Single-Molecule Studies of Transcription Initiation

Diego Armando Duchi Llumigusín

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy at the University of Oxford



St. Cross College
University of Oxford
Trinity Term 2014

Disclaimer

I declare that the work presented in this thesis is mine alone, expect where indicated in the main text, and as described below.

- Kristofer Gryte wrote the data visualisation software used to generate single-molecule fluorescence time trajectories
- A number of figures in chapters 6 and 7 were prepared using data visualisation software written by Geraint Evans
- Nicole Robb purified DNA substrates
- Alexandra Tomescu performed *in vitro* transcription reactions
- Laurent Fernandez assisted in the collection of data presented in chapter 6

These contributions are stated at the end of relevant chapters. In addition, because my work has involved help, feedback, discussions, and ideas from members of the Kapanidis group, I use the inclusive *we* to reflect that my achievement results from a collective effort.

Diego Duchi

September 2014

Acknowledgements

I dedicate this thesis to Catalina Cardona. I would not have gotten this stage without your love, support, and sacrifices.

I would like to thank Achillefs for his support during my application to embark on this DPhil, and for his support during these four years. A special 'thank you' goes to Finn. His support during my summer internship initiated my interest in scientific research and the Kapanidis lab.

Thanks to Pouya for keeping me laughing during my first year. For his unconditional help with all things biochemistry, thank you to Javier. Thanks to Nicole for her help with experiments, RNAP discussions, and of course, DNA preparation. For all things slides and gaskets, thank you Sarah, without you the Kapanidis lab would collapse! A huge 'thank you' goes to Laurent Fernandez for his assistance with experiments. His contribution accelerated the progress of my DPhil, and I am very grateful for his dedication. Thank you to the teams at Target Schools, OUSU, and the Oxford Student Consultancy. The work I undertook with you broadened my horizons and provided me with opportunities to recharge my batteries for the next problems awaiting me in my DPhil.

Thank you to Athan aka Kris. Your passion for personal development encouraged me to embark on new challenges and enter (kind of) the world of programming. Thank you for your friendship and support with understanding data. Thank you Muffy & Pacha for your friendship and distractions.

Finally, thank you to the other women in my life, 'mami' & 'abuelita'. Your determination in life has inspired me to always persevere. This attitude helped me a great deal during these four years. Thank you.

Abstract

Single-molecule Förster resonance energy transfer (smFRET) has emerged as an important tool for studying biological reactions. This thesis describes smFRET investigations into the mechanism of bacterial transcription initiation.

We developed protocols to immobilize RNAP-DNA initiation complexes using vesicles and antibodies. We used these techniques to show that the transcription bubble conformation in immobilized complexes exhibits inter-molecular heterogeneity. We observed large FRET changes that we attribute to transcription bubble opening and closing dynamics. We found that σ^{70} region 3.2 (σR3.2) influences the kinetics of the bubble dynamics, which supports proposals that σR3.2 interacts with the transcription bubble template strand.

We extended our investigations to RNA synthesis and were able to observe abortive initiation cycles directly. We observed RNAP pausing and backtracking for the first time in transcription initiation. We obtained data suggesting that σR3.2 stabilises short RNAs at the active centre and forms a barrier to the extension of RNAs longer than 5-nt in length.

We extended our abortive initiation assay to observe signal changes that we attribute to promoter escape. Our data revealed the number of abortive cycles that occur prior to escape, the kinetics of promoter escape, and pausing events that may have some regulatory function.

We investigated the conformational dynamics of the RNAP β' clamp and observed dynamic conformational changes between clamp-open and clamp-closed states. Our work confirms proposals that the clamp remains stably closed once the open complex (RP_O) is formed. We investigated what affect the antibiotics Myxopyronin and Lipiarmycin have on the clamp conformation. Our results revealed that Myxopyronin traps the clamp in a closed conformation, while Lipiarmycin traps it in an open conformation.

Overall, we made a number of novel observations that we believe advance our understanding of the mechanism of transcription. We hope that the discoveries reported here will direct future research efforts into RNAP function.

Publications arising from DPhil

N.C. Robb, T. Cordes, L.C. Hwang, K. Gryte, **Diego Duchi**, T.D. Craggs, Y. Santoso, S. Weiss, R.H. Ebright and A.N. Kapanidis (2013) "The Transcription Bubble of the RNA Polymerase–Promoter Open Complex Exhibits Conformational Heterogeneity and Millisecond-Scale Dynamics: Implications for Transcription Start-Site Selection", J Mol Biol; 425 (5), pp.875-885.

Diego Duchi, K. Gryte, N.C. Robb, K. Brodolin, S. Wigneshweraraj and A.N. Kapanidis (2014) "The Transcription Bubble of the RNA Polymerase–Promoter Open Complex Exhibits Conformational Static and Dynamic Heterogeneity", (Manuscript in preparation).

Diego Duchi, L. Fernandez, G. Evans, N.C. Robb, A. Tomescu, K. Gryte, L.C. Hwang, K. Brodolin and A.N. Kapanidis (2014) "Real-Time Observations of Abortive Initiation and Promoter Escape", (Manuscript in preparation).

Conference Presentations

26th Annual RNAP Workshop
Newcastle University, Newcastle, UK **April 2-3, 2014**
"Single-molecule studies of transcription initiation: monitoring real-time conformational changes": Presented research showing real-time abortive cycling and the role of σ^{70} region 3.2 in abortive initiation.

Mechanisms and Regulation of Prokaryotic Transcription
Vermont, US **June 23-28, 2013**
"Real-time observation of initial transcription and promoter escape at base-pair resolution": Poster presentation in which I reported novel observations for abortive initiation and promoter escape.

25th Annual RNA Polymerase Workshop
Milton Keynes, UK **March 18-19, 2013**
"Real time observation of transcription initiation dynamics": Presented preliminary results of the dynamics of abortive initiation.

24th Annual RNA Polymerase Workshop
Oxford University, Oxford, UK **March 19-20, 2012**
"Real-time observation of dynamics of the open and initial transcribing complex": Presented research demonstrating transcription bubble dynamics seen in transcription initiation complexes.

56th Annual Biophysical Society Meeting

San Diego, California, USA **Feb 25-29, 2012**

"Single-molecules studies of transcription initiation using a vesicle approach": Poster presentation in which I reported transcription bubble dynamics in transcription initiation complexes.

17th PicoQuant International Workshop

Berlin, Germany **Sept 7-9, 2011**

"Surface-immobilization strategies for observing real-time dynamics of DNA polymerase": Poster presentation in which results of vesicle immobilization of 'DNA polymerase I' were presented.

Contents

Chapter 1: Introduction **10**

Chapter 2: RNA polymerase **13**
2.1 Introduction 13
2.2 RNAP core 13
2.3 RNAP core structure 14
2.4 σ^{70}-dependent Promoters 16
2.5 RNAP holoenzyme 17
2.6 Transcription initiation 19
2.7 RP_O 19
2.8 RNA synthesis *de novo* 22
2.9 Abortive initiation 25
2.9.1 Causes of abortive initiation 25
2.9.2 Possible functions of abortive initiation 28
2.10 Productive initiation 29
2.11 RNAP β' clamp 30
2.12 Single-molecule studies of bacterial RNAP 32

Chapter 3: Single-molecule fluorescence **35**
3.1 Introduction 35
3.2 Principles of fluorescence 35
3.3 Förster Resonance Energy Transfer 36
3.4 Single molecule detection techniques 38
3.4.1 Confocal spectroscopy and alternating laser excitation 39
3.4.2 Total Internal Reflection Fluorescence (TIRF) microscopy 39
3.5 Photophysics 41

Chapter 4: Immobilization of proteins for single-molecule studies **44**
4.1 Introduction 44
4.2 Materials and Methods 45
4.2.1 DNA and protein preparation 45
4.2.2 Vesicle encapsulation of KF 47
4.2.3 Vesicle immobilization and imaging 48
4.2.4 Porous vesicle formation 49
4.2.5 Antibody-immobilization of KF and imaging 49
4.2.6 Instrumentation 50
4.2.7 Data analysis and visualisation 51
4.3 Results and discussion 52
4.3.1 Immobilization of vesicle-encapsulated KF and comparison with solution data 52
4.3.2 KF fingers opening and closing 54
4.3.3 Permeabilising the vesicle lipid membrane 56
4.3.4 Antibody-immobilization of KF 58
4.4 Conclusions 63
4.5 Contributions 64

Chapter 5: RNAP transcription bubble dynamics **67**
5.1 Introduction 67
5.2 Materials and Methods 68
5.2.1 DNA and protein preparation 68

5.2.2 RP_O formation 70
5.2.3 Vesicle encapsulation and immobilization of RP_O 70
5.2.4 Antibody-immobilization of RP_O 71
5.2.5 Instrumentation 71
5.2.6 Single-molecule experiments 71
5.2.7 Data analysis and visualisation 71
5.2.8 HMM Analysis of dynamic smFRET time trajectories 72
5.3 Results 73
5.3.1 Vesicle immobilization of RNAP-promoter DNA complexes 73
5.3.2 Antibody-immobilization of RNAP-DNA complexes 77
5.3.3 Investigating the origin of transcription bubble dynamics: DNA determinants 80
5.3.4 Investigating the origin of transcription bubble dynamics: Protein determinants 83
5.3.5 Kinetic analysis of transcription bubble dynamics 86
5.4 Discussion 87
5.4.1 The transcription bubble exhibits static heterogeneity 87
5.4.2 The transcription bubble exhibits dynamic heterogeneity 91
5.4.3 Possible functions of the static and dynamic heterogeneity 92
5.5 Conclusions 93
5.6 Contributions 93

Chapter 6: Initial transcription monitored in real-time 95
6.1 Introduction 95
6.2 Materials and Methods 97
6.2.1 DNA and protein preparation 97
6.2.2 RP_O formation 97
6.2.3 *In vitro* transcription (IVT) reactions 97
6.2.4 Antibody-immobilization of RP_O 98
6.2.5 Instrumentation 98
6.2.6 Single-molecule experiments 98
6.2.7 Data analysis and visualisation 99
6.2.8 HMM Analysis of cycling smFRET time trajectories 100
6.2.9 Accessible volume (AV) modelling 101
6.3 Results 102
6.3.1 Initial attempts to monitor RNA synthesis produces unexpected results 102
6.3.2 An alternative labelling scheme 106
6.3.3 RNA synthesis observed in real-time 108
6.3.4 Calibrating our FRET signal 112
6.3.5 Structural interpretations of complex $RP_{ITC\leq7}$ behaviour 114
6.3.6 Kinetics of abortive cycling 115
6.3.7 σR3.2 in abortive initiation 117
6.3.8 Promoter escape observed in real-time 119
6.3.9 Abortive cycling in the presence of all four NTPs 122
6.4 Discussion 124
6.4.1 Designing a DNA substrate to monitor initial transcription in real-time 124
6.4.2 RNA synthesis observed in real-time 126
6.4.3 Extracting abortive cycling kinetics 127
6.4.4 RNAP pauses and backtracks in initial transcription 128
6.4.5 σR3.2 as a barrier to RNA extension 128
6.4.6 Pausing and backtracking: an additional explanation 131
6.4.7 Direct observations of promoter escape 131
6.4.8 Abortive cycling does not appear to be a pre-requisite for promoter escape 132
6.4.9 RNAP-promoter DNA complexes can be trapped in abortive initiation 133

