Imperial College London

Single-Molecule studies of eukaryotic RNA Polymerase II

George A. Moore

Department of Infectious Disease

Imperial College London

Thesis submitted for the degree of Doctor of Philosophy

September 2022

Copyright Declaration

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC). Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose. When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes. Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law

Declaration of Originality

I hereby declare that all data, results, analysis and text within this thesis are my own except for the following places. Experiments in Chapter 4 carried out on the AT-rich sequence was in collaboration with the Jesper Svejstrup laboratory who originally conceived the project. Data in relation to Figure 4.7 was obtained by Zhong Han, a post-doc in his lab. All single-molecule data for this project was collected by myself. The purification of yeast and mammalian RNAPII was also carried out predominantly by Zhong Han. I have included this in my thesis as the smFRET data collected provided valuable insight into the project as part of a collaboration and adds extra insight into the dynamics of RNAPII. Chapter 5 includes both biochemical and single-molecule experiments carried out by myself on damaged DNA templates using the yeast repair factor Rad26, which was purified and gifted to us by Prof. Dong Wang.

I also declare that any use of sources, publications, figures or text has been appropriately cited in the text and the bibliography. Similarly, the copyright and permission for the reuse of figures or images have been obtained for each relevant figure and publication where needed.

Acknowledgements

I would like to thank my supervisor, Prof. David Rueda for giving me this amazing opportunity to work in his lab and for all the support he has given me over the past 6 years. His guidance, experience and mentorship has been invaluable throughout in helping me to drive this project forwards and guide my career. I would also like to thank my co-supervisor Prof. Xiaodong Zhang and our collaborators Prof. Jesper Svejstrup, Prof. Dong Wang and Dr Zhong Han for providing me with valuable knowledge, experience and support throughout my PhD. It has been a truly amazing opportunity working in this lab, one that I will not soon forget.

I would also like to thank every member of the Rueda lab who I have met along the way, who have always been so friendly and kind, providing support whenever I needed it. Dr. Adam Cawte, Elena Crameri, Dr. Maru Fuentes-Perez, Dr. Hailey Gahlon, Dr. Paul Girvan, Dr. Mohammed Ghoneim, Dr. Artur Kaczmarczyk, Dr. Haruki Lino, Marialucrezia Losito, Paul-Poudevign-Durance, Quentin Smith, Dr Sheila Xie, Dr. Anita Meier, Dr. Erin Cutts and Dr. Tianyi Zhang. To all of you, I say thank you for always making the lab a great place to work over these past 4 years.

I would like to make a special thanks to Hailey who trained me when I first arrived in the lab and inspired me to pursue academic research and my PhD, without you I would not be where I am today. A big thanks to Paul and Marilú, who have become two of my closest friends. Thank you for all the support you have given me both within and outside the lab, without you guys I am not sure how I would have endured such an experience.

Finally, a big thank you to all of my friends and family who have always been a constant source of encouragement and support throughout my PhD, I have appreciated it more than you can know.

Abstract

RNA transcription is the essential first step in the conversion of genetic information stored in DNA into fully functioning proteins. RNA polymerases read this genetic information and make a complimentary form, called RNA. RNA Polymerase II is responsible for reading and converting most of the protein coding genes. However, eukaryotic DNA is highly complex and contains multiple roadblocks to transcription including nucleosomes and DNA damage. If these damages are not resolved, mutations can arise causing genome instability and also cancers including melanoma and adenocarcinoma. How these roadblocks affect RNAPII on a molecular level has yet to be probed deeply at a single-molecule level due to the dynamic nature of this process. Here, I develop novel single-molecule based assays to directly monitor RNAPII elongation dynamics without the need to genetically modify the polymerase. Using the smFRET based assay I was able to visualise RNAPII elongation on different templates including AT-rich DNA. I showed that AT-rich DNA not only induces premature termination of transcription, but also that severe pausing precedes this step. Addition of different damages into the template also caused destabilising effects on RNAPII to varying amounts. While abasic lesions caused complete stalling of RNAPII, UV-induced damage and 8-Oxo Guanine groups led to more dynamic pausing and backtracking behaviour, with RNAPII bypassing the latter damage. Finally, Rad26, a recovery factor, was shown to promote bypass of oxidative damage by physically blocking the backtracked state and keeping the polymerase around the site of damage. The assays developed here can be utilised for a variety of purposes where other roadblocks, e.g. nucleosomes, can be inserted readily into the template. Furthermore, the results shown here will be important in increasing our understanding of how DNA damage and other roadblocks interfere with the transcription pathway.

Contents

1 Introduction **21**

1.1 The Central Dogma . . . 22

1.2 Transcription and RNA Polymerases . . . 24

1.2.1 Background . . . 24

1.2.2 Single-subunit RNA Polymerases . . . 25

1.2.3 Multi-subunit RNA Polymerases . . . 26

1.3 Eukaryotic RNA Polymerase II Cycle . . . 33

1.3.1 Overview . . . 33

1.3.2 Initiation . . . 35

1.3.3 Elongation . . . 42

1.3.4 Termination . . . 50

1.4 Roadblocks to Transcription . . . 54

1.4.1 Overview . . . 54

1.4.2 Abasic lesions . . . 56

1.4.3 Oxidative damage . . . 58

1.4.4 UV-induced damage . . . 60

1.4.5 Damage response pathways . . . 61

1.5 Single-molecule studies of transcription . . . 67

1.6 Thesis Objectives . . . 73

2 Background Theory **103**

2.1 Single-Molecule Imaging . . . 104

2.2 Single-Molecule TIRF Microscopy . . . 105

2.2.1 Principles of Total Internal Reflection Microscopy . . . 106

2.2.2 Single-Molecule Forster Resonance Energy Transfer (FRET) 108
2.3 Optical Trapping Microscopy 113
2.3.1 Properties of Optical Tweezers 114
2.3.2 Properties of optically trapped DNA molecules 117
2.4 Single-Molecule studies of RNAPII elongation 118

3 Materials and Methods **127**
3.1 Preparation of DNA and RNA Oligonucleotides 128
3.1.1 Deprotection of DNA oligonucleotides 128
3.1.2 Purification of DNA and RNA oligonucleotides 129
3.1.3 Fluorescent labelling of DNA and RNA oligonucleotides 129
3.2 Purification of eukaryotic RNA Polymerase II 130
3.2.1 Purification of yeast RNAPII 130
3.2.2 Purification of mammalian RNAPII 131
3.3 *In vitro* transcription assay 132
3.4 TIRF Microscopy 133
3.4.1 Slide passivation 133
3.4.2 smFRET experiments 134
3.4.3 smFRET Analysis 136
3.5 Optical Tweezer Spectroscopy 137
3.5.1 Biotinylation of DNA 137
3.5.2 Generation of hybrid ss-/ds-DNA constructs 138
3.5.3 Ligation method to generate long TEC constructs 140
3.5.4 Autosticky PCR method to generate long TEC constructs 140
3.5.5 Optical tweezers preparation 142
3.5.6 Optical tweezers assay 142
3.5.7 Optical tweezers analysis 142

4 Visualisation of RNAPII transcription elongation dynamics via single-molecule FRET **147**
4.1 Abstract 148
4.2 Results 149
4.2.1 Visualisation of FRET dependent on RNA primer length 149

4.2.2 Visualisation of RNAPII dynamics using smFRET 152
4.2.3 AT-rich DNA induces premature disengagement from DNA 158
4.2.4 AT-rich DNA induces RNAPII pausing before disengagement from DNA 160
4.2.5 NTS substitution allows permissive transcription with no premature termination 162
4.3 Discussion 165
4.4 Supplementary Figures 172
4.5 DNA and RNA Sequences 174

5 DNA damage induces heterogeneous stalling and backtracking of RNA Polymerase II 179
5.1 Abstract 180
5.2 Results 181
5.2.1 RNAPII dynamically stalls in absence of all nucleotides 181
5.2.2 Abasic lesions cause static stalling of RNAPII 185
5.2.3 Oxidative damage and bulky lesions induce dynamic stalling of RNAPII 187
5.2.4 Rad26 stimulates bypass of DNA lesions by preventing backtracked state 193
5.3 Discussion 196
5.4 Supplementary Figures 206

6 Visualisation of RNAPII dynamics via optical tweezer based assays 213
6.1 Abstract 214
6.2 Results 215
6.2.1 DNA handle overhang length drives efficient TEC ligation 215
6.2.2 TEC with long DNA handles permits RNAPII transcription 217
6.2.3 16 kb TEC permits transcription in bulk but not at single-molecule level 220
6.2.4 Generation of hybrid ss-/ds-DNA constructs 221
6.2.5 Spurious binding of RNAPII driven by DNA accessibility 224
6.3 Discussion 227

7 Conclusion **239**

7.1 Thesis Discussion 240

7.2 Future Work 245

List of Figures

1.1 The Central Dogma 23
1.2 Different forms of RNA 24
1.3 Architecture of T7 RNA Polymerase 26
1.4 Architecture of RNA Polymerases 29
1.5 Comparison of evolutionary conserved polymerases 31
1.6 Structure of RNA Polymerase II 34
1.7 RNAPII Transcription Cycle 36
1.8 Initiation of RNAPII transcription 37
1.9 Role of Mediator complex at promoters and enhancers 40
1.10 Structure of RNAPII elongation complex 43
1.11 Nucleotide Addition Cycle 44
1.12 Structure of RNAPII EC bound to elongation factors 49
1.13 Models of RNAPII transcription termination 51
1.14 Effect of DNA roadblocks on transcription 55
1.15 Structures of DNA damage 57
1.16 Base Excision Repair (BER) Pathway 62
1.17 Transcription Coupled Nucleotide Excision Repair (TC-NER) Pathway . . 65
1.18 smFRET assay revealed scrunching mechanism of RNAP initiation 69
1.19 Optical tweezer assay to monitor RNAPII elongation 72

2.1 Principles of TIRFM 107
2.2 Principles of smFRET 110
2.3 Schematic of smFRET TIRF setup 112
2.4 Principles of Optical Tweezers 116

3.1 Chemical structures of fluorescent dyes . . . 130
3.2 Assembly of smFRET slides . . . 135
3.3 DNA Constructs for Optical Tweezers . . . 138
3.4 Schematic of DNA constructs generated via autosticky PCR . . . 141
3.5 Schematic of C-trap microfluidic system . . . 143

4.1 Long RNA primer leads to poor FRET . . . 150
4.2 Decreasing RNA length increases FRET efficiency . . . 151
4.3 RNAPII transcribes in an NTP dependent manner . . . 153
4.4 Schematic of smFRET TIRF based assay . . . 154
4.5 RNAPII remains in pre-transcription state without NTPs . . . 155
4.6 RNAPII transcription leads to loss of FRET . . . 156
4.7 RNAPII disengages on AT-rich DNA . . . 159
4.8 AT-rich DNA perturbs transcription . . . 161
4.9 AT-rich DNA induces RNAPII pausing . . . 162
4.10 NTS substitution does not induce RNAPII pausing in AT-rich region . . . 163
4.11 Immobilised TEC with NTS labelled . . . 172
4.12 RNA calibration . . . 173

5.1 Artificially stalled RNAPII dynamically stalls . . . 183
5.2 Dwell time analysis of RNAPII transitions . . . 184
5.3 Abasic lesions cause static stalling of RNAPII . . . 187
5.4 Oxidative damage and Bulky lesions induce stalling of RNAPII . . . 189
5.5 RNAPII dynamically stalls at oxidative damage . . . 190
5.6 TT-Dimer causes both dynamic and static stalling of RNAPII . . . 191
5.7 RNAPII stalling dynamics are not NTP dependent . . . 193
5.8 Rad26 induces bypass of DNA damage . . . 194
5.9 Rad26 prevents long range backtracking of RNAPII . . . 195
5.10 RNAPII stalling mechanism . . . 196
5.11 RNAPII transcribes in an NTP dependent manner . . . 206
5.12 Kinetic Rates of Rad26 . . . 207
5.13 Rad26 binds to RNAPII containing TEC . . . 208

6.1 8 kb TEC Formation . . . 216

6.2 16Kb TEC Formation 219
6.3 Optical Tweezers of 16Kb TEC 220
6.4 Hybrid ss/ds-DNA Validation 223
6.5 Hybrid ss-/ds-DNA validation via fluorescence 224
6.6 RNA binding to hybrid DNA Construct 225
6.7 RNAPII binding non-specifically to ssDNA perturbs transcription 226
6.8 T7 Polymerase Transcription on optical tweezers 231

List of Tables

2.1 Biophysical parameters of DNA 118

3.1 DNA Sequences for tortionally constrained DNA 139
3.2 Thermocycler Conditions for autosticky PCR 141

4.1 DNA and RNA sequences used in initial optimisation of smFRET assay . . 174
4.2 DNA and RNA Sequences used for smFRET assay 175
4.3 DNA and RNA Sequences used for AT-rich DNA assay 175

5.1 DNA and RNA Sequences used 182
5.2 Dwell time analysis of RNAPII Transitions 185

6.1 DNA Sequences used for 8Kb TEC 217
6.2 DNA Sequences used for 16Kb TEC 218
6.3 DNA Sequences used for Hybrid DNA 222
6.4 Expected and Actual Gap Sizes 223

Chapter 1

Introduction

1.1 The Central Dogma

All life depends on the production of fully functional proteins to carry out essential biological processes and the replication of their genetic information. In general, life is said to be made up of three main components. DNA, where the genetic information is stored in 4 different bases, Adenine (A), Guanine (G), Cytosine (C) and Thymine (T), and the order of these bases determines the protein to be produced. Proteins, made up of amino acids forming poly-peptide chains which are encoded for by DNA. And finally RNA, another nucleic acid consisting of 4 bases (A, G, C and Uracil (U)) that is an essential intermediate in the conversion of DNA into proteins. DNA replication is the synthesis of more copies of the DNA template and is essential for the reproduction of an organism. However, by itself DNA cannot synthesise the proteins essential for life, but instead must first be converted into RNA via transcription. This RNA molecule now encodes the DNA sequence of a particular protein and will often undergo modifications before being translated into proteins (Figure 1.1) (Koonin, 2012). This most basic conversion in life is termed by many biologist as 'The Central Dogma' and first coined by Francis Crick (Crick, 1970). Although these are essential for DNA based organisms, organisms such as viruses often have RNA genomes and must replicate them like our DNA based genome (RNA replication) or convert it into a DNA based structure to use host DNA polymerases (reverse transcription) (Figure 1.1).

RNA is one of the most important biomolecules of life, playing a vital role not only in the transfer of genetic information (messenger RNA, mRNA) but playing a functional role also. There are several types of RNA molecules that are essential for life, mRNA, transfer RNA (tRNA), ribosomal RNA (rRNA) and small nuclear RNA (snRNA) (Brosius and Raabe, 2016). tRNA is an essential RNA molecule in the synthesis of proteins, where they

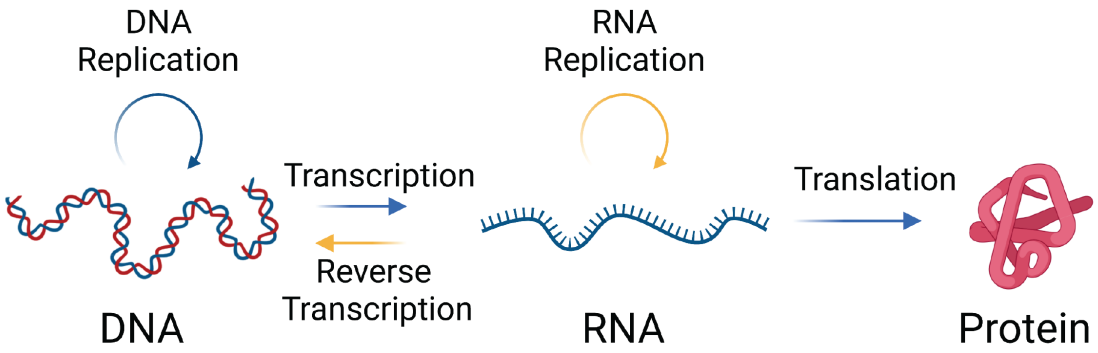


Figure 1.1: **The Central Dogma.** The three main components of life DNA, RNA and proteins and the processes that help to synthesise and convert them. Blue arrows depict essential processes in all organisms. Orange arrows show more specialised processes. Made with BioRender

carry amino acids to the ribosome catalysing the polypeptide formation. While snRNA is a class of RNA molecules that associate with the spliceosome and are important in the splicing of RNA to form mature RNA (Matera et al., 2007). rRNA also associates with peptides and forms part of the ribosome, which in turn is used to translate mRNA into a polypeptide chain. While these RNA molecules are some of the best understood, other RNA molecules such as long non-coding (lncRNA) have a vital role in gene regulation but have no protein coding information and are poorly understood. Furthermore, RNA has a propensity to form secondary structures in the forms of hairpins, ribozymes and riboswitches all of which contribute to the regulation of our genes (Figure 1.2 (Walter and Engelke, 2002; Statello et al., 2020; Ganser et al., 2019).

The catalysis of almost all RNA molecules is carried out by DNA dependent RNA polymerases (RNAP), where the genetic code stored in double-stranded DNA (dsDNA) is read by the polymerase and transcribed using the 4 nucleoside triphosphates (NTPs) to generate an RNA molecule containing the correct sequence. RNA polymerases are highly conserved proteins and have evolved in almost all forms of life, including bacteria, viruses, archaea and eukaryotes (Werner and Grohmann, 2011).

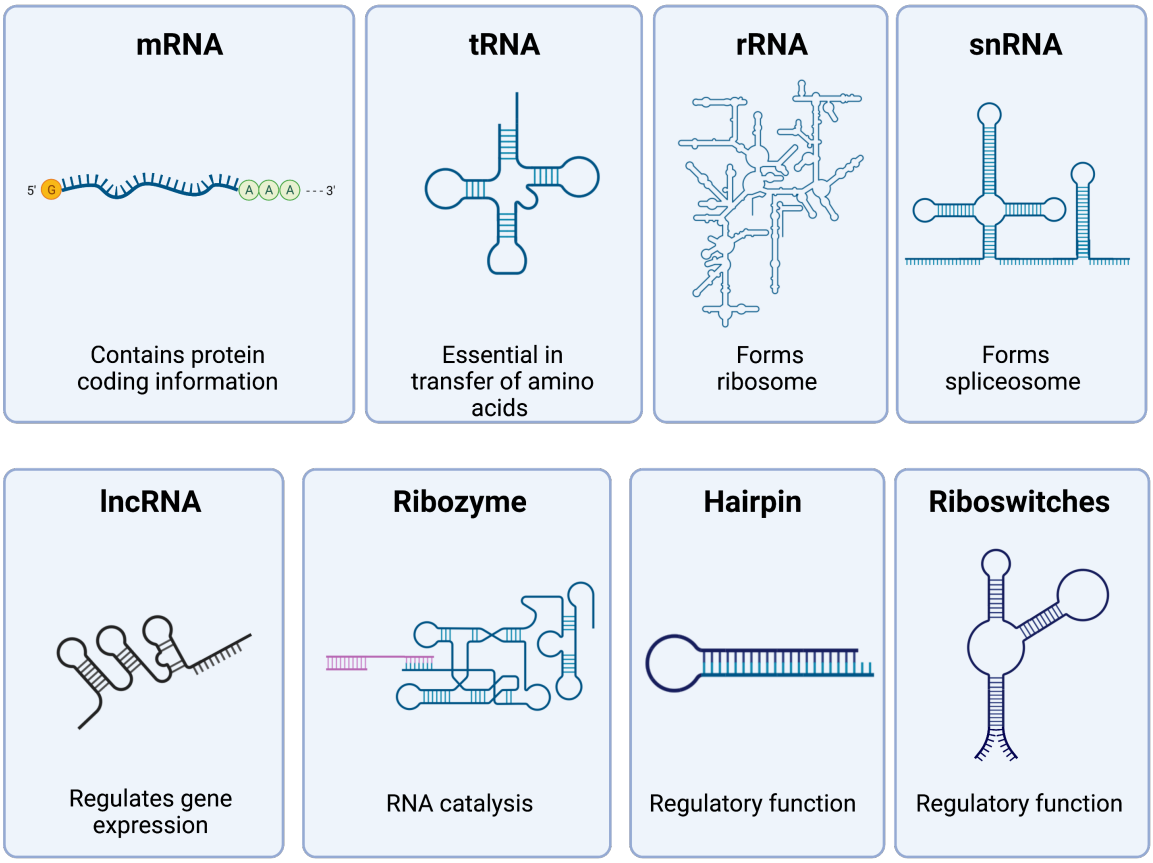


Figure 1.2: **Different forms of essential RNA and their function in the cell.** Examples of different types of RNA found within living organisms and their main role within the cell. Made with BioRender

1.2 Transcription and RNA Polymerases

1.2.1 Background

RNA polymerases are essential factors in the synthesis of RNA from the DNA template. They were first hypothesised to exist in the early 1950s but it was not till 1961 where it was shown that the RNA sequence encoded the same sequence as the DNA and the first enzyme to show a DNA dependent RNA synthesis was isolated from *Micrococcus lysodeikticus* (FURTH et al., 1961; Weiss and Gladstone, 1959).

RNA polymerases have been studied heavily due to their vital role in protein synthesis since their initial discovery with polymerases being identified in organisms ranging from basic prokaryotic life all the way to humans. Early structural studies focused on phage

and bacterial polymerases and soon after polymerases from eukaryotic samples were isolated where three distinct classes of polymerases were found (Cramer, 2019b; Roeder and Rutter, 1969; Burgess, 1971). Since then, our understanding of the structure and function of polymerases has deepened even further by the use of new structural, biochemical and biophysical assays, with Roger Kornberg winning the Nobel prize in 2006 for his pioneering work on transcription.

1.2.2 Single-subunit RNA Polymerases

Due to the importance of RNA synthesis, organisms have developed processive and accurate polymerases to catalyse this reaction. Some of the most basic forms of RNA polymerases are found in microorganisms such as bacteriophages where RNAP is made up of a single subunit, for example T7, T3 and SP6 polymerase (Werner and Grohmann, 2011). T7 polymerase has a molecular weight of around 100kDa consisting of a single modular component containing several sub-domains (Cheetham and Steitz, 2000). The biggest of these sub-domains is the N-terminal domain, which is common between most bacteriophage polymerases and is responsible for DNA binding and melting. Three further sub-domains termed the palm, fingers and thumb make up the rest of T7 and form a 'right-handed' structure and help to carry out the catalytic activity of RNA synthesis (Cheetham et al., 1999) (Figure 1.3).

By itself, the core of T7 polymerase is able to synthesise RNA with only a DNA template and metal ions and does not need extra factors to associate, unlike for more complex organisms (Cheetham et al., 1999). Although they do not require initiation or elongation factors to function, bacteriophage polymerases do exhibit a sequence-dependence where initiation only occurs at a specific start site, with different phage polymerases exhibiting

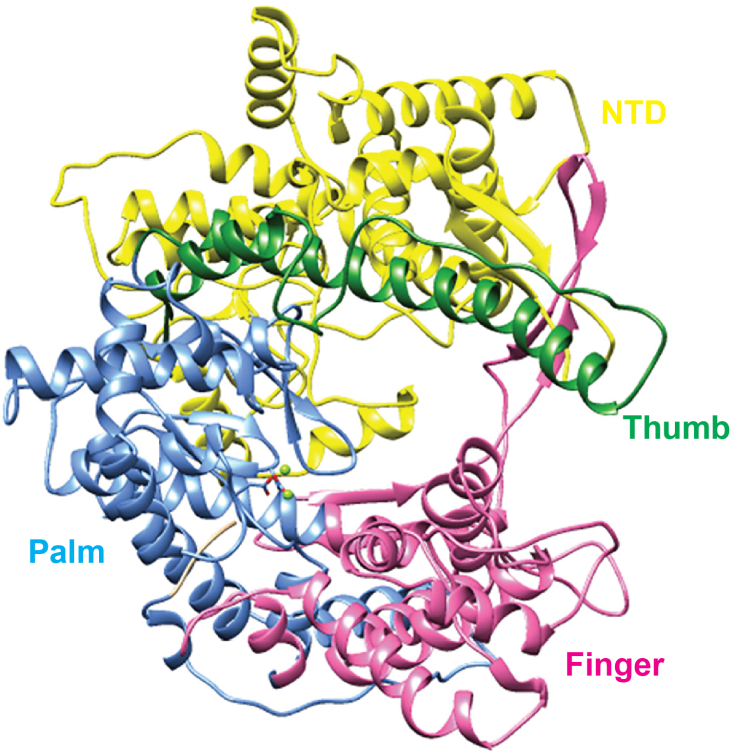


Figure 1.3: **Architecture of T7 RNA Polymerase.** Structure of T7 RNA Polymerase showing the core N-terminal domain (NTD, yellow) and the fingers (pink), palm (blue) and thumb (green) domains. Figure adapted from (Borkotoky and Murali, 2018)

specific sequence dependency (Gholamalipour et al., 2018). The T7 RNAP start site is 5'TAATACGACTCACTATAG, where transcription begins at the final G. It is due to this simplistic nature of bacteriophage polymerases that T7 polymerase has become widely used in molecular biology assays to directly synthesise RNA of any desired length and sequence.

1.2.3 Multi-subunit RNA Polymerases

Unlike single subunit polymerases, multi-subunit polymerases are more complex not only in their structure but also in their function. For example, although multi-subunit polymerases do have sequence specificity like T7 or T3, they cannot bind to DNA in a transcriptional competent way by themselves and require a host of auxiliary factors to bind in

concert to initiate transcription (Steitz, 2009; Cramer, 2019a). Multi-subunit polymerases are usually also bigger in size and structure ranging up to 500-600 kDa. Although multi-subunit polymerases are found in a whole range of organisms they do contain a high amount of conservation in relation to their structure and core function suggesting that bacteria, archaea and eukaryotes all use an RNA polymerase based off their last universal common ancestor (LUCA)(Werner, 2007; Werner and Grohmann, 2011). In general, multi-subunit polymerases form a 'claw-like' structure where multiple subunits form a pincer interacting with downstream DNA from the binding site (Zhang et al., 1999; Cramer et al., 2001; Korkhin et al., 2009) (Figure 1.4A). This downstream DNA is zipped apart and fed into the active site of the polymerase, allowing for NTPs to bind and begin RNA synthesis. The active site contains two major regions, the bridge helix and trigger loop both of which are essential in the catalytic function of RNA polymerases. As this nascent RNA increases in size, it will exit the active site perpendicular to the downstream DNA through the exit channel of the polymerase (Gnatt et al., 2001). In archaea and eukaryotes, this will interact with the stalk domain of the polymerase, whereas in bacteria although there is no polymerase subunit to associate with auxiliary factors are often bound to the exiting RNA to stabilise it (Hirtreiter et al., 2010). RNA polymerase will then continue along the template DNA continually synthesising RNA until a stop region is reached.

Bacterial RNA Polymerase

Bacterial RNA polymerases are one of the simplest forms of the multi-subunit class of RNA polymerases weighing in at just under 400 kDa (Murakami, 2015). The core of the polymerase has 5 subunits, a dimer of α subunits, β, β' and ω and it is these subunits that form the 'claw' like structure discussed earlier (Ghosh et al., 2010). These core

subunits are also found conserved in archaea and eukaryotes forming the core catalytic unit. β and β' form the pincers of this claw and are the largest subunits of bacterial RNAP (around 150 kDa each), separated by a 3 nm deep cleft that forms the active site (Zhang et al., 1999) (Figure 1.5A). α domains are made up of a N-terminal region (NTR) connected to the C-terminal region (CTR) via a flexible linker. The NTR of each subunit dimerise to each other and interact with the β and β' subunits causing stabilisation of the core polymerase (Gourse et al., 2000). Although the smaller CTR does not aid stability, it does interact with DNA and proposed to be important for promoter recognition (Haugen et al., 2008). The ω subunit is the smallest subunit of bacterial RNAP and its function is not as well understood, but is thought to be important for stabilisation of the core unit and may also have a role in the stress response pathway (Mathew and Chatterji, 2006; Weiss et al., 2017). In general, bacterial RNAP is formed of two types of domains, those that are stable (both α subunits, ω and the active site regions of β and β') and those that are more mobile and transition during transcription. These include many sub-regions of the β and β' subunits including the β'-clamp, β'-jaw and the β-flap. These subdomains will transiently move during the initiation and elongation phase of transcription to accommodate the DNA and translocate along the template (Ghosh et al., 2010).

As discussed earlier, bacterial RNAP cannot bind to promoter DNA by itself but instead requires additional factors to find its target before opening the DNA. Sigma (σ) factors are proteins found in bacteria which associate transiently with RNAP forming a holoenzyme to induce promoter recognition and association with necessary transcriptional activators and repressors (Burgess et al., 1969; Feklístov et al., 2014).

Usually, a single σ-factor is responsible for promoter recognition and basal transcription,

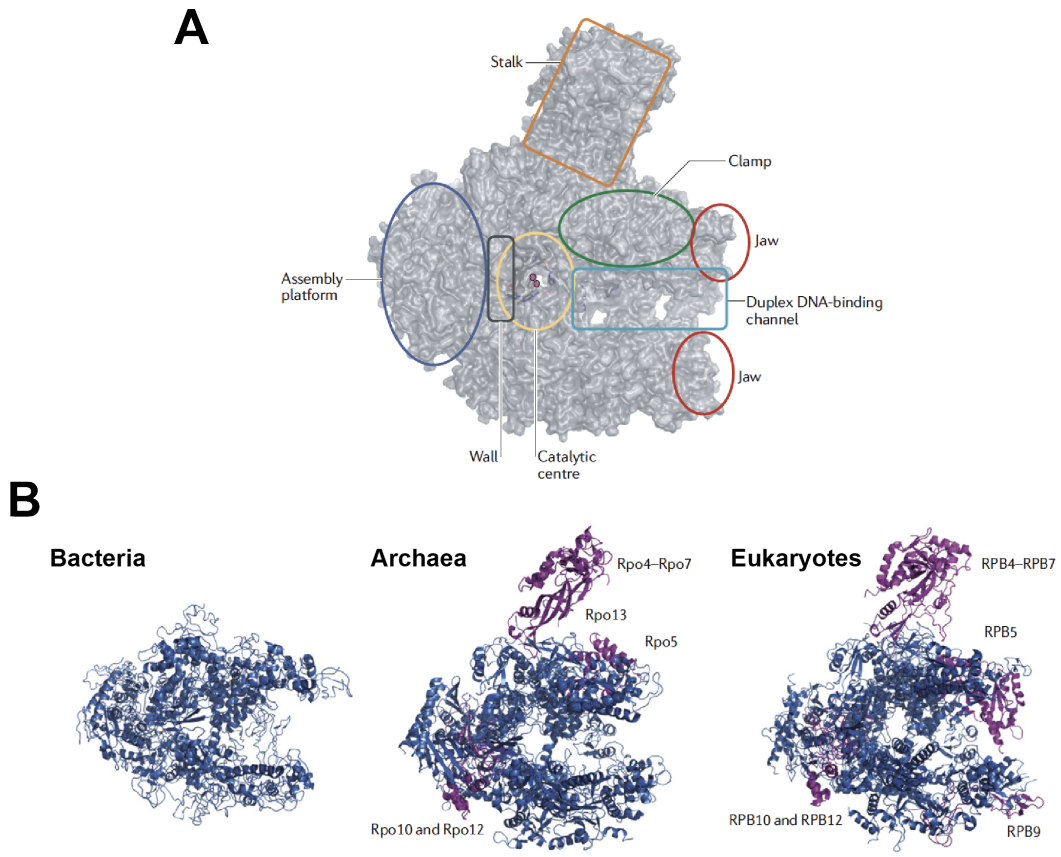


Figure 1.4: **Overall structure of RNA polymerase from bacteria, archaea and eukaryotes.** (A) Structure of RNA polymerase showing key structural features found in archaeal and eukaryotic polymerase including stalk, catalytic centre, clamp, jaw and assembly platform. PDB 1Y1W (B) Comparison of polymerase structure between bacteria, archaea and eukaryotes showing which domains are conserved (blue) and which ones are unique to archaea and eukaryotes (purple). Figure adapted from (Werner and Grohmann, 2011)

although this can be changed upon an external or internal stress (Buck et al., 2000). For this reason, bacteria generally contain several σ-factor genes encoded in their genome, (*Streptomyces coelicolor* encodes 63) but often only express one at a time (Gruber and Gross, 2003). σ- factors are broken down into two main classes based on their function and sequence. σ^{70} is the largest class of σ-factors and is responsible for basal transcription containing multiple sub-classes in *E.coli*($\sigma^{70,38,32,28,24,19}$) (Gruber and Gross, 2003). The second minor class of σ-factors, σ^{54}, has little conservation to the σ^{70} class in respect to sequence homology and promoter specificity. Furthermore, unlike the major class of σ-factors, σ^{54} only contains one σ-factor. σ^{54} has been shown to be induced during the

transcription of genes under various stresses including nitrogen starvation, phage shock and cytotoxicity (Buck et al., 2000). Although both classes bind to RNAP in similar positions, they can alter the function and target of RNAP.

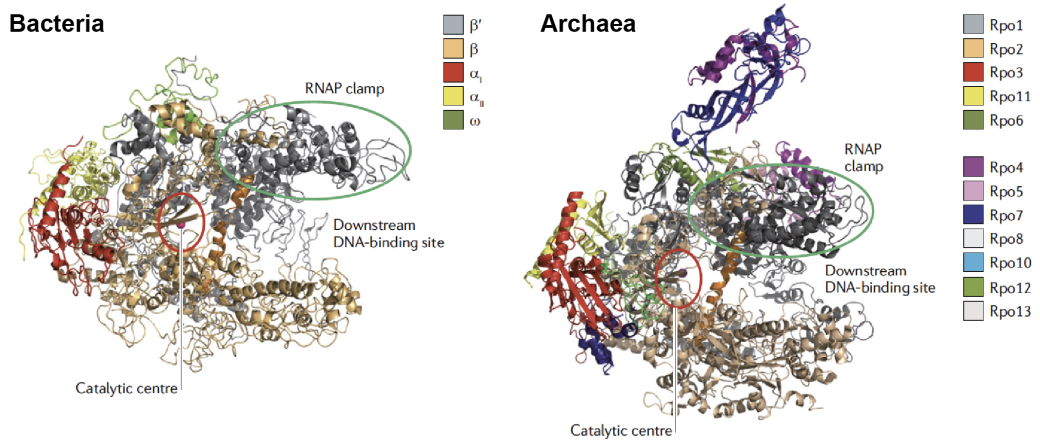
Our understanding of bacterial RNAP structure originated from the thermophillic bacteria *Thermus aquaticus* and *thermophilus* but the structure from *Escherichia coli* was later solved and is now more commonly used for structural, biochemical and biophysical analysis to probe its function (Zhang et al., 1999). The core function of RNA synthesis is conserved between bacteria and eukaryotes and will be discussed in more detail later.

Archaea and Eukaryotic RNA Polymerases

Although archaeal and eukaryotic polymerases are larger and more complex (made up of 12 subunits, over 500 kDa) than bacterial RNAP, the 'core' of each polymerase is highly conserved between them (Werner and Grohmann, 2011) (Figure 1.5). The pincers of the polymerase (β and β' in bacteria) correspond to Rpb1 and Rpb2 in eukaryotes and Rpo1 and Rpo2 in archaea and make up over 50% of the total polymerase weight (350 kDa). The α subunits correspond to Rpb/Rpo3 and 11 while the ω subunit corresponds to Rpb/Rpo6. While these may form the core of the polymerase carrying out the polymerisation function, there are other essential subunits found in archaea and eukaryotes that encode accessory functions (Werner, 2007) (Figure 1.4B). One of these main differences is the 'stalk' region found on the periphery of the archaeal and eukaryotic polymerase (Rpb4 and Rpb7/Rpo4 and Rpo7). Early structural studies often did not contain these subunits as in yeast they can reversibly bind to the core polymerase, although this is not true in archaea where it is stably bound as one complex (Cramer et al., 2001; Grohmann et al., 2009; Orlicky et al., 2000). The role of the stalk varies during the transcription

cycle including promoting structural changes to allow initiation and during elongation increases processivity (Hirtreiter et al., 2010).

A



B

Bacteria	Archaea	Eukaryotes		
		RNAPII	RNAPIII	RNAPI
RNAP subunits				
β-subunit	Rpo1 (RpoA)	RPB1	C160	A190
β-subunit	Rpo2 (RpoB)	RPB2	C128	A135
α-subunit	Rpo3 (RpoD)	RPB3	AC40	AC40
α-subunit	Rpo11 (RpoL)	RPB11	AC19	AC19
ω-subunit	Rpo6 (RpoK)	RPB6	RPB6	RPB6
	Rpo5 (RpoH)	RPB5	RPB5	RPB5
	Rpo8§ (RpoG)	RPB8	RPB8	RPB8
	Rpo10 (RpoN)	RPB10	RPB10	RPB10
	Rpo12 (RpoP)	RPB12	RPB12	RPB12
	Rpo4 (RpoF)	RPB4	C17	A14
	Rpo7 (RpoE)	RPB7	C25	A43
		RPB9	C11	A12
	Rpo13§			

Figure 1.5: **Comparison of RNA polymerase structures and domains between evolutionary conserved polymerases.** (A) Structure of bacterial (right) and archaeal (left) RNA polymerase showing the different subunits found within the structure and which subunits are conserved between them (see colour label). PDB 1L6V and 2WAG respectively (B) Table showing the nomenclature and conservation of subunits found within bacterial, archaeal and the three main eukaryotic RNA polymerases. Figure adapted from (Werner and Grohmann, 2011)

Other subunits of the polymerase are also involved in promoting initiation by aiding DNA melting. The C-terminal domain of Rpb5 has high homology to Rpo5 in archaeal polymerase, and both subunits form multiple discrete contacts with the largest subunit Rpb1 (Rpo1 in archaea), and in turn forms the 'jaw domain' (Grünberg et al., 2010).

This jaw interacts with downstream DNA from the promoter region and induces DNA bubble formation, this region will then undergo severe structural changes throughout the transcription cycle from initiation to elongation. Rpb8 is found on the underside of the polymerase near to the exit pore and has also been shown to have a similar role in DNA melting (Kang et al., 2006). It contains a nucleic-acid binding domain (OB-fold) which can interact not only with the DNA but also possibly the 3' end of backtracked RNA.

While these subunits play important functional roles, other subunits of these polymerases play more of a structural role, such as Rpb3, 10, 11 and 12 which are part of the 'assembly platform' (Hantsche and Cramer, 2017). Together these subunits form a stable complex associated with the core. Rpb10 and 12 bind to one of the pincer domains (Rpb2) and act as a bridge connecting it to Rpb3 inducing substantial inter-domain contacts increasing the stability of this large complex. Furthermore, structural studies have shown that they can interact with initiation factors such as TFIIB thereby tethering the polymerase to the larger initiation complex (Kostrewa et al., 2009; Hantsche and Cramer, 2017).

Although many of these subunits share homology and have conserved functions between organisms, eukaryotes have evolved further complexity to this process by having multiple polymerases carrying out distinct functions, whereas bacteria and archaea encode only one to carry out their entire RNA synthesis needs. Most eukaryotic organisms contain three main polymerases, RNAPI, II and III (Vannini and Cramer, 2012). The reason for this is likely due to the complex nature of RNA synthesis, where various types of RNA need to be synthesised in a regulated manner (Section 1.1). While RNAPI is mainly responsible for synthesising rRNA and non-coding RNA, RNAPIII synthesises smaller RNAs such as tRNA. RNAPII is the main polymerase in eukaryotes and is responsible for the synthesis of mRNA from the protein coding genes. Structurally they are all quite

similar, but RNAPI and III do associate with up to 2 or 5 extra factors respectively and often make distinct interactions with unique regulatory factors (Vannini and Cramer, 2012) (Figure 1.5B). Aside from these three, eukaryotes also encode other polymerases such as RNAPIV and V found in plants and also intriguingly a single-subunit polymerase found in mitochondria, often used by many viruses when attacking a host.

Just as bacterial polymerases bind to sigma-factors, eukaryotic RNAPII also associates with a multitude of regulatory factors during the transcription cycle to aid in its core polymerisation function. These proteins will bind dynamically and reversibly to the polymerase to encode different functions such as initiation, elongation, termination or damage response. The Cramer group has been heavily involved in furthering our understanding of the structure of RNAPII and its associated regulatory factors in both the initiation and elongation state having recently solved several structures (Farnung et al., 2018; Vos et al., 2018b,a; Bernecky et al., 2016).

1.3 Eukaryotic RNA Polymerase II Cycle

1.3.1 Overview

RNAPII is one of the most vital polymerases in eukaryotes, synthesising the bulk of protein coding genes. Our current understanding of the structure and function of RNAPII originated over 20 years ago when the first X-ray crystallography structure of the 10-subunit core RNAPII was solved, aided by the ability to purify large amounts of endogenous RNAPII (Thompson et al., 1990; Cramer et al., 2000). It was then only a few years later that the complete 12 subunit structure of RNAPII was obtained by addition of the Rpb4/7 subunits (Armache et al., 2003; Bushnell and Kornberg, 2003).

As mentioned above, RNAPII is made up of 12 subunits which interact together to form essential functional domains. One of the most important features of RNAPII is the cleft that is formed by the two largest subunits Rpb1 and Rpb2 (Osman and Cramer, 2020). At the base of this cleft lies a funnel-like pore that enables incoming NTPs to enter the polymerase and bind to the highly conserved and essential active site found next to this pore (Osman and Cramer, 2020; Cramer et al., 2001; Wang et al., 2006). The active site can accommodate up to 9nt of RNA forming a hybrid with the DNA also in the active site in an A-form structure (Figure 1.6).

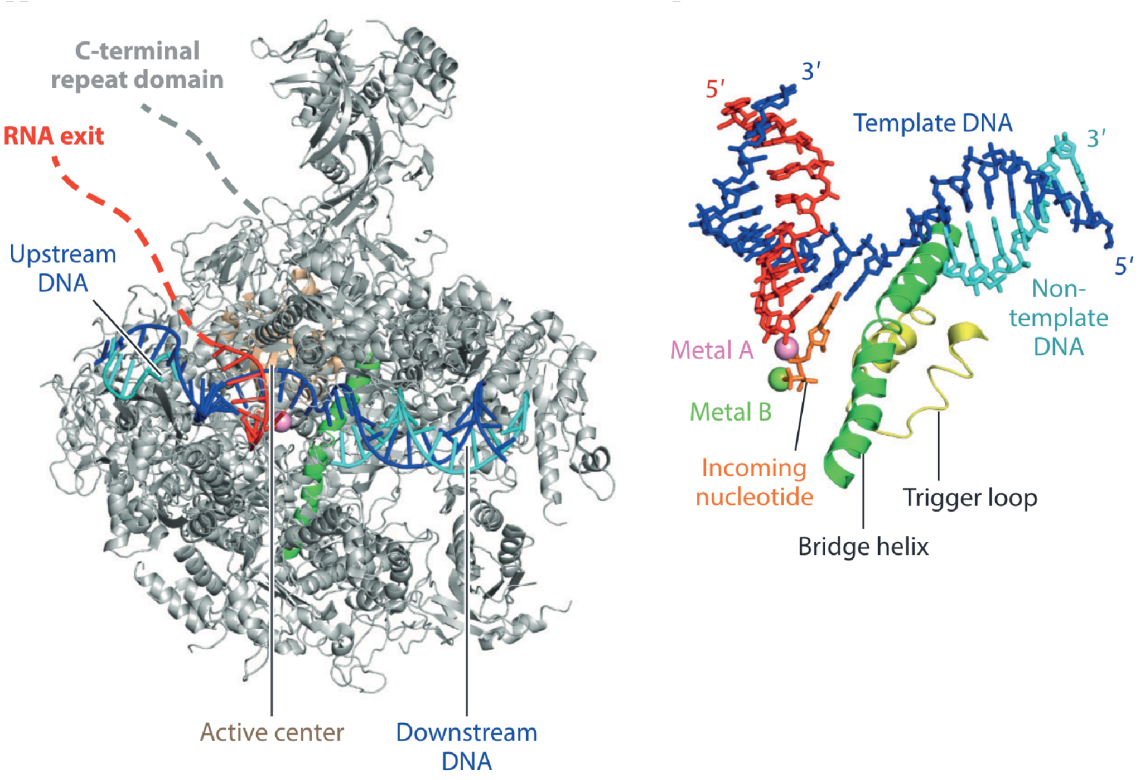


Figure 1.6: **Structure of RNA Polymerase II.** Overall structure of RNAPII showing the key structural domains important for its function including the bridge helix (green), the trigger loop (yellow) and C-terminal domain (grey). The upstream and downstream DNA is also shown (blue) as is the nascent RNA (red). Figure adapted from (Osman and Cramer, 2020)

Another important domain unique to RNAPII is the C-terminal domain of Rpb1, which is a long flexible region often not resolved in structures. The CTD contains multiple repeats (26 for yeast and 52 for humans) of a discrete heptapeptide sequence found in

all eukaryotes consisting of Lysine, Serine, Proline, Threonine, Serine, Proline and Serine (YSPTSPS) (Hsin and Manley, 2012). The CTD is vital for the regulation of RNAPII function and is often targeted throughout the transcription cycle by kinases which can phosphorylate Y1, S2, T4, S5 or S7 (Osman and Cramer, 2020). These phosphorylation marks can be used as recognition signals for transcription cycle specific factors to bind to the polymerase. For example, S5 phosphorylation is often a mark found on RNAPII in the initiation stage and S2 during the elongation phase (Hsin and Manley, 2012; Cho et al., 1997, 2001).

RNAPII transcription can broadly be broken down into three main stages. Firstly, RNAPII will bind to a transcription start site (TSS) during the **Initiation** stage followed shortly by promoter proximal pausing. RNAPII will then transcribe the DNA template synthesising a nascent RNA strand complimentary to the DNA sequence in a predominately unidirectional manner, the **Elongation** stage. Finally, upon reaching a termination region such as Poly-Adenylation (Poly-A) sites the polymerase will undergo **Termination** and release the RNA to be used in downstream reactions (Figure 1.7) (Osman and Cramer, 2020). Throughout all these stages RNAPII will become associated with multiple factors to aid in its function, providing both specificity and regulation.

1.3.2 Initiation

Promoter binding and PIC formation

RNA synthesis must be tightly controlled to accurately and efficiently convert our genome into RNA. In order to achieve this, DNA contains promoter regions where transcription initiation begins. Unlike in bacterial systems, eukaryotic promoters do not contain a consensus DNA sequence that triggers the initiation, but often contain one of the many

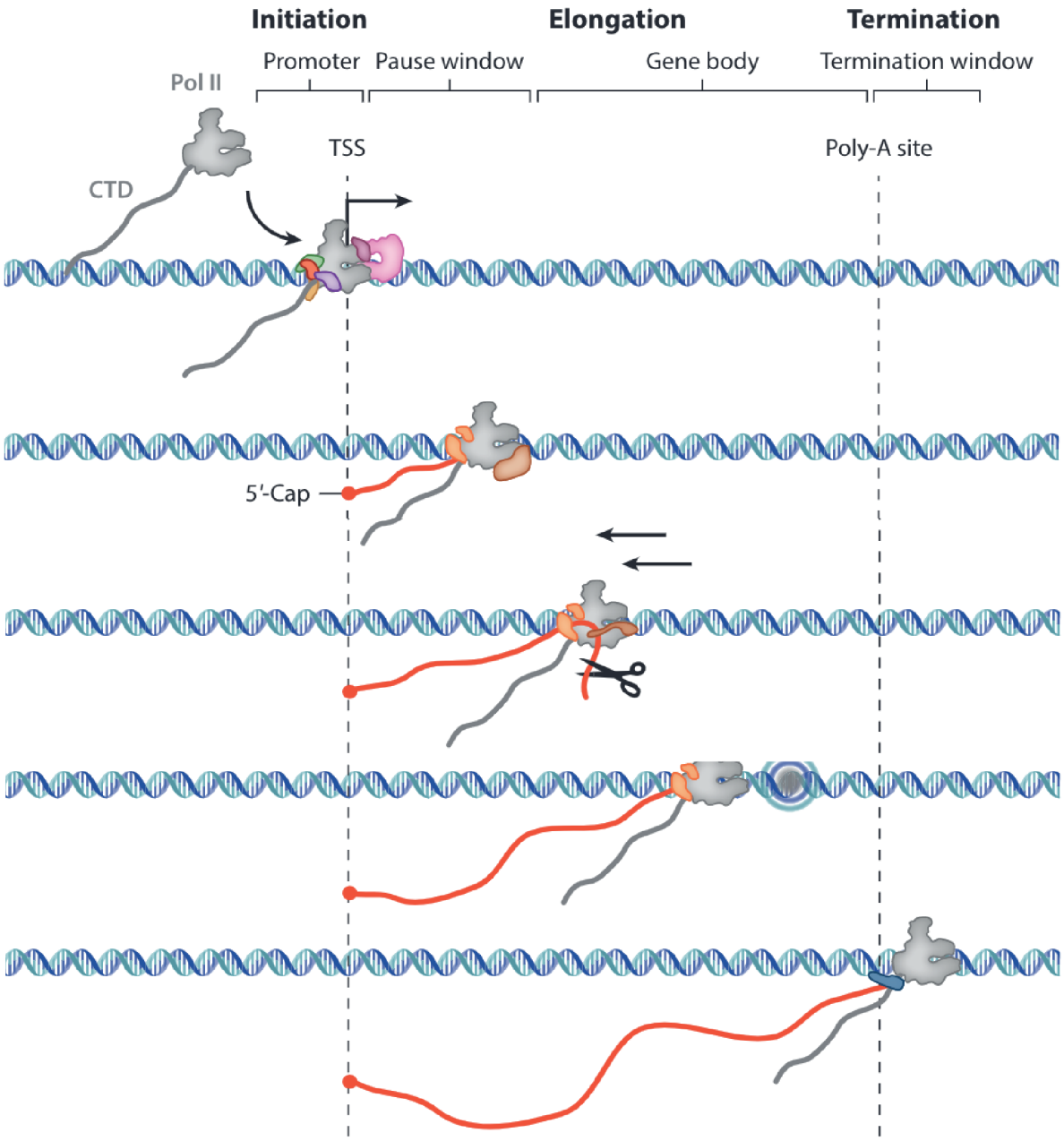


Figure 1.7: **RNAPII transcription cycle.** Transcription cycle of RNAPII moving through the key three stages. 1. Initiation - where GTFs bind and assemble the initiation complex. 2. Elongation - where RNAPII transcribes the template often backtracking or pausing. 3. Termination - RNAPII terminates at a poly-A signal. RNA shown in red, DNA in blue. Figure adapted from (Osman and Cramer, 2020)

motifs found across the promoter region (Kadonaga, 2012). One of the most famous eukaryotic promoter motifs is the TATA-box, discovered in 1979 it is found anywhere between 30 -120 base pairs upstream of the TSS (Smale and Kadonaga, 2003). Despite it being so well known, it is only found in approximately 10-20% of human promoters (Osman and Cramer, 2020). Instead, our promoters contain multiple other elements such as the

initiator element, downstream promoter element (DPE) and TFIIB-recognition elements (BRE). Like all multi subunit polymerases, RNAPII cannot bind to these promoter regions by itself but instead requires binding of general transcription factors (GTFs) first, which then act as a nucleation site for the formation of the pre-initiation complex (PIC) (Nogales et al., 2017).

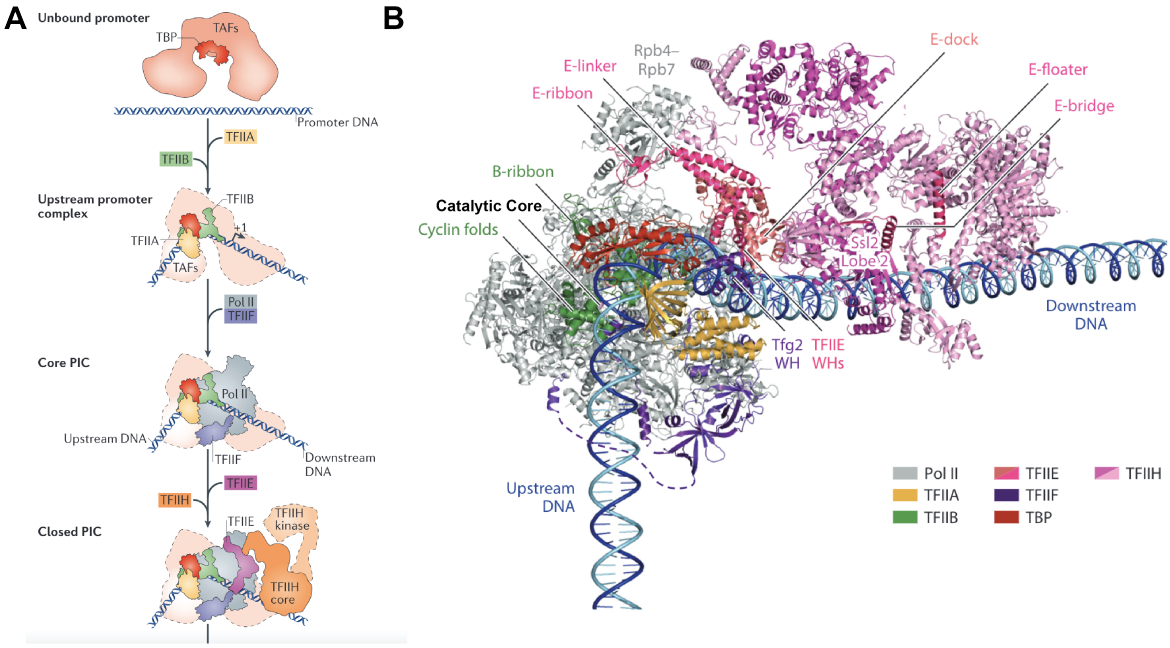


Figure 1.8: **Initiation of RNAPII transcription.** (A) Assembly pathway of GTFs in the canonical model. GTFs (various colours) bind to DNA in a concerted manner allowing binding of RNAPII (grey) forming the cPIC. (B) Structure of the yeast PIC with RNAPII bound to the template DNA along with the GTFs bound to upstream DNA (TFIID, TFIIB), RNAPII (TFIIF, TFIIE) and downstream DNA (TFIIH). PDB 5OQJ. Figures adapted from (A) (Sainsbury et al., 2015) and (B) (Osman and Cramer, 2020)

There are 5 main GTFs important for the initiation stage, TFIIB, TFIID, TFIIF, TFIIE and TFIIH all of which bind in a concerted manner. The process of GTF binding has been studied extensively over the past 20 years, with a canonical model being generated (Figure 1.8A). In this canonical model, a TATA-binding protein (TBP) is first loaded onto the TATA-box inducing a severe bend in the DNA (90°), this in turn is then stabilised by TFIIB (Sainsbury et al., 2015). TFIIB binds to TBP spanning the promoter region interacting with the downstream BRE. TFIID is also important in this step as it contains

a TBP subunit, although TFIID does not traditionally bind TATA-box elements it has been shown to bind to DPEs inducing the same bend that TBP alone does (Osman and Cramer, 2020). Once TBP and TFIIB have bound, TFIIF is next to arrive. TFIIF is often thought to associate with RNAPII and act as bridge between DNA bound TFIIB and free RNAPII, bringing RNAPII into the initiation complex on the DNA. However, recent single-molecule and biochemical studies have shown that RNAPII can bind to the promoter DNA by itself and that only 50% of yeast RNAPII is associated with TFIIF (Baek et al., 2021). Instead, it is now proposed that TFIIF acts as a regulator ensuring RNAPII binds specifically to the promoter region and not non-specifically along the DNA (Sainsbury et al., 2015). The final step in the formation of the PIC is binding of TFIIE and TFIIH. TFIIE has been shown to bind to one side of the RNAPII clamp, across from TFIIF on the other side (Chen et al., 2010; Eichner et al., 2010). TFIIE will make extensive contacts with RNAPII via its WH-domains stabilising the association, as well as interacting with TFIIF and TFIIH (He et al., 2013; Chen et al., 2007; Allepuz-fuster et al., 2019).TFIIH follows and binds on one end to TFIIE and on the other to downstream DNA, approximately 10 base pairs away (He et al., 2013) (Figure 1.8A). Although TFIIH does not have an assembly role in bringing RNAPII to the initiation site, it does have a very important function in opening of promoter DNA (Kim et al., 2000; Sainsbury et al., 2015). RNAPII associated with these 5 GTFs now forms the closed PIC (cPIC) which is a mega-dalton sized complex and upon TFIIH activity will transition from the cPIC to an open complex capable of transcribing DNA (Osman and Cramer, 2020) (Figure 1.8B).

Enhancers and the Mediator complex

Although promoter elements are essential binding sites for RNAPII to activate transcription, there are other regulatory regions of DNA, called enhancers, that are essential in

regulating gene expression and preventing RNAPII from transcribing the wrong gene (Serfling et al., 1985). Enhancers are very variable in their size and function, where they can range from 10s of base pairs up to 10s of kilobase pairs (Hnisz et al., 2013; Parker et al., 2013) (Figure 1.9). What is most surprising about these enhancers is that they can often be found many kilobases away from the gene which they activate, raising multiple questions which is the target of many sequencing and *in vivo* studies (Field and Adelman, 2020). One hypothesis is the idea of DNA looping, and chromatin is often found heavily compacted in 3D space making multiple contacts to regions that are not close in a genomic 1D sense but are close topologically, topologically associated domains (TADs) (Andrey and Mundlos, 2017; Schoenfelder and Fraser, 2019). The formation of these TADs can be driven by multiple transient factors including CTCF and cohesin which will physically bring enhancers close to the gene promoter allowing more direct control over the desired gene (Field and Adelman, 2020).

Another factor that can drive this looping is the Mediator complex, a multi-subunit mega-dalton complex that has been shown to bind both TFs and RNAPII (Osman and Cramer, 2020). The interaction between Mediator and RNAPII has been shown extensively in structural studies but *in vivo* assays suggest that this interaction is more dynamic and transient (Schilbach et