF microscopy to monitor the clamp conformation in

real-time at different stages of transcription. The work presented in this chapter was carried out in collaboration with the authors of (27) who provided the labelled RNAP shown in Fig. 7.1.

Our results are the first to show that the clamp undergoes dynamic opening and closing transitions in the holoenzyme state. Our results also reveal considerable heterogeneity in the clamp conformation reminiscent of that observed for the DNA polymerase I Klenow fragment fingers subdomain (see chapter 4). This observation leads us to speculate that static and dynamic heterogeneity may be a common feature of polymerase mobile domains.

Our results confirm many of the observations made earlier using confocal spectroscopy, and also provide new insight into the mechanism of action of switch-region-target inhibitors. The results presented in this chapter complement those in chapters 5 and 6, and we believe considerably advance our understanding of the transcription initiation mechanism.

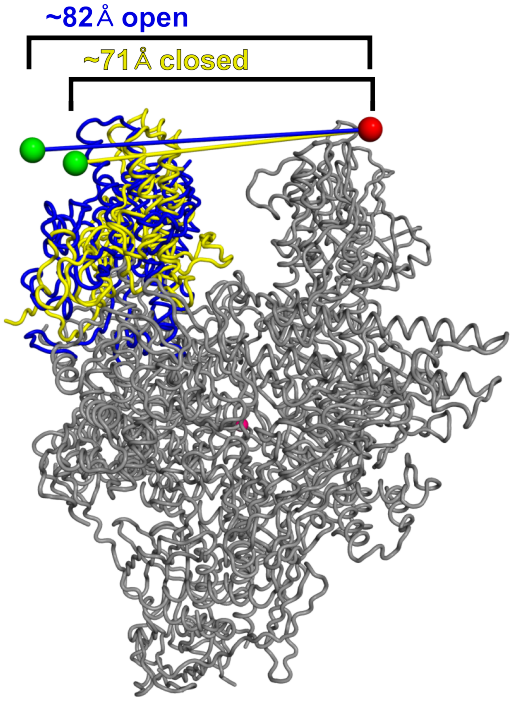

Figure 7.1. Labelling scheme used to monitor the clamp conformation. Structure of *E. coli* RNAP labelled with a donor dye (Cy3B, green) at position 284 on the β' subunit, and an acceptor dye (Alexa647, red) at position 106 on the β subunit. The clamp open (blue) and closed (yellow) conformations (PDB files 1I3Q and 4IGC, respectively) are shown. The remainder of the protein is shown in grey. σ^{70} is omitted for clarity. The Mg^{2+} ion at the active site is shown as a pink sphere. The inter-fluorophore distances in the open and closed clamp conformations are shown and labelled in blue and yellow, respectively. Dye positions were approximated using accessible volume modelling (see *Materials and Methods*). Figure adapted from (27).

7.2 Materials and Methods

7.2.1 DNA and protein preparation

Oligonucleotides were purchased from IBA Life Sciences (Germany) and annealed as described in chapter 5. The DNA sequences used in this chapter are shown in Fig. 7.2. RNAP holoenzyme labelled with Cy3B at position 284 on the β' subunit, and Alexa647 at position 106 on the β subunit was provided by our collaborator Dr. Richard Ebright (Rutgers, USA); fluorescently labelled holoenzyme samples used throughout this chapter were prepared as described in (27).

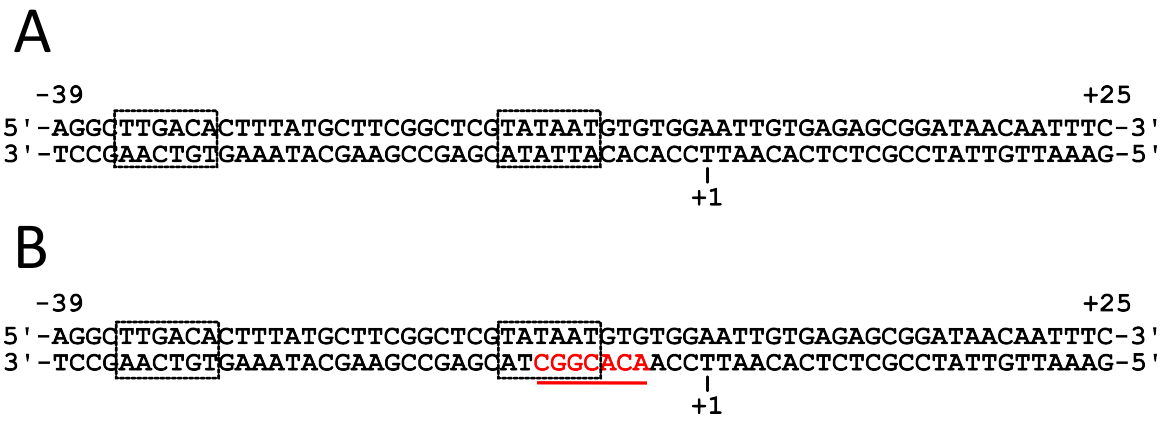

Figure 7.2. DNA sequences used to form RP_O. (A) lacCONS+2 promoter DNA fragment used in this chapter. (B) lacCONS+2 promoter DNA fragment with a mismatch in the transcription bubble region shown in red and underlined in red. The black boxes indicate the -10 and -35 elements. The transcription start site is labelled +1.

7.2.2 RP_O formation

RP_O was formed as described in chapter 5 except with the following minor modification: the substrate DNA was added to a final concentration of 250 nM instead of 10 nM; the RNAP concentration was unchanged. This modification was made to minimise the likelihood that we would immobilize RNAP not bound promoter DNA in RP_O experiments since in these experiments we are not able to directly monitor the presence of the unlabelled promoter DNA.

7.2.3 Antibody-immobilization of RNAP

A PEG-passivated glass surface was functionalized with penta-His antibody (Ab) as described in chapter 4. Then, a 100 pM solution of fluorescently labelled RNAP (as holoenzyme or pre-formed RP_O) was incubated on the surface for 5 min. Binding was monitored until ~60 molecules were deposited on the field of view surface. Excess unbound RNAP molecules were washed away.

7.2.4 Instrumentation

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

7.2.5 Single-molecule experiments

For RNAP holoenzyme and RP_O experiments, T8 imaging buffer (see chapter 5) was added to the observation chamber. For $RP_{ITC\leq7}$ and RD_{e11} experiments, T8 imaging buffer supplemented with 500 µM ApA dinucleotide was added to the observation chamber. NTP reaction mixtures were then added to the observation chamber manually during data acquisition by pipetting 3 µL into the well. For $RP_{ITC\leq7}$ experiments, the reaction mixture consisted of imaging buffer supplemented with UTP and GTP at 80 µM. For RD_{e11} experiments, the reaction mixture consisted of imaging buffer supplemented with UTP, GTP, and ATP to achieve [UTP/GTP]=100 µM and [ATP]=200 µM. The CRIFF autofocusing system was used during data acquisition to ensure focal stability. Temporal resolutions are given in relevant figure captions.

Lipiarmycin was purchased from BioAustralis, Australia. Myxopyronin was provided by our collaborator Dr. Ebright who prepared it as described in (27). For experiments assessing the effect of these inhibitors on the clamp conformation, 20 µM Lipiarmycin or Myxopyronin was added to 50 nM RNAP holoenzyme and incubated for 10 mins at 37°C. RNAP was then immobilized in T8 imaging buffer supplemented with 20

μM of the appropriate inhibitor. These experiments were performed at a temporal resolution of 20 ms. For real-time inhibitor binding experiments, RNAP holoenzyme was immobilized as described above, and then the inhibitor manually added to the observation chamber during data acquisition by pipetting 3 μL into the well to achieve a final concentration of 20 μM. Real-time inhibitor binding experiments were performed at a temporal resolution of 200 ms. All experiments were performed at 21°C. 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.

7.2.6 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 100 molecules for analysis according to the criteria described in chapter 5. All E* histograms were constructed using 250 frames of data from 100 selected molecules. 38% of spots passed all selection criteria.

We manually classified the intensity time trajectories of individual RNAP-promoter DNA complexes molecules into groups. Static open clamp signals were defined as time trajectories showing a stable E*~0.23 signal with no fluctuations to higher FRET states during the observation period. Static closed clamp signals exhibited a stable E*~0.42 signal with no fluctuations to lower FRET states. Dynamic clamp signals were defined as time trajectories showing any anti-correlated E* changes during the observation periods. Molecules exhibiting confounding photophysical fluctuations were excluded from analysis.

7.2.7 HMM Analysis of dynamic smFRET time trajectories

HMM analysis was performed on time trajectories displaying dynamic FRET fluctuations using custom-written MATLAB software described in chapter 5. Each time

trajectory was fitted with 1 to 2 states and the best model (number of states) was selected automatically using maximum evidence criteria (see chapter 5). The two possible states represent the two populations seen in our E* histograms (see *Results*). The states extracted by HMM were categorized according to their FRET efficiency and the states that preceded them; low FRET efficiency states were classified as "open clamp states", and high FRET efficiency states were classified as "closed bubble states". The dwell time distributions for each state category were fitted with an exponential decay curve as described in chapter 5. The mean open clamp dwell times (τ_{open}) and mean closed clamp dwell times (τ_{closed}) were extracted from the fits for the open and closed clamp states, respectively.

7.2.8 Accessible volume (AV) modelling

The *E. coli* RP_O model described in *Appendix A* was used for AV modelling of the donor and acceptor dyes in the open and closed clamp conformations shown in Fig. 7.1. The accessible volumes of the donor and acceptor dyes attached to amino acid Cβ atoms via a flexible linker were calculated according to (197) and as described in chapter 6. Visualisations and distance measurements between the donor and acceptor average dye positions were performed in Pymol.

7.3 Results

7.3.1 Clamp conformations in RNAP holoenzyme

A previous smFRET study revealed that the clamp in RNAP holoenzyme adopts multiple conformations (27). However, as this study was performed using confocal spectroscopy, it remained unclear whether the heterogeneity observed was due to: 1) subpopulations of the RNAP clamp adopting different static conformations, or 2) RNAP molecules in which the clamp undergoes dynamic conformational changes. In an attempt to address this unknown, we immobilized fluorescently labelled RNAP

holoenzyme (Fig. 7.1) on an Ab-coated surface and visualised immobilized molecules for many seconds (~20s at a temporal resolution of 20ms) using TIRF microscopy. Control experiments showed minimal non-specific adsorption of RNAP to a PEG-passivated surface lacking the penta-His antibody (Fig. 7.3).

One hundred immobilized RNAP holoenzyme molecules were selected for analysis according to the criteria outlined in *Materials and Methods*, and used to generate a FRET (E*) histogram, which showed two populations centred at E*=0.22 and E*=0.41 (Fig. 7.4A). We assigned these populations as the open and closed clamp conformations, respectively. Repeat experiments showed that the open clamp population was consistently more abundant than the closed clamp population.

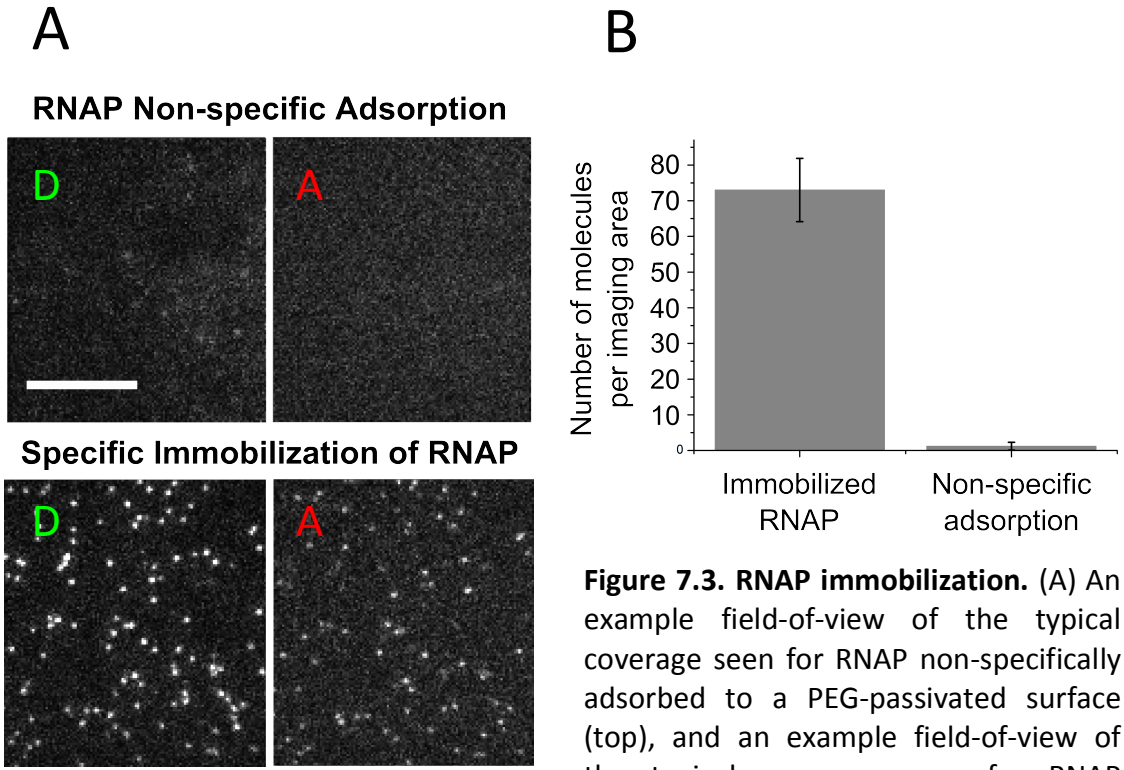

Figure 7.3. RNAP immobilization. (A) An example field-of-view of the typical coverage seen for RNAP non-specifically adsorbed to a PEG-passivated surface (top), and an example field-of-view of the typical coverage seen for RNAP specifically immobilized to a PEG-passivated surface via penta-His-antibody/His-tag interactions (bottom). The left image shows the donor channel, the right image shows the acceptor channel. Scale bar is 5 µm. (B) Bar graph with average number of fluorescent molecules found using TwoTone per imaging due to specific immobilization and non-specific adsorption. Error bars represent the standard deviation of the mean across 5 images.

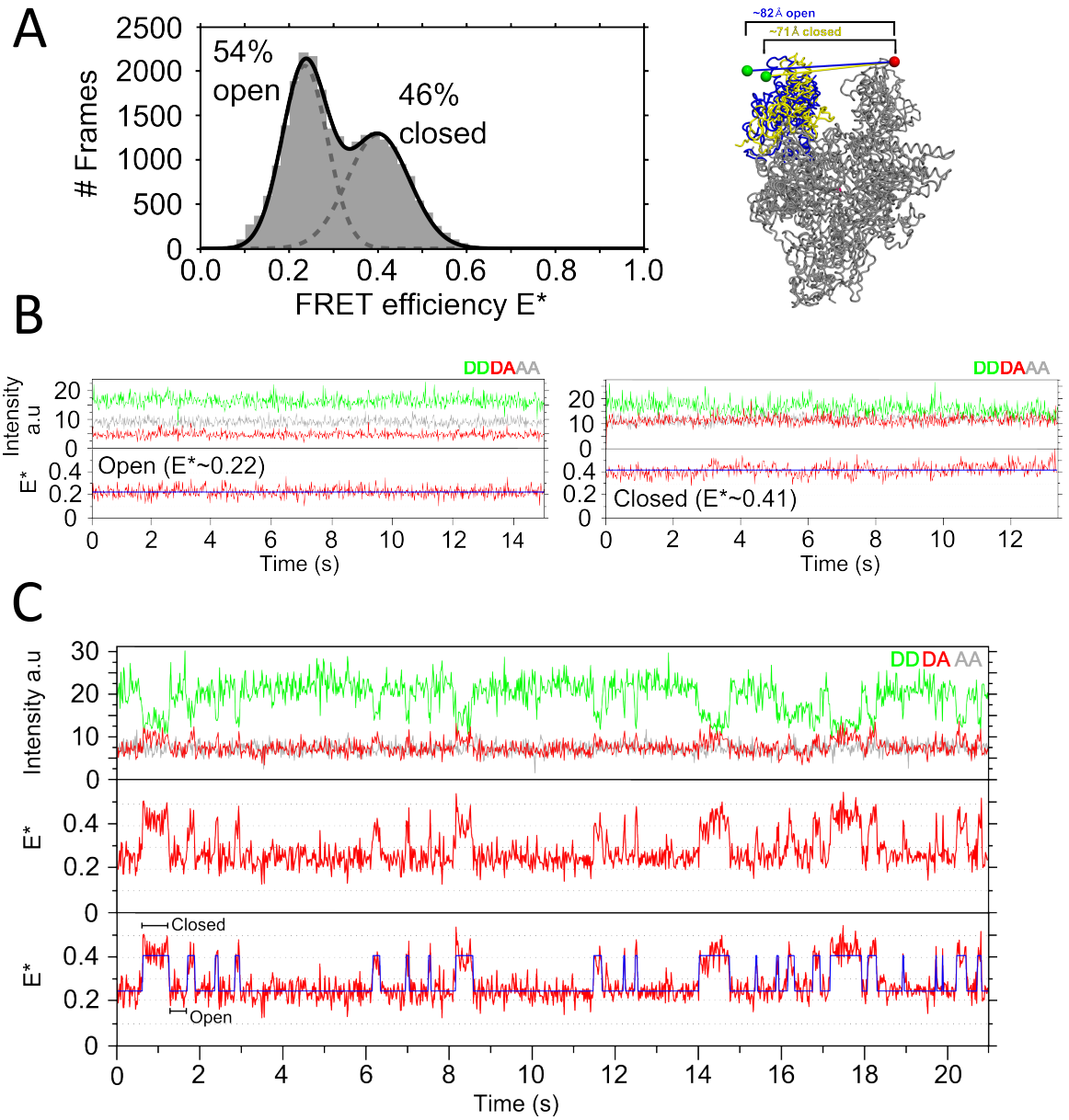

Figure 7.4. RNAP holoenzyme clamp dynamics. (A) RNAP holoenzyme E* histogram fit to a 2 Gaussian profile centred at E*=0.22 (open clamp) and E*=0.41 (closed clamp). The percentages of the populations are shown. The right hand figure shows the overlayed open and closed clamp conformations of *E. coli* RNAP from Fig. 7.1. 100 molecules are shown in the histogram. (B) Example time trajectories of RNAP molecules exhibiting a stable open clamp state (left) and a stable closed clamp state (right). (C) Example time trajectory of an RNAP molecule exhibiting clamp opening and closing dynamics. In figures B–C the top panel shows the fluorescence intensity of the donor under donor excitation (DD; green line), the acceptor under acceptor excitation (AA; grey line), and the acceptor under donor excitation (DA; red line). In (B), the bottom panels show the E* trace calculated from the fluorescence intensity. The blue line shows the mean E* value of the trace. In (C), the top panel shows the fluorescence intensity, the middle panel shows the E* trace, and the bottom panel shows the E* trace overlayed with the Hidden Markov Model fit shown in blue. Temporal resolution is 20 ms.

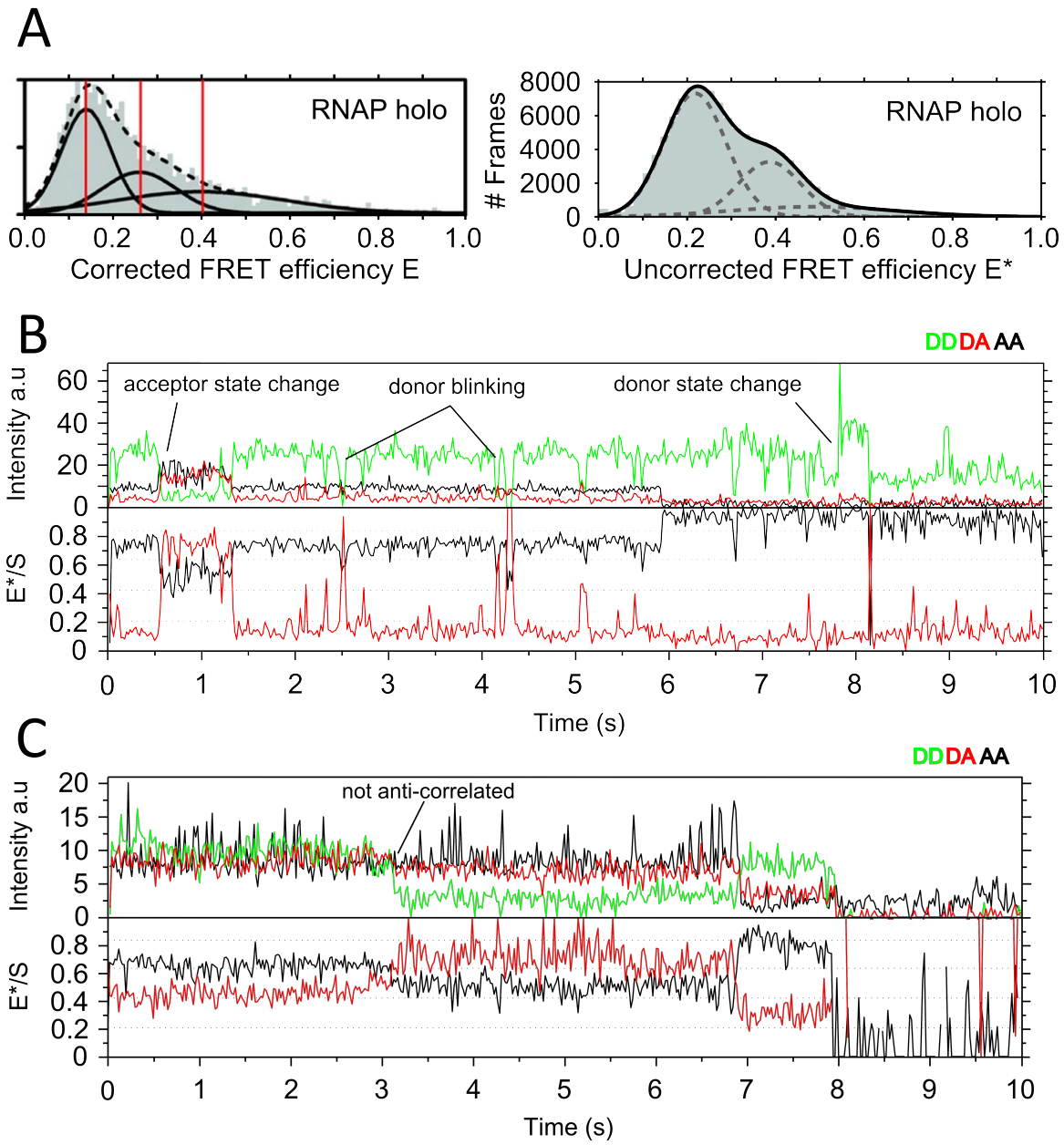

Figure 7.5. Accounting for different FRET distributions between TIRF and confocal data. (A) The corrected FRET efficiency E histogram for diffusing RNAP holoenzyme molecules from (27) is shown on the left. The uncorrected FRET efficiency E* histogram for 382 immobilized RNAP holoenzyme molecules before applying selection criteria (see main text) is shown on the right. Only the DA populations contribute to both histograms. (B–C) Example time trajectories showing behaviour that do not meet all selection criteria. The top panel has the same colour scheme as Fig. 7.4, except for AA, which is shown in black to be seen more clearly. The bottom panels show fret efficiency (E*) in red and stoichiometry (S) in black. (B) Shows examples of photophysical phenomena that induce high FRET states. (C) Shows an example of a non anti-correlated event that induces a high FRET state.

Our surface-based TIRF microscopy results were broadly in agreement with the previous solution-based confocal spectroscopy results (27). There was however, a slight discrepancy between the two in that the confocal E* histogram showed an additional population of molecules with FRET values above that of the closed clamp conformation (Fig. 7.5A, left). This population was assigned to a clamp collapsed state in (27), which has not been observed in any crystal structure to date.

We investigated possible explanations for this discrepancy and found that if we generated an E* histogram of RNAP holoenzyme molecules before selecting 100 molecules for analysis, we obtained a distribution that closely resembled the published confocal spectroscopy distribution (Fig. 7.5A, right)[19]. Furthermore, an inspection of individual time trajectories showed that molecules that contribute to the high FRET (E*>0.5) population in the TIRF histogram display behaviours that do not meet our selection criteria (see *Materials and Methods*) and are therefore not taken forward for further analysis (Fig. 7.5B–C). We propose that molecules that display aberrant photophysical behaviour and/or aggregates containing multiple fluorophores are the cause of the small high FRET population seen in Fig. 7.5A. As confocal spectroscopy studies do not permit the manual filtering of molecules in the same way in which TIRF studies do, it is likely that molecules that would not fit our selection criteria are included in the confocal histogram and contribute to the small high FRET population. We therefore conclude that the RNAP holoenzyme clamp adopts open and closed conformations only.

7.3.2 Clamp static and dynamic heterogeneity in RNAP holoenzyme

Our time trajectory data of individual RNAP holoenzyme molecules revealed that the clamp conformation exhibits both static and dynamic heterogeneity (Fig. 7.4B–C). Behaviours could be classified into three groups: 1) a stable open clamp population (E*~0.22) that rarely samples higher FRET states (Fig. 7.4B, left), 2) a stable closed clamp

[19] Differences in absolute FRET values can be attributed to the fact that the confocal histogram shows corrected FRET efficiencies while the TIRF histogram shows uncorrected FRET efficiencies.

population (E*~0.41) that makes rare transient transitions to the low FRET state (Fig. 7.4B, right), and 3) a dynamic population that undergoes numerous transitions between the E*~0.22 and E*~0.41 states (Fig. 7.4C). This is the first direct observation of clamp conformational dynamics in any transcriptional system, and confirms the long-standing hypothesis that the RNAP clamp is able to switch dynamically between open and closed states. We quantified the observed clamp dynamics by fitting a Hidden Markov Model (HMM) to the E* time trajectories of the dynamic subpopulation of molecules (Fig. 7.4C) to extract the dwell times of the open and closed conformations (Fig. 7.6). The dwell times of the two conformations were plotted in dwell time histograms and fit with a monoexponential function. The clamp was found to exhibit a slow closing rate of 3.6±0.4 s^{-1}, and a faster opening rate of 5.8±0.5 s^{-1}.

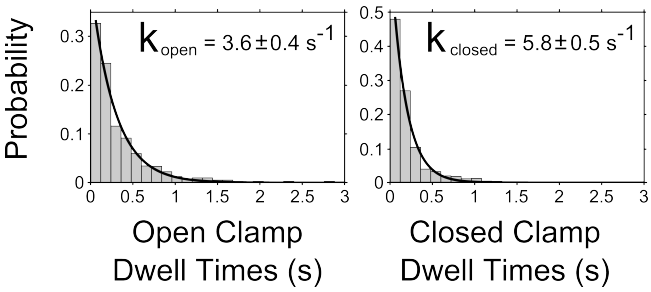

Figure 7.6. Holoenzyme clamp dynamics. Left hand figure shows the closed clamp dwell-time histogram (1052 events, 107 molecules). Right hand figure shows the open clamp dwell-time histogram (997 events, 107 molecules). The clamp opening and closing rates are shown inset.

7.3.3 Clamp conformations in RP_O

Next, we investigated whether the clamp undergoes conformational fluctuations in RP_O. We did this by imaging immobilized RP_O (formed *in vitro* using an unlabelled *lac* consensus promoter DNA fragment shown in Fig. 7.2A) using TIRF microscopy. The E* histogram showed that the RP_O clamp exists primarily in a closed conformation (Fig. 7.7A). This observation is in agreement with the previous confocal spectroscopy study (27); however, in contrast to this previous study, we observed a small population of RNAP molecules that have an open-clamp conformation. Individual time trajectories showed that the data frames corresponding to the E*=0.22 population originate from molecules with a stable open clamp state. No molecules were found to exhibit clamp opening and closing dynamics (Fig. 7.7B). We therefore conclude that the clamp is locked in a stable closed conformation in RP_O.

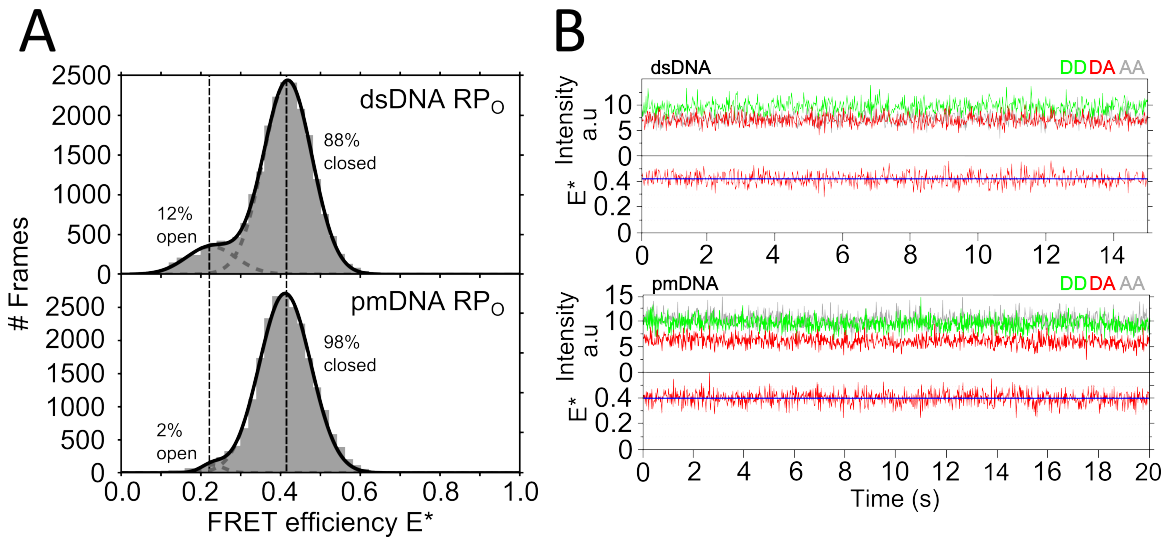

Figure 7.7. RNAP clamp conformation in RP$_O$. (A) E* histogram of RP$_O$ formed using dsDNA (top), and pmDNA (bottom). The vertical dashed lines indicate the centres of the open and closed clamp populations. Data for 100 molecules is shown per histogram. (B) Example fluorescence intensity and E* traces for dsDNA RP$_O$ (top figure), and pmDNA RP$_O$ (bottom figure). Temporal resolution is 20 ms.

We hypothesised that the presence of the E*=0.22 population in the RP$_O$ histogram could be explained by the following scenario. Immobilized RP$_O$ may have moved backwards in the RP$_O$ formation pathway and reached states in which the clamp adopts the open conformation, or even released the promoter DNA. To test whether this situation was a possibility, we formed RP$_O$ using pre-melted promoter DNA (pmDNA; sequence is shown in Fig. 7.2B). The mismatch in the pmDNA stabilises RP$_O$ (176, 189) and was therefore expected to slow the dissociation of DNA from the polymerase and reduce the size of the E*=0.22 population.

The FRET histogram for RP$_O$ formed with pmDNA shows a large E*=0.41 population and an E*=0.22 population that is smaller than that seen for fully complementary DNA (Fig. 7.7A). This suggests that the clamp in RP$_O$ exclusively adopts a stable closed conformation and exhibits no dynamic transitions between open and closed clamp states. This finding has implications for our understanding of the RP$_O$ formation mechanism, which is discussed further in section 7.4. We note however, that further investigations with fluorescently labelled DNA are needed to test our hypothesis.