6.5 Conclusion 134
6.6 Contributions 134

Chapter 7: RNAP clamp dynamics 136
7.1 Introduction 136
7.2 Materials and Methods 138
7.2.1 DNA and protein preparation 138
7.2.2 RP_O formation 138
7.2.3 Antibody-immobilization of RNAP 139
7.2.4 Instrumentation 139
7.2.5 Single-molecule experiments 139
7.2.6 Data analysis and visualisation 140
7.2.7 HMM Analysis of dynamic smFRET time trajectories 140
7.2.8 Accessible volume (AV) modelling 141
7.3 Results 141
7.3.1 Clamp conformations in RNAP holoenzyme 141
7.3.2 Clamp static and dynamic heterogeneity in RNAP holoenzyme 145
7.3.3 Clamp conformations in RP_O 146
7.3.4 Clamp conformation during initial RNA synthesis 148
7.3.5 Effect of switch-region-target inhibitors on the RNAP clamp conformation 149
7.4 Discussion 153
7.4.1 RNAP clamp adopts a primarily open conformation in holoenzyme 153
7.4.2 RNAP clamp is closed in RP_O and remains closed during RNA synthesis 155
7.4.3 Switch-region-target inhibitors differentially modulate the clamp conformation 155
7.4.4 Integrating work from different chapters to understand the RP_O formation mechanism 156
7.5 Conclusion 157
7.6 Contributions 158

Chapter 8: Conclusion 160
8.1 Future work 163

Appendix A: *E. coli* RNAP-promoter DNA complex structural model 167
A.1 RP_O model 167
A.2 RP_{ITC} models 168
A.3 RNAP His-tag location 169

Appendix B: Additional real-time RNA synthesis results 170
B.1 Leading edge-FRET (LE-FRET) shows similar heterogeneity in abortive initiation 170
B.2 Actively transcribing complexes occasionally exhibit switching behaviour 172

Appendix C: Additional dwell time distributions 173
C.1 RP_O bubble dynamics 173
C.2 Abortive cycling dynamics 175

Bibliography 176

Chapter 1: Introduction

Single-molecule techniques have a number of advantages over classical ensemble approaches. These include: the ability to observe transient intermediates in a reaction pathway, to disentangle static heterogeneity from dynamic heterogeneity, and to observe real-time trajectories of biological reactions without having to synchronise the reactions. It is therefore not surprising that these techniques have in recent years been utilized to answer long-standing questions in a number of areas of biological research. One particular field of research that has benefited from the emergence of single-molecule tools is that of transcription.

Single-molecule optical detection and nanomanipulation studies have provided new insight into all three stages of transcription: initiation, elongation, and termination (1). Below I mention just some examples of how single-molecule techniques have advanced our understanding of transcription.

The high spatial resolution afforded by optical tweezers has permitted researchers to observe transcription elongation at base-pair resolution (2), and also to investigate the mechanism and regulation of transcriptional pausing and backtracking (3). Single-molecule Förster Resonance Energy Transfer (smFRET) experiments combined with a trilateration/triangulation method have provided new insight into the structure of the eukaryotic elongation complex. Two notable contributions include mapping the paths of 1) the exiting nascent RNA, and 2) the non-template DNA strand, which were not resolved by X-ray crystallography (4, 5). Studies using smFRET and magnetic tweezers have revealed that initial transcription proceeds via a scrunching mechanism (6, 7), and fluorescence colocalization experiments have been used to extract the kinetics of initiation at an activator-dependent promoter (8).

Despite the major contributions made by single-molecule techniques to the field of transcription, many questions still remain unanswered and therefore open to investigation. In this thesis, we use smFRET to tackle some of the remaining questions concerning the molecular mechanism of transcription initiation.

This thesis is organised as follows: Chapter 2 provides an introduction to RNA polymerase (RNAP), and highlights the relevant studies that have shaped our current understanding of the mechanism of transcription initiation. Chapter 3 provides a brief introduction into the principles of fluorescence, FRET, and total internal reflection fluorescence (TIRF) microscopy, which are the concepts and techniques that are implemented in our investigations. Chapter 4 describes how we developed protocols to immobilize proteins for single-molecule measurements using TIRF microscopy, and demonstrates its use in revealing inter-molecular heterogeneity in DNA polymerase I. Chapter 5 presents the use of the immobilization strategies developed in chapter 4, to monitor transcription bubble conformational dynamics in the RNAP-promoter DNA initiation complex. This chapter also describes our investigations into the function of the observed transcription bubble dynamics. Chapter 6 extends our smFRET assay to monitor the initial transcription reaction in real-time with an unprecedented level of detail. Chapter 7 describes our ability to monitor the RNAP β' clamp conformation at different stages of transcription, and how two specific antibiotics modulate the clamp conformation. Finally, Chapter 8 provides concluding remarks and suggestions for future investigations.

Chapter 2: RNA polymerase

2.1 Introduction

Transcription is the process by which the information encoded in DNA is transferred to RNA. It is the first and most regulated step in gene expression (9) and is carried out by the enzyme RNA polymerase (RNAP) (10). Since its discovery over 40 years ago (11–13), the *Escherichia coli* RNAP enzyme has been the focus of intense investigations that have led to a vast accumulation of structural, biochemical, and genetic information regarding its function.

2.2 RNAP core

The RNAP 'core' enzyme is a multi-subunit assembly consisting of five core subunits (α^{I}, α^{II}, β, β', and ω) with a molecular mass of ~380 kDa (Fig. 2.1A). These subunits are evolutionarily conserved throughout the three domains of life and confer onto RNAP its catalytic activity (14).

The α subunits are identical in both sequence and structure, and form a dimer that acts as the assembly platform for RNAP. Each α subunit consists of two independently folded domains connected by a flexible linker (15–18). The N-terminal domain (αNTD) is responsible for dimerization and forms the interface with the rest of the polymerase. The C-terminal domain (αCTD) is primarily responsible for interactions with both *cis*- and *trans*-acting regulatory elements. The flexible linker joining the two domains enables the αCTD to interact with factors removed from the main RNAP body. The two largest RNAP subunits are β and β', which form the majority of the polymerase-DNA interface, the active centre, and also a number of the functionally important RNAP mobile elements (19). The ω subunit is the smallest of the five subunits and forms interactions primarily with the β' subunit. It has been shown to promote RNAP core assembly, but is not essential for enzymatic activity *in vitro* (20, 21).

2.3 RNAP core structure

Over the past 15 years a number of crystal structures of bacterial RNAP have been solved (17, 18, 22–26). These structures have enabled the reconciliation of much of the functional data available for bacterial RNAP, and have proved invaluable in guiding new experimental studies into RNAP function (6, 27, 28). The structures have shown that RNAP resembles a crab claw with two pincers (Fig. 2.1A–B). One of the pincers is formed by the β subunit, while the other is formed by the β' subunit. The two pincers form a DNA binding cleft that harbours the enzyme active centre, where RNA is synthesised via multiple rounds of nucleotide addition cycles (NAC).

The active centre contains two noteworthy structural elements that play key roles in NAC: the bridge helix and trigger loop (Fig. 2.1C). The bridge helix interconnects the β and β' subunits, and separates the enzyme active centre from the downstream DNA. The trigger loop forms a flexible loop and makes close contacts with incoming nucleoside triphosphates (NTPs). Both structural elements have been crystallised in multiple conformations. Conformational changes in the bridge helix/trigger loop module are thought to play a major role in catalysis and substrate selectivity (25, 28–30).

Each NAC involves phosphodiester bond formation via a two-metal-ion reaction mechanism in which three evolutionarily conserved aspartate residues at the active centre coordinate the tightly bound Mg^{2+}–I ion, while a second Mg^{2+}–II ion is recruited to the active site with incoming NTPs (31, 32). In this mechanism, Mg^{2+}–I activates the oxygen of the 3'-OH group so that it can perform a nucleophilic attack on the α-phosphate of the incoming nucleotide. Both metal ions then stabilize the resulting transition state and accelerate phosphodiester bond formation (Fig. 2.1D). The cycle is completed by the release of pyrophosphate (PP_i). In addition, the RNAP active site possesses phosphodiester bond hydrolysis activity, which is important for the rescue of arrested transcription complexes, and in the repair of misincorporated NTPs (33, 34).

RNAP contains three cavities connecting the active site of the enzyme to the external aqueous environment. They are the main DNA binding cleft, the secondary/NTP-entry channel, and the RNA-exit channel (Fig. 2.1A–B). The secondary channel is generally accepted as the route by which NTPs reach the active centre, and is

a target for inhibitor molecules (35). The channel also serves as a regulatory site for proteins such as Gre factors, which exert their regulatory effects via coiled-coil motifs that insert into the active centre via the secondary channel and stimulate the cleavage of backtracked RNA transcripts (36, 37).

Single-molecule studies have confirmed proposals, based on crystal structures, that the RNA-exit channel is the route by which nascent RNAs exit the polymerase (Fig. 2.1B) (4, 25). Transcripts up to ~9nt in length can be accommodated in the active centre before entering the exit channel. Then, transcripts must reach ~12nt in length before they are able to exit the polymerase.

RNAP contains functionally important mobile structural elements such as the β' clamp and β' jaw (Fig. 2.1A–B). The β' clamp plays key roles at all three stages of transcription. In initiation, it modulates DNA accessibility to the binding cleft, and once DNA is bound, it closes to prevent DNA release from the cleft, thus stabilising the RNAP-DNA complex (38). In elongation, the clamp is an important target for regulatory molecules such as NusG, which locks the clamp in a closed conformation, and enhances the processivity of RNAP (39). In fact, clamp opening during elongation has been proposed to be the key conformational change that results in RNAP pausing. In termination, clamp opening is thought to be a pre-requisite conformational change for both intrinsic and rho-dependent termination (23). The RNAP switch region is located at the base of the β' clamp and functions as the hinge that controls clamp opening and closing.

The β' jaw has also been implicated in stabilising the RNAP-DNA complex in transcription initiation. It is thought that the jaw domain undergoes a conformational change in later stages of initiation that enable it to form favourable interactions with downstream regions of the promoter DNA, which stabilise the complex (40, 41).

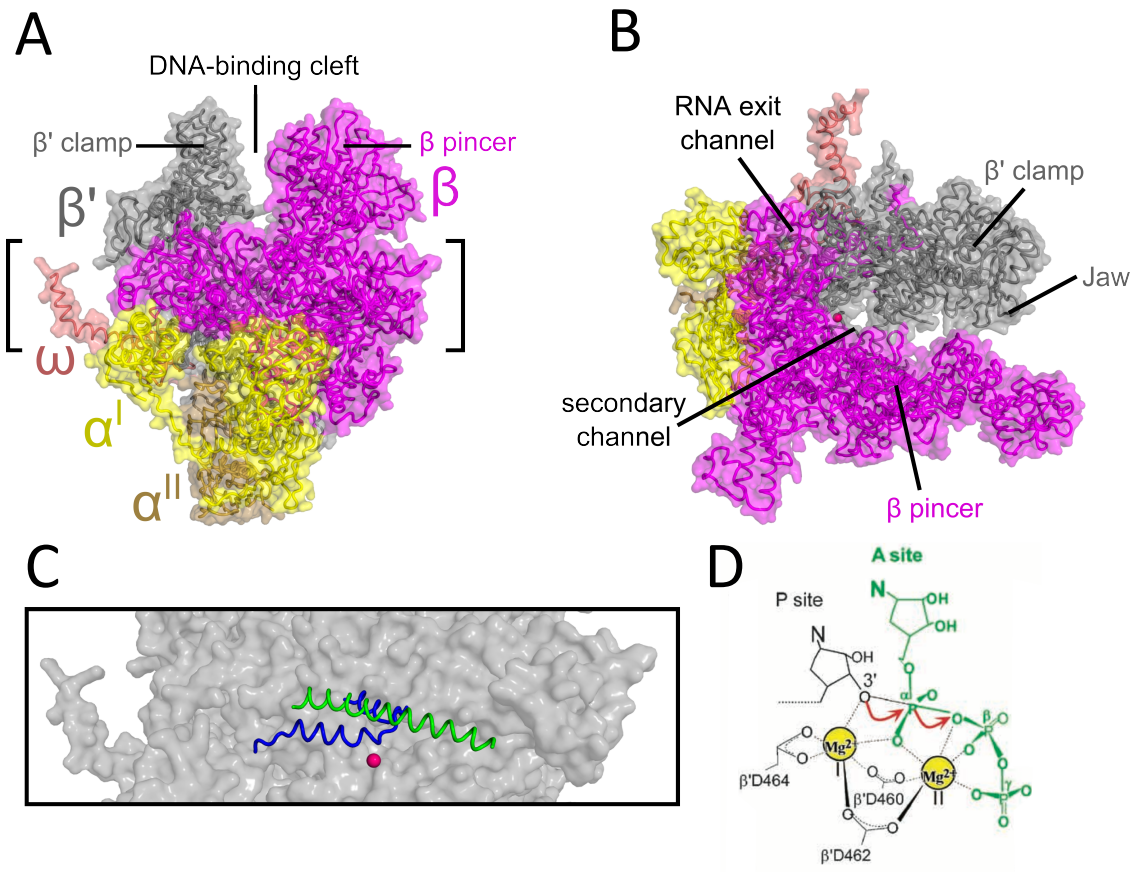


Figure 2.1. RNAP core. The colour scheme is the same for A and B. α^{I} is yellow, α^{II} is dark yellow, β is magenta, β' is grey, and ω is light red. (A) Front view of RNAP core. (B) Side view of RNAP core. (C) Zoomed in view of active site region marked with square parentheses in A. The protein surface is shown in grey. The bridge helix (green) and trigger loop (blue) are shown as cartoons. The active site Mg^{2+} ion is shown as a pink sphere. (D) Two-metal-ion reaction mechanism, figure from (42).

2.4 σ^{70}-dependent Promoters

Promoters are specific DNA sequences that recruit RNAP to the beginning of genes in order to initiate transcription. Bacterial promoters at the front of genes required for exponential growth are recognised by a protein factor called σ^{70}. These promoters are said to be σ^{70}-dependent. Other σ factors such as σ^{54} recruit RNAP to different promoters during different growth conditions and/or periods of stress (43). σ^{70}-dependent promoters consist of two core promoter elements known as the -10 and -35 elements (Fig. 2.2) (44–46). These two hexameric sequences are named according to their location relative to the transcription start site position (+1). The -10 element has

the consensus sequence 5'-TATAAT-3', while the -35 element has the consensus sequence 5'-TTGACA-3' (47). The sequence elements are separated by a spacer region that has a consensus length of ~17 bp but no conserved sequence (44). Structural studies have shown that the length of the spacer is critical in positioning the -35 and -10 promoter elements correctly for optimal binding to RNAP (48, 49).

Sequence variations in the core promoter elements influence the basal strength of promoters (47). In general, the closer the core promoter elements resemble the consensus sequences, the stronger the binding of the polymerase to the promoter. A number of promoters also contain additional sequence elements that can increase the binding affinity and compensate for -10 and -35 elements that deviate from the consensus (46). One such sequence element is the 'extended -10 element', which is found at some promoters immediately upstream of the -10 element and has the consensus sequence 5'-TGN-3' (47). The 'UP element' is another such sequence region that is ~20 bp long and is located upstream of the -35 element. It is recognised by the C-terminal domains of the RNAP α^{I} and α^{II} subunits, and can increase the strength of a promoter up to 50-fold (50, 51) (Fig. 2.2).

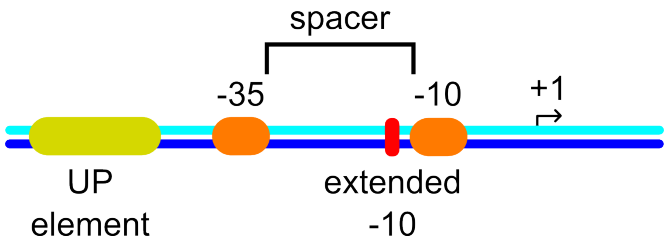


Figure 2.2 Simple schematic of a bacterial promoter. The template strand is shown in dark blue and the non-template strand is shown in light blue. See main text for details on the sequence elements.

2.5 RNAP holoenzyme

RNAP core enzyme possesses non-specific catalytic activity but is not able to initiate transcription at promoters in a sequence-dependent manner (52). To bind to a promoter and initiate transcription, RNAP core must first associate with a sigma (σ) factor subunit to form a six-subunit assembly referred to as RNAP holoenzyme ($\alpha^{I}\alpha^{II}\beta\beta'\omega$ σ) (53–55). The most abundant σ factor in *E. coli* is σ^{70}, which directs the catalytic

activity of the RNAP core enzyme to promoter sequences at the housekeeping genes; however, most bacteria contain multiple σ factors (49, 56, 57).

σ^{70} is a 70 kDa protein consisting of five structural regions: σR1.1, σR2, σR3, σR3/σR4 linker (hereafter referred to as σR3.2), and σR4 (Fig. 2.3A–B). These regions are highly conserved among σ^{70} family σ factors, and each plays a specific role in initiation (19, 58). σR2 and σR4 are structured domains that make specific interactions with the promoter -10 and -35 elements, respectively (49, 59). σR2 is positioned between the β and β' pincers, while σR4 is located close to the RNA-exit channel. σR1.1 and σR3.2 are unstructured and negatively charged domains that play critical roles in post-promoter-recognition steps such as transcription bubble melting, abortive initiation, and promoter escape (49, 60). σR1.1 binds in the RNAP DNA binding cleft, while σR3.2 binds in the RNA exit channel. It is thought that these two domains act as molecular placeholders for DNA by binding in the active centre cleft where they prevent RNAP-DNA non-specific interactions from forming (17). However, their positioning also means that before DNA can enter the binding cleft, σR1.1 must be ejected, and similarly, for RNA transcripts to progress into the RNA-exit channel, σR3.2 must be displaced (61).

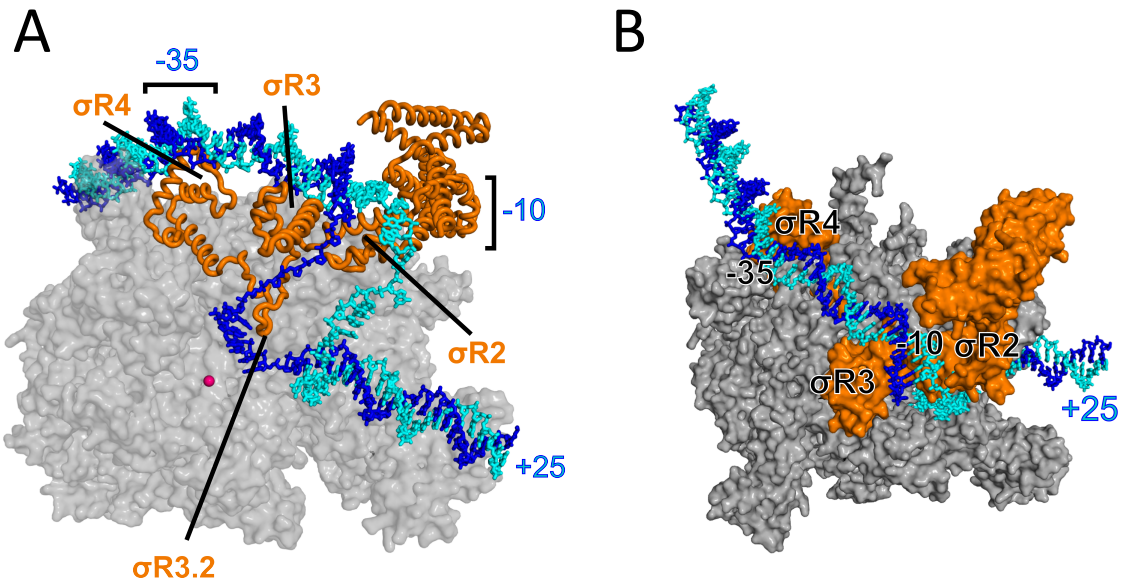


Figure 2.3. Structural model of *E. coli* RNAP holoenzyme bound to promoter DNA. RNAP core is shown in grey, σ^{70} is shown in orange, the promoter DNA template strand is shown in dark blue, the non-template strand is shown in light blue. The active site Mg^{2+} ion is shown as a pink sphere. The promoter -10 and -35 elements are labelled. σR1.1 is omitted for clarity. (A) Side view with RNAP core shown partially transparent. (B) Top view. See *Appendix A* for details on how the structural model was created.

2.6 Transcription initiation

Transcription initiation has been the focus of intense investigations for a number of years. Studies involving structural, chemical footprinting, and rapid mixing kinetic approaches have shown that initiation is a multi-step process, which begins with RNAP holoenzyme binding to a promoter, and ends when the initiation complex becomes an elongation complex (Fig. 2.4) (17, 41, 62–64). Each step along the pathway involves structural re-arrangements or 'isomerizations'. The first few isomerization steps culminate in the formation of the RNAP-promoter DNA open complex (RP_O) (65). This is followed by a second set of isomerization steps in which the polymerase synthesises RNA *de novo* until it is able to transition to elongation.

The transcription initiation mechanism is shown in Fig. 2.4. This view of initiation was largely developed using biochemical investigations. Notably, mutational studies identified aromatic residues that play a key role in bubble melting, temperature shift experiments combined with footprinting techniques permitted the characterisation of RP_C and RP_i structures, and *in vitro* transcription assays revealed that abortive cycling occurs at almost all promoters (66).

2.7 RP_O

The first step in transcription initiation involves the binding of RNAP holoenzyme to promoter DNA, to yield the RNAP-promoter closed complex (RP_C) (Fig. 2.4). RP_C is an unstable heparin-sensitive complex that either rapidly dissociates, or isomerizes to the next intermediate in the initiation pathway. The transcription bubble region remains double-stranded (dsDNA) in RP_C. In the next isomerization step at the *lac* promoter, RNAP-DNA interactions are extended downstream of the transcription start site, which results in the formation of a heparin-resistant RNAP-promoter intermediate complex (RP_i) where promoter DNA remains double-stranded[1]. The final isomerization step sees the unwinding of 12-14 base pairs (bp) of DNA surrounding the transcription start site

[1] The structural characteristics of the intermediate vary depending on the promoter in question.

(positions -12 to +2), to form an open transcription bubble region. The unwound template (T) strand is inserted into the active centre cleft of the polymerase, while the nontemplate (NT) strand is placed in its binding track to yield the stable and catalytically competent RP_O (43, 54, 66). There is currently no full RP_O structure available; however, a detailed structural model of a thermophilic RP_O has been pieced together using a combination of available crystal structures (49) and a wealth of genetic and biochemical data; these include cross-linking studies, which reveal the close interactions between RNAP and DNA (67), DNase and hydroxyl radical footprinting studies of intermediates trapped at various temperatures, which reveal large scale DNA conformational changes, and MnO_4^- footprinting studies, which reveal the presence of single-stranded DNA regions (66). In addition, we have created a structural model of *E. coli* RP_O that is used for structural considerations throughout this thesis (Fig. 2.3). Details of how this model was created can be found in *Appendix A*.

The RP_O structural model combined with functional studies have considerably advanced our understanding of the DNA melting process. However, the exact mechanism by which melting occurs still remains a matter of debate (66, 68). In the current model, DNA melting proceeds via a sequential 'nucleation-propagation' mechanism in which DNA opening is nucleated at position -11 by σ^{70}, and is then propagated downstream to the +2 position (Fig. 2.4B). σR2.3 has several highly conserved aromatic residues that assist in initial nucleation by binding in a sequence-specific manner to NT strand bases at the -10 element (54, 69, 70). One of these residues is the absolutely conserved W433 (*E. coli* numbering), which acts as a wedge to disrupt the A/T base pair at position -11. Recent crystal structures showing the interactions between σ^{70} and the -10 element demonstrate that the -11 A base and the -7 T base are buried in pockets on the σ^{70} surface. This observation explains why these two bases are the most highly conserved positions in the -10 element (71–73). However, it remains unclear whether RNAP actively destabilises base pairing at the -10 element or opportunistically traps the single-stranded conformation of a transiently denatured -10 element.

The steps that follow initial nucleation are more contentious. There are currently two models for bubble propagation known as the 'bend-load-open' and 'open-bend-load' models (Fig. 2.4C) (66, 68). In the former model, full DNA unwinding is proposed to occur after the duplex DNA has entered the DNA binding cleft. In the latter model, DNA unwinding must occur before single-stranded DNA (ssDNA) is able to enter the binding cleft (74).

In the bend-load-open model, DNA nucleation facilitates the bending of the DNA double helix at the -10 element. This bend brings the downstream DNA to the vicinity of the jaw so that it can be loaded into the DNA binding cleft where interactions between the downstream DNA and the β/β' subunits can promote active DNA melting *inside* of the binding cleft. This model is supported by experiments that stabilise putative bend-load-open intermediates by varying conditions such as temperature (63, 75–78). However, it should be noted that it remains to be shown whether the intermediates identified truly represent *bona fide* on-pathway intermediates.

The open-bend-load model posits that after initial nucleation, the T strand enters the DNA binding cleft and diffuses to the active site as ssDNA. Thus, full DNA melting precedes DNA entry into the binding cleft. This model was originally proposed based on crystal structures of holoenzyme and an RNAP-fork junction DNA complex, which showed the DNA binding cleft in a conformation too narrow to accommodate dsDNA (49, 62). Further support for this model has come from real-time kinetic measurement at the T7 *A1* promoter as well as computational simulations of DNA melting (74, 79). However, it should be noted that a recent single-molecule study of RNAP holoenzyme has shown that the β' clamp adopts primarily an open conformation in solution (27). This raises the possibility that crystal-packing forces may have artificially stabilised the closed conformation in the holoenzyme crystal structure mentioned above.

These conflicting observations regarding the bubble melting process, combined with the fact that RP_O exhibits varying stabilities, structural intermediates, and levels of activity depending on the promoter sequence it binds to, suggest that the RP_O formation mechanism may not necessarily follow a single linear pathway at all promoters. In fact, it has been suggested that the exact sequence of events may vary depending on the

promoter sequence context and on environmental conditions such as temperature and DNA supercoiling (66, 68).

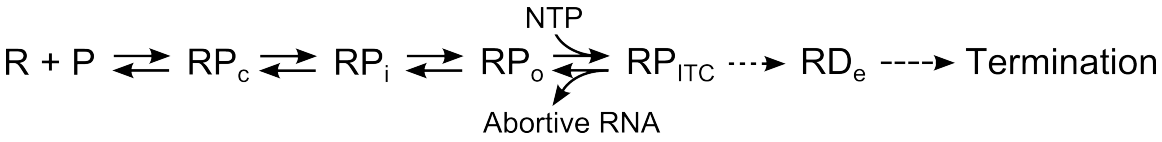
Finally, it is worth mentioning that a recent structural study of a *Thermus thermophilus* transcription initiation complex demonstrated that σR3.2 forms a loop that protrudes towards the RNAP active centre and comes in close contact with the T strand (72). A current hypothesis is that interactions between σR3.2 and the melted T strand are involved in the strand's correct positioning at the active site, and in the stability of the transcription bubble (73). This hypothesis is supported by $KMnO_4$ footprinting studies that show a change in the accessibility of the transcription bubble thymines when σR3.2 is deleted (80).

2.8 RNA synthesis *de novo*

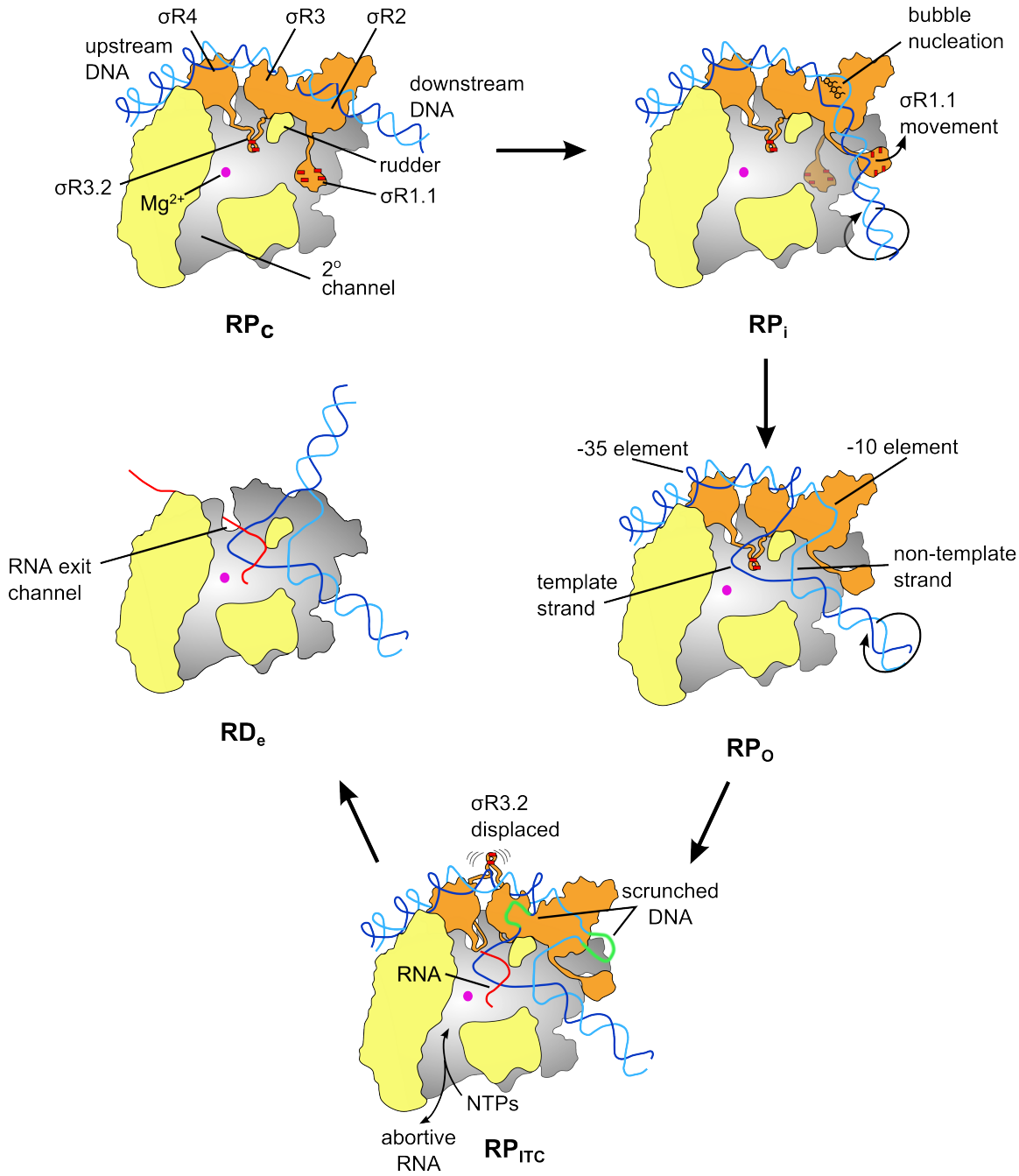
In the presence of NTPs, RP_O begins synthesising RNA *de novo* to yield the RNAP-promoter DNA initial transcribing complex (RP_{ITC}) (Fig. 2.4). This initial phase of RNA synthesis proceeds via a mechanism distinct from that in elongation known as 'scrunching'. Here, RNAP remains stably bound to the promoter and pulls downstream DNA towards itself while synthesising RNA. Since the polymerase does not move forward during this process, DNA from the expanding transcription bubble is accumulated as ssDNA bulges i.e. DNA is scrunched into the polymerase (6, 7).

Initial RNA synthesis can be classified as either abortive or productive. In abortive initiation, RNAP fails to escape the promoter and progress to elongation. Instead, the polymerase iteratively synthesises and releases short RNA products of varying lengths (81, 82). Abortive RNAs are generally 2–9-nt long; however, some promoters produce longer abortive products known as 'very long abortive transcripts' (VLATs) (83). It is hypothesised that strain generated from DNA scrunching destabilises short RNAs at the active centre and drives their ejection from the polymerase. A current model proposes that abortive RNAs are displaced from the active centre in a backward direction such that the 3'-single stranded RNA tail becomes extruded into the secondary channel, which temporarily arrests the RNAP in a backtracked position. This view of RNA

A


R + P
RP$_c$
RP$_i$
RP$_o$
NTP
RP$_{ITC}$
RD$_e$
Termination
Abortive RNA

B


σR4
σR3
σR2
upstream DNA
downstream DNA
σR3.2
Mg^{2+}
rudder
σR1.1
2° channel
RP$_C$
bubble nucleation
σR1.1 movement
RP$_i$
-35 element
-10 element
non-template strand
template strand
RP$_O$
σR3.2 displaced
scrunched DNA
RNA
NTPs
abortive RNA
RP$_{ITC}$
RNA exit channel
RD$_e$

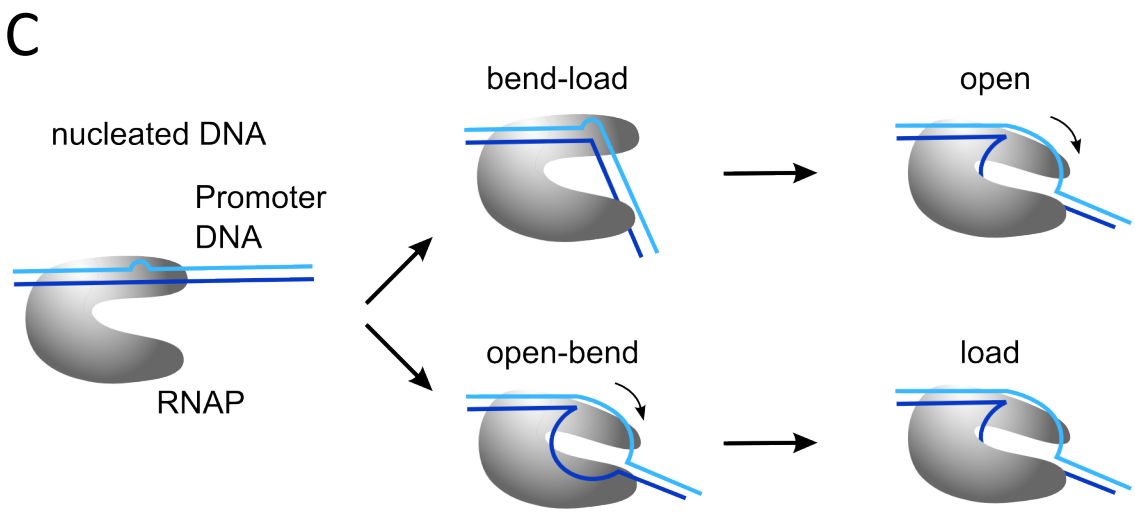


Figure 2.4. Schematic of transcription initiation. (A) Simple schematic representation of transcription showing free RNAP (R) binding to promoter DNA (P) to form the RNAP-DNA closed complex (RP_C). RPc then isomerizes to form the RNAP-DNA intermediate complex (RP_i) where DNA melting has been nucleated. RP_i then isomerizes to form the RNAP-DNA open complex (RP_O), which subsequently undergoes multiple rounds of abortive initiation as an RNAP-DNA initial transcribing complex (RP_{ITC}). The transition to elongation involves the formation of an elongation complex (RD_e). (B) Structural changes that occur at each step between RP_C and RD_e in Fig. 2.4A. Curly arrows in RP_i and RP_O indicate DNA rotation due to transcription bubble melting. Figure adapted from (54). Evolutionarily conserved negatively charged residues are indicated as red negative signs in RP_C and RP_i. Aromatic residues involved in bubble nucleation are shown in RP_i. (C) Key structural changes in the bend-load-open (top) and open-bend-load (bottom) models of promoter opening. The curly arrow indicates closing of the RNAP clamp.

release is supported by the fact that Gre factors – which stimulate the cleavage of backtracked RNAs – are able to reduce the level of abortive products while increasing that of the full-length RNA (84–87). Since Gre factors appear to only rescue abortive transcripts longer than 4–5-mers, it is postulated that shorter backtracked abortive transcripts are so unstably bound at the active centre that they are quickly ejected from the polymerase (88). Upon RNA release, RP_{ITC} reverts back to RP_O to begin another round of RNA synthesis.

In productive initiation, once the nascent RNA becomes ~9–11-nt long, the polymerase is able to leave the promoter and transition to elongation via a process known as promoter escape. Once the polymerase has escaped the promoter, it no longer proceeds via a scrunching mechanism, and instead translocates forward along the DNA with each NTP incorporated into the nascent RNA chain. Once the nascent RNA

exceeds ~12-nt in length, it is extruded out of the RNA-exit channel to form a highly stable complex known as the RNAP-DNA elongation complex (RD_e). During this process, the likelihood of σ^{70} release increases (Fig. 2.4B) (66).

2.9 Abortive initiation

Abortive initiation has been conserved through evolution and is carried out by RNAPs from all three domains of life (89, 90). Furthermore, abortive initiation has been shown to occur *in vivo* (91). Despite these indicators of biological importance, the exact cause and function of abortive initiation remains unclear. Let us first review what is known regarding the causes of abortive initiation, and then consider possible biological functions.

2.9.1 Causes of abortive initiation

The simplest model to account for abortive initiation is that for each potential NTP addition event there is a direct competition between an NTP being incorporated into the nascent RNA and the release of that RNA (92). This model predicts that the amount of abortive products produced would decrease with an increasing NTP concentration ([NTP]), and that the abundance of a specific abortive product could be reduced if the concentration of the next complementary NTP were to be increased. However, experiments testing these predictions have shown that abortive products are still produced at saturating [NTP], and RNAs of specific sizes are abortively released with characteristic probabilities that vary depending on the promoter sequence (92). This indicates that the model above does not provide an adequate explanation for abortive initiation, and that the promoter DNA sequence is an important factor that influences whether RNAP will perform abortive initiation.

As a general rule, the closer a promoter resembles the consensus sequence, the more stable the resulting RP_O is, and the more likely it is to undergo abortive cycling (93–97). This is because RNAP forms strong interactions with consensus promoter

sequences that are difficult to break during the transition to elongation, and so the polymerase becomes stalled at the promoter. This relationship helps explain why the *rrnB* P1 promoter, which has non-consensus features, forms an unstable RP_O and yet is transcriptionally very active and produces almost no abortive RNA products (84, 98, 99). It appears that evolution has selected for non-consensus sequence features in this promoter to enable the polymerase to bind weakly to it and escape with relative ease. These observations suggest that the strength of the binding interactions between the polymerase and the promoter is a cause of abortive initiation.

However, mutational studies have shown that it is not only the core promoter recognition region (PRR; -40 to -1) that influences whether RP_O undergoes abortive initiation. Two DNA sequences with the same PRR and different initial transcribed sequences (ITSs; +1 to +20) can produce distinct but reproducible ladders of abortive RNA products (known as abortive profiles) when visualised on a gel (Fig. 2.5A) (92). The band intensities of an abortive profile reflect the abundance of the aborted RNA products, and thus the inherent abortive potential of RNAs of different lengths. While one might expect that shorter RNAs would be unstably bound at the active site and therefore have a higher abortive potential compared to longer RNAs, abortive profiles do not show an absolute correlation between the abundance of an RNA product and its length (81, 84). Furthermore, there is no obvious correlation between the abundance of an abortive RNA product and its sequence i.e. GC-rich RNA-DNA heteroduplexes do not invariably result in fewer abortive products than AT-rich heteroduplexes (92). These observations indicate that abortive RNA release is dependent on the underlying DNA sequence in a manner more complex than that simply dictated by heteroduplex stability, or the strength of the RNAP-promoter DNA interactions.

Structural studies of RNAP have shown that σR3.2 is located within the RNA exit channel where it apparently forms a physical impediment to the advancing nascent RNA (Fig. 2.3–2.4) (16, 17). Furthermore, functional studies have shown that deleting σR3.2 can reduce abortive initiation (17). These observations indicate that another possible cause of abortive initiation is the presence of σR3.2 in the active centre. However, as discussed above, the abortive profiles of different promoters vary considerably despite

the fact that σR3.2 would be expected to influence promoters equally in a transcript-length-dependent manner. Therefore, although there is evidence to suggest that the steric hindrance created by σR3.2 may promote abortive initiation, other factors that we have already considered such as the PRR and the ITS also play important roles.

So far we have seen that a number of factors influence the likelihood that RNAP will undergo abortive cycling. This situation is further complicated by the fact that at a number of promoters, a sub-population of RP_O molecules exist that are trapped in abortive initiation and are unable to proceed to elongation (Fig. 2.5B) (100, 101). These molecules are referred to as moribund complexes, and their abundance varies between promoters (102). Although there have been some suggestions that moribund complexes may play some regulatory role in gene expression (see section 2.9.2), very little is known about the origin of this inter-molecular heterogeneity.

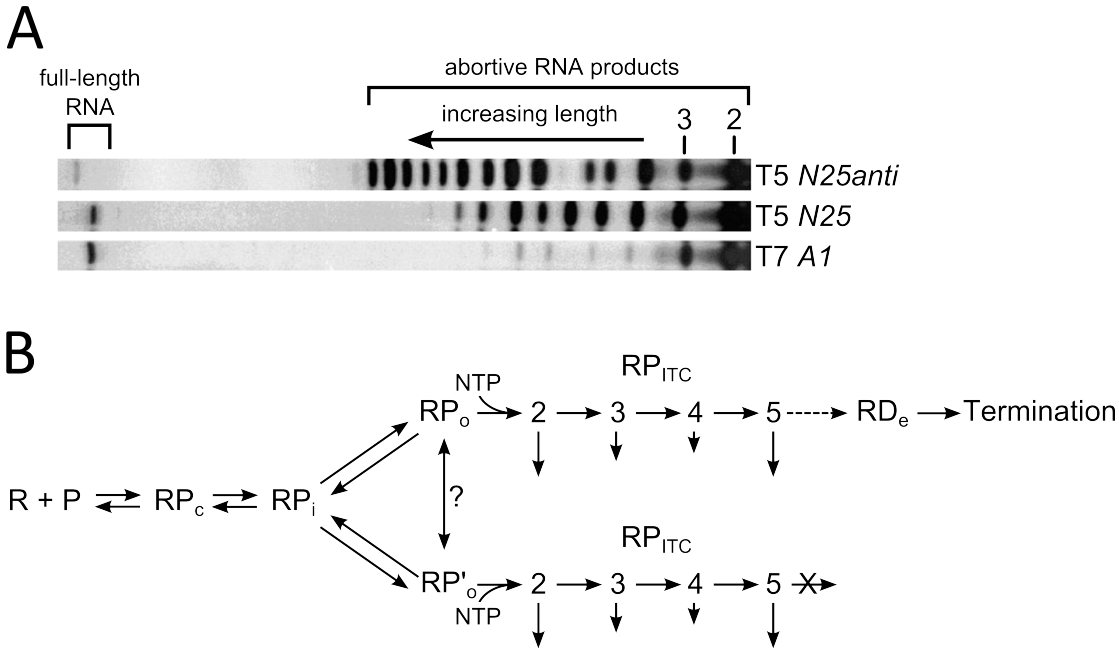


Figure 2.5. Abortive initiation. (A) Abortive profiles of the T7 *A1*, T5 *N25,* and T5 *N25anti* promoters showing variability in the abundance of abortive products. The T5 promoters have the same PRR but different ITSs, which highlights the importance of the ITS in influencing abortive initiation. 2-mer and 3-mer RNAs are labelled. (B) Branched pathway model of transcription initiation showing the formation of productive and moribund open complexes (RP_O and RP'_O, respectively). RP_O is able to escape the promoter after abortive initiation. RP'_O is unable to escape the promoter, and therefore performs multiple rounds of abortive cycles. It is unclear whether RP_O can isomerize into RP'_O and vice versa. Figures adapted from (84).

2.9.2 Possible functions of abortive initiation

It is possible that abortive initiation may simply be an inescapable consequence of the mechanism by which RNAP binds to a promoter and initiates transcription. However, it is difficult to understand why evolution would have conserved such a futile cycle in all branches of life, if it were not in some way useful. Furthermore, the *rrnB* P1 promoter provides an example of how evolution can select for promoter sequences that initiate transcription with almost no abortive initiation, so why then do all promoters not exhibit this behaviour? One possibility based on the fact that abortive RNA release and productive RNA extension are in competition with one another, is that abortive initiation constitutes a mechanism for gene regulation (91). In the simplest scenario, a regulatory signal would influence whether RP_O is able to progress past the abortive phase of initiation to elongation, and would therefore influence how much full length RNA is synthesised. This model is supported by the regulatory mechanisms observed at the *lac* and *malT* promoters (84, 103). At the *lac* promoter, the DNA-binding protein LacI represses the synthesis of full-length RNA by permitting the synthesis of only short abortive products (104, 105). At the *malT* promoter, cAMP receptor protein increases the production of full-length mRNA at the expense of abortive RNA products (106).

The observation that Gre factors are able to promote the formation of productive initiation complexes at the expense of moribund complexes at the λP_R promoter has led to proposals that moribund complexes may also play a regulatory role in gene expression (107). In this model, moribund complexes would act as a source of 'ready-to-go' transcription complexes that are able to respond quickly to regulatory signals. However, this hypothesis remains a topic of debate.

A final possibility discussed in the literature and supported by experimental evidence is that short abortive products (2–4-mers), once released, may rebind to RP_O and prime RNA synthesis *in vivo*. The exact function of such a priming activity is unclear; however, researchers have shown that short RNAs that rebind to RP_O can alter the selection of transcription start sites and drive changes in gene expression, which suggests that the priming activity may play a regulatory role in gene expression (108).

2.10 Productive initiation

Productive initiation involves RNAP escaping the promoter and proceeding to elongation. It is only once this transition has occurred that a full-length RNA transcript will be produced. Investigations into the tendency of a promoter to produce a full-length RNA transcript have revealed that promoters are generally either rate-limited at the RP_O formation stage or at the promoter escape stage. *In vitro* studies have shown that promoters that are rate-limited at escape are readily able to form RP_O but have difficulties escaping the promoter (109). Furthermore, genome-wide ChIP-chip studies have shown that RNAP molecules are particularly enriched at promoter regions, which suggests that promoter escape may also be rate-limiting *in vivo* (110, 111). These observations highlight the fact that understanding the mechanism of promoter escape is key for our understanding of how gene expression might be regulated inside cells.

DNA scrunching is an obligatory process in promoter escape and has been proposed to play a functional role in the escape mechanism (6, 7). In one model, mechanical strain is built up during DNA scrunching until the nascent RNA becomes ~9–11-nt long, at which point the accumulated energy is sufficient to drive the release of RNAP-promoter DNA contacts. As the contacts are released, RNAP moves forwards and the upstream edge of the transcription bubble is re-annealed (5). This reduces the size of the transcription bubble to ~8–9-bp, which is then maintained during elongation (7). σ^{70} is responsible for the primary contacts made between RNAP holoenzyme and the promoter. It is therefore believed that for the polymerase to escape the promoter, σ^{70}-DNA contacts must be disrupted. How might this occur? One possibility is that the re-annealing of the upstream edge of the transcription bubble may compete with DNA-σ^{70} interactions. This hypothesis is supported by the fact that interfering with the re-annealing process of the upstream transcription bubble affects the forward translocation of the T7 RNA polymerase (112).

The observation that σ^{70} is often released during promoter escape has also led to suggestions that RNAP core contacts with σ^{70} are weakened during promoter escape, and this may help disrupt RNAP-promoter DNA contacts (17, 113, 114). A structural

explanation for the weakening of contacts between RNAP core and σ^{70} has come from X-ray crystallography and footprinting studies showing that for nascent RNAs to be extended into the RNA exit channel, σR3.2 must be displaced from the active site (Fig. 2.3–2.4) (25, 115). These findings have led to a model in which the displacement of σR3.2 weakens the interactions between RNAP core and σ^{70}, initiates the disruption of RNAP-promoter DNA interactions, and thereby facilitates promoter escape (54). In this model σR3.2 plays a major role in promoter escape, which is supported by functional studies showing: 1) that mutant complexes lacking σR3.2 exhibit a decreased efficiency in promoter escape (60), and 2) that DNA scrunching is not a major driving force in abortive initiation and promoter escape (116). The central role played by σR3.2 in initial transcription is further highlighted by the fact that its function has been conserved through evolution. In eukaryotes, the TFIIB B-reader loop is an ortholog of σR3.2 that also physically blocks the extension of nascent RNAs and must be displaced for RNA polymerase II to progress to elongation (117). This has led to suggestions of a common promoter escape mechanism for polymerases from all three domains of life.

Although a detailed picture of the promoter escape mechanism has been formulated over the years, questions still remain regarding the relationship between the abortive and productive phases of initiation. It also remains unknown whether abortive initiation, like DNA scrunching, is an obligatory intermediate on the path to promoter escape. It is challenging to address these questions using conventional bulk techniques because both abortive cycling and promoter escape are transient events that exhibit temporal heterogeneity. Single-molecule techniques on the other hand are well suited to investigate the kinetics of promoter escape.

2.11 RNAP β’ clamp

The RNAP β’ clamp has been captured in a number of conformations that range from a fully open state to one that is fully closed (Fig. 2.6A) (23, 25, 49, 118–120). The functional significance of these conformations was originally uncertain due to the possibility that crystal packing forces might be responsible for the differences observed.

Nevertheless, groups speculated with regards to the possible roles that the different conformations might play in transcription. For example, it was proposed that the closed conformation might stabilise both RP_O and RD_e by preventing the release of DNA from the binding-cleft (121). Recently, a single-molecule fluorescence study produced evidence supporting this proposal by showing that the clamp does indeed undergo conformational changes in solution that appear to function to prevent DNA release from RNAP-DNA complexes (27).

Conformational changes of the clamp are mediated by a hinge region at the base of the clamp known as the switch region (Fig. 2.6A). A number of small molecule inhibitors have recently been identified that specifically target the RNAP switch region (80, 122, 123). Myxopyronin (Myx) and Lipiarmycin (Lpm) are examples of switch-region-target inhibitors that prevent the formation of the catalytically competent RP_O (Fig. 2.6B).

Myx is an antibiotic (MW = 431 Da) produced by *Myxococcus fulvus* Mf50 (124). Myx inhibits bacterial RNAPs but not eukaryotic RNA Pol II (125). Biochemical studies have shown that Myx interacts with switch region 2, and prevents the formation of a fully open transcription bubble (122, 123). In one proposal, the inhibitory effect on RP_O formation was due to Myx-induced clamp conformational changes. This was later confirmed to be correct using a single-molecule approach (27).

Lpm is also an antibiotic (MW = 1, 073 Da) that is produced by *Actinoplanes deccanensis* and has been shown to be effective against rifampicin-resistant forms of *Myobacterium tuberculosis* (126, 127). Lpm – like Myx – also binds to switch region 2, and inhibits the formation of RP_O by completely blocking the formation of an open transcription bubble at an earlier step in the RP_O formation pathway than that targeted by Myx. It has been proposed, based on DNase I protection studies, that Lpm maintains the clamp in an open conformation (80); however, this has yet to be shown experimentally.

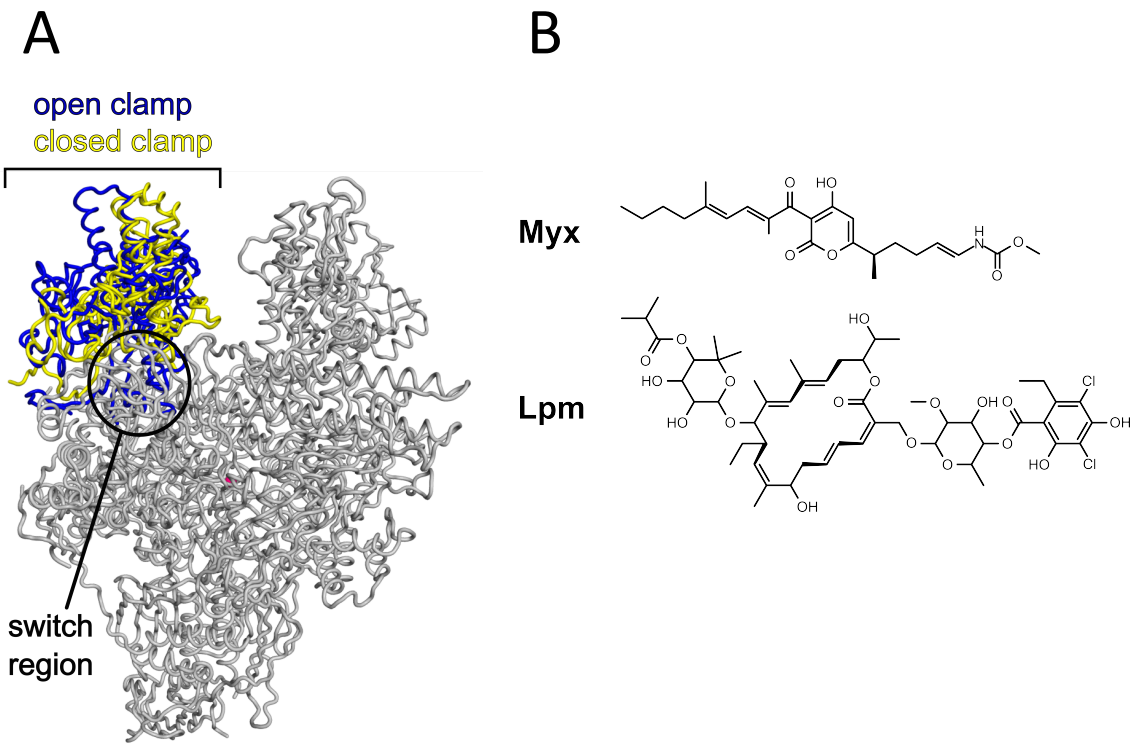


Figure 2.6. RNAP clamp structure and switch-region-target inhibitors. (A) The open and closed clamp conformations (PDB 1I3Q and 4IGC) are shown in blue and yellow, respectively. The remainder of the protein is shown in grey. (B) Chemical structures of Myxopyronin and Lipiarmycin. Figures adapted from (27, 80).

2.12 Single-molecule studies of bacterial RNAP

Single-molecule techniques have considerably advanced our understanding of the molecular mechanisms of multi-subunit RNA polymerases (1, 128, 129). Bacterial RNAP has been studied extensively in the elongation stage of transcription using nanomanipulation techniques (128, 130). These types of investigations have revealed the translocation mechanism of RNAP in elongation and the velocity at which RNAP moves forwards at this stage of transcription. Nanomanipulation techniques have also enabled investigations into the pausing and backtracking behaviour of RNAP, which are thought to play important regulatory roles in elongation (2, 131, 132). Although transcription initiation has been studied less extensively using nanomanipulation approaches, magnetic tweezer studies have revealed the scrunching mechanism of initial transcription (7). One possible reason for why nanomanipulation studies have been used less extensively in initiation compared with elongation may be the relative

instability of initiation complexes that makes it technically challenging to apply nanomanipulation techniques to the former. Instead, transcription initiation has been studied extensively using an alternative single-molecule approach that relies on fluorescence (1).

Single-molecule fluorescence techniques have made a number of notable contributions to our understanding of the mechanism of transcription initiation. For example, studies of diffusing molecules have shown that σ^{70} does not always dissociate from RNAP core during promoter escape (114), and also that initial transcription proceeds via a scrunching mechanism (6). More recently, single-molecule fluorescence investigations of immobilized molecules have permitted the elucidation of the kinetics of transcription initiation at an activator-dependent promoter (8). Furthermore, the recent development of mutagenesis using unnatural amino acids has permitted the targeted labelling of RNAP core with fluorophores, which in turn has made it possible to monitor RNAP protein conformational changes using single-molecule fluorescence tools (27).

Throughout this thesis we use single-molecule fluorescence approaches to investigate the detailed mechanism of bacterial transcription initiation and tackle long-standing questions regarding abortive initiation and promoter escape. In the next chapter we provide an introduction to the principles that underlie the single-molecule fluorescence techniques utilized in this work.

Chapter 3: Single-molecule fluorescence

3.1 Introduction

Over the past two decades, single-molecule techniques have emerged as powerful tools in the fields of biochemistry and biophysics, with an increasing number of researchers turning to single-molecule fluorescence approaches to answer biological questions that are either extremely challenging, or impossible to answer using classical ensemble techniques (133–135).

Single-molecule techniques have contributed to our understanding of a number of biological processes including protein-nucleic acid interactions (4, 6, 7, 136), the stoichiometry and oligomerization of important biological assemblies (137, 138), and the role of protein conformational dynamics in replication and transcription (27, 139, 140). More recent advances in the field have seen the application of single-molecule techniques to live cells (141–143).

In this chapter, I will introduce the concept of fluorescence and describe how biological processes are studied through the application of single-molecule Förster Resonance Energy Transfer (smFRET). I shall also describe the principal microscopy method used to monitor molecules of interest in this thesis.

3.2 Principles of fluorescence

Molecules can be described in terms of electronic states, which are associated with different amounts of energy. When a molecule absorbs photon energy, an electronic transition may occur such that the molecule reaches an electronically excited state. Electronic transitions tend to occur between the ground state and one of the higher vibrational levels of an excited state (Fig. 3.1A). A molecule can return to the ground state by several competing processes. One of these processes is vibrational relaxation, which will allow the molecule to reach the lowest vibrational level of the

excited state and involves the transfer of excess vibrational energy to nearby solvent molecules. A second pathway is photon emission, which is the process that gives rise to fluorescence. Here, the transition generally occurs between the lowest vibrational level of the excited state and a higher vibrational level of the ground state. Excess energy will again be dissipated via vibrational relaxation. Due to the dissipation of energy via pathways other than photon emission, the energy gaps for the electronic transitions from an excited state to a ground state are smaller than the energy gaps from a ground state to an excited state. This difference in energy gives rise to the Stokes Shift – the observation that the peak absorption of a molecule occurs at a shorter wavelength than that of the peak in emission (Fig. 3.1B).

3.3 Förster Resonance Energy Transfer

Förster resonance energy transfer (FRET) is the non-radiative transfer of energy from a donor fluorophore to an acceptor fluorophore via a dipole-dipole interaction that occurs when the emission spectrum of the donor fluorophore overlaps with the excitation spectrum of the acceptor fluorophore (Fig. 3.1B).

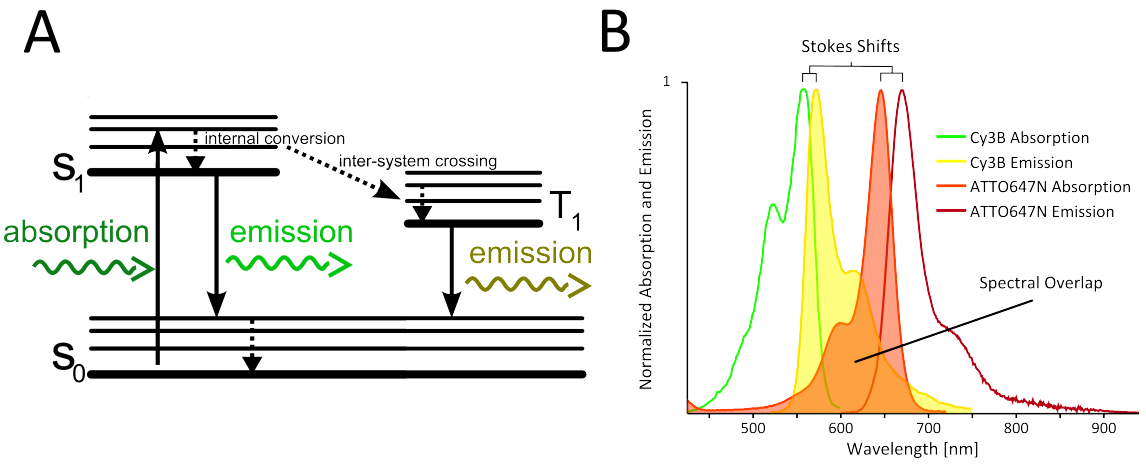


Figure 3.1. Principles of fluorescence. (A) Jablonski diagram showing the ground and excited electronic states of a molecule (S_0 and S_1) as well as the vibrational levels. Dashed arrows represent non-radiative transitions, while solid arrows represent radiative transitions. Inter-system crossing can occur to the triplet state (T_1), which leads to phosphorescence. (B) The normalised absorption and emission spectra of the fluorophores, Cy3B and ATTO647N. Figures adapted from Kapanidis Group communal figures.

The rate of energy transfer between the donor and acceptor dyes displays the following distance dependence:

$$k_{ET} = \frac{1}{\tau_D}\left(\frac{R_0}{R}\right)^6$$

where τ_D is the fluorescence lifetime of the donor in the absence of the acceptor, R is the distance between the fluorophores, and R_0 is the Förster Radius, which is the distance at which energy transfer efficiency is 50%, and is defined by:

$$R_0 = \frac{9000 \ln 10\, Q_D \kappa^2 J}{128\pi^5 n^4 N_A}$$

$$J = \int f_D(\lambda)\epsilon_A(\lambda)\lambda^4\, d\lambda$$

where κ is the orientation factor between the two fluorophore dipoles and averages to 2/3 for freely rotating dyes, n is the refractive index of the medium, N_A is Avogadro's constant, Q_D is the donor quantum yield, and J is the normalised spectral overlap between donor emission and acceptor absorption, and contains terms for the normalised donor emission f_D, and the molar extinction coefficient of the acceptor ϵ_A at a wavelength λ (144).

Since FRET competes with the normal radiative decay process of a fluorophore, which occurs at a rate of $k_r = \frac{1}{\tau_D}$, the efficiency of FRET (E) can be defined as

$$E = \frac{k_{ET}}{k_{ET} + k_r}$$

$$E = \frac{1}{1 + \left(\frac{R}{R_0}\right)^6}$$

R_0 is typically constant for most biological experiments (145), which means that E exhibits a strong inverse correlation to the sixth power of the inter-fluorophore distance, R. Consequently, FRET is highly sensitive to small inter-fluorophore distance

changes. For biological applications, donor-acceptor FRET pairs are generally chosen to detect distance changes between 1-10 nm, the distance range at which many biologically relevant conformational changes occur.

It is relatively straightforward to calculate an *uncorrected* FRET efficiency (E^*) from empirically observed fluorescence intensities of donor and acceptor fluorophores using

$$E^* = \frac{F_{DA}}{F_{DA} + F_{DD}}$$

Where F_{DD} and F_{DA} are donor and acceptor fluorescence intensities under donor excitation. However, in order to calculate accurate FRET values one must correct for factors such as background fluorescence, spectral cross-talk, and differences in the detection efficiencies and quantum yields of the donor and acceptor (146, 147). Although accurate FRET values can be used to extract true inter-fluorophore distances, in most cases, only relative changes in FRET efficiency are required to make biological interpretations of carefully designed experiments. As such, we will only consider uncorrected FRET efficiencies in the work presented in this thesis.

3.4 Single molecule detection techniques

Since the first observation of single molecules at room temperature (148), there have been steady technological advances that have increased the temperospatial resolution afforded by single-molecule techniques (149). Single-molecule fluorescence techniques generally deal with molecules diffusing in solution, or with molecules physically immobilized to a surface. The bulk of the work presented in this thesis was carried out using Total Internal Reflection Fluorescence (TIRF) microscopy to visualise surface-bound molecules. However, some control experiments were performed using confocal spectroscopy to probe diffusing molecules. I will therefore briefly discuss confocal spectroscopy, and then focus in more detail on TIRF microscopy.

3.4.1 Confocal spectroscopy and alternating laser excitation

In confocal spectroscopy, fluorescent molecules diffuse through a femtolitre-sized illumination volume where the fluorophores are excited and emit photons. Diffusion through the volume typically lasts ~1ms and gives a burst of fluorescence that is detected using an avalanche photo Diode (APD). In our experiments we combine single molecule detection techniques with Alternating Laser Excitation (ALEX), which as the name suggests, involves the alternation of donor and acceptor excitation lasers in order to directly monitor the fluorescence of the acceptor dye (150). This probing of the acceptor fluorescence introduces a new ratiometric variable, the stoichiometry (S), which is defined as:

$$S = \frac{F_{DA} + F_{DD}}{F_{DA} + F_{DD} + F_{AA}}$$

Where F_{DD} and F_{DA} are donor and acceptor fluorescence intensities under donor excitation, and F_{AA} is the acceptor fluorescence intensity under acceptor excitation. The stoichiometry permits us to virtually sort molecules (Fig. 3.2) and therefore to focus our analysis on those molecules containing the appropriate number of donor and acceptor dyes. Furthermore, because ALEX permits us to monitor the behaviour of the acceptor dye directly, we can detect whether any FRET changes we observe are due to photophysical phenomena of the acceptor as opposed to true FRET change due to changes in the distance separating the two fluorophores (151).

3.4.2 Total Internal Reflection Fluorescence (TIRF) microscopy

TIRF microscopy is a technique that enables the extended observation of fluorescent molecules immobilized on a surface (Fig. 3.3). Using this technique, one can monitor multiple molecules in parallel and thereby observe hundreds of biological reactions in real-time. This feature has made TIRF microscopy a favourite tool of single-molecule researchers investigating biological processes.

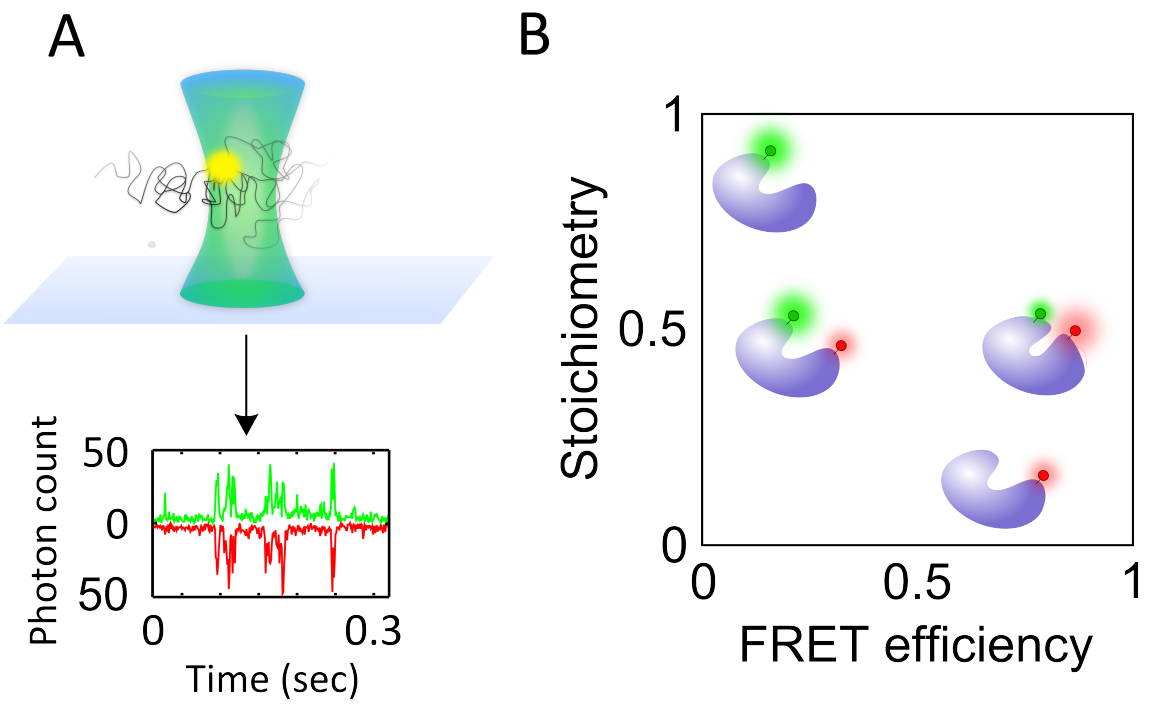


Figure 3.2. Confocal spectroscopy and ALEX. (A) Schematic representation of a fluorescent protein diffusing through a small illumination volume. The photons collected from the diffusing molecule produce bursts of fluorescence seen in the bottom figure. Here, the donor photon count is green and the acceptor count is red. (B) The fluorescence intensities in Fig. 3.2A are used to calculate FRET efficiency and stoichiometry, which are then plotted in a 2D-histogram. Molecules that contain only a donor or only an acceptor dye have stoichiometry values close to 1 and 0, respectively. Molecules labelled with both a donor and acceptor dye have stoichiometry values close to 0.5. The different species can therefore be virtually sorted along the y-axis. Figures adapted from Kapanidis Group communal figures.

Total internal reflection at a glass/solution interface generates an evanescent wave that illuminates the glass surface and a small volume that extends 100-200 nm into the solution. This permits the observation of molecules close to the surface while reducing background fluorescence from the surroundings. Total internal reflection occurs when light from a laser beam passes into a medium of a refractive index (n_1) from one of a higher refractive index (n_2), and when the angle of incidence (θ_i) of the light is greater than the critical angle (θ_c).

$$\theta_c = \sin^{-1}\left(\frac{n_1}{n_2}\right)$$

The resulting evanescent wave decays exponentially according to:

$$I(z) = I(0)e^{-z/d}$$

$$d = \frac{\lambda_0}{4\pi}(n_2^2 \sin\theta_i - n_1^2)^{-1/2}$$

Where z is the position above the surface. Two approaches are generally used to achieve total internal reflection. The first uses a prism and is known as prism-type TIRF, the second uses a sufficiently high numerical aperture (NA) objective, and is known as objective-type TIRF. Objective-type TIRF is technically more straightforward to implement than prism-type TIRF, and is the technique used throughout this thesis (Fig. 3.3). In TIRF microscopy, the fluorescence emitted from immobilized molecules is separated from the excitation light using a dichroic mirror and divided into donor and emission channels (Fig. 3.3C). The emitted light is detected using an electron-multiplying charged-couple device (EMCCD).

3.5 Photophysics

Organic fluorophores that are routinely used for biological investigations often display photophysical phenomena that can complicate data analysis and interpretation. The most notable phenomena are photobleaching and blinking.

Photobleaching is an irreversible process in which the photochemical properties of a dye are altered such that its ability to emit fluorescence is abolished. Performing experiments in the presence of an enzymatic oxygen scavenging system can often reduce the rate of photobleaching by removing oxygen that would otherwise induce photobleaching (152, 153). For experiments in which long time trajectories are necessary, it is often preferential to use exceptionally photostable dyes such as Cy3B and ATTO647N. Blinking is defined as transient periods during which a dye ceases to emit fluorescence. It occurs mainly when the dye enters the triplet state, and can be avoided by performing experiments in the presence of triplet state quenchers such as Trolox (154).

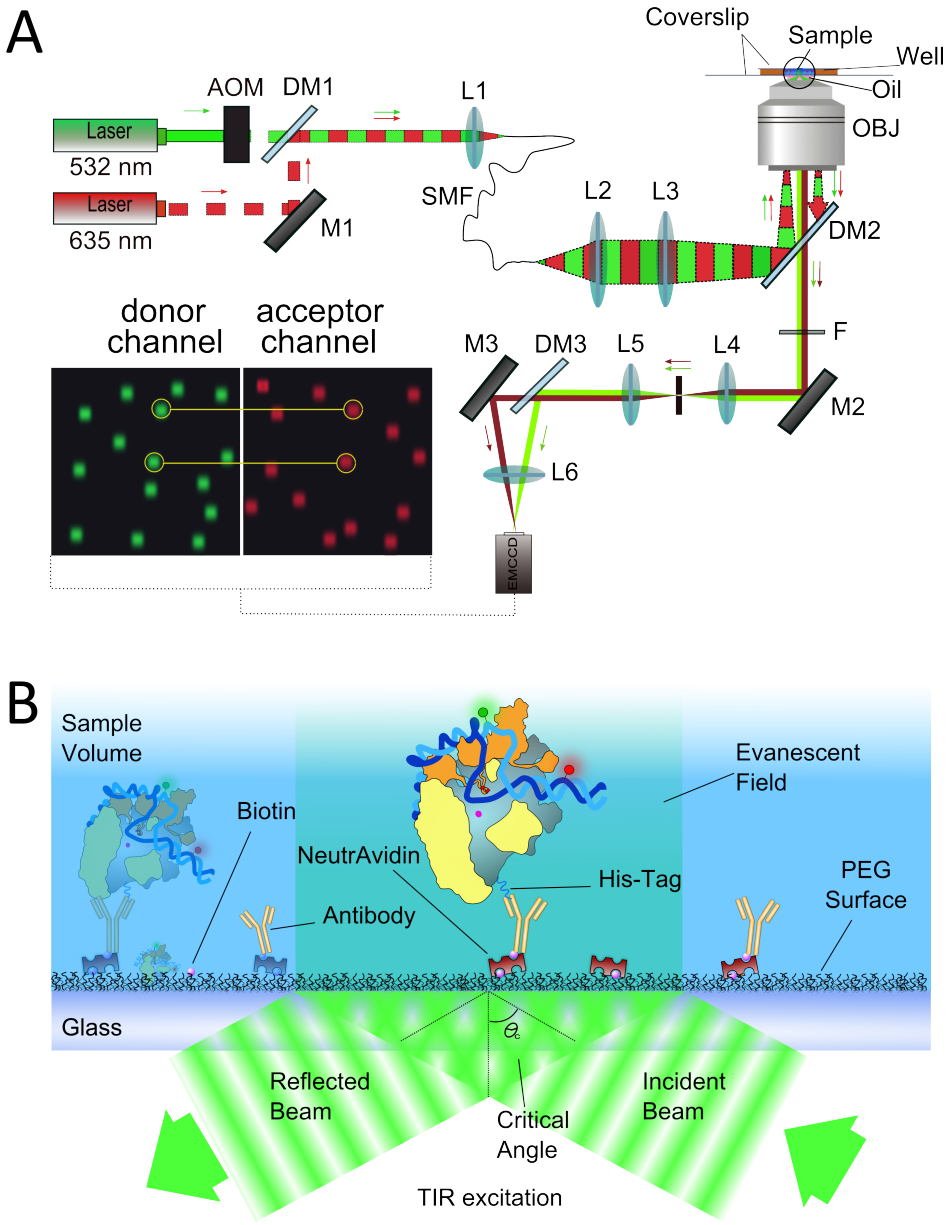


Figure 3.3. Schematic illustrations of TIRF microscopy. (A) Experimental set-up in objective-type TIRF. Green and red lasers are alternated and coupled into a single-mode fiber (SMF). The fiber output is then collimated and focused in the back-focal plane of the objective (OBJ), which produces a totally internally reflected beam. The fluorescence emission is filtered using a dichroic mirror (DM2) and cleanup filters (F). The emission signal is focused on a slit to crop the image, spectrally separated, and then imaged on an EMCCD camera, which produces a field-of-view consisting of single-molecule point-spread-functions in the donor and acceptor channels. These images are used to extract fluorescence intensities of molecules containing both a donor and an acceptor dye; two examples of such molecules are shown linked by straight yellow lines. (B) Zoomed view of the glass/solution interface in (A). A laser beam is totally reflected when it is incident at an angle greater than the critical angle. This produces an evanescent field in the solution above the glass surface, which excites multiple immobilized molecules that are imaged in parallel. The schematic shows RP_O immobilized using a penta-His antibody (see chapter 5). Figures adapted from Kapanidis Group communal figures. Not to scale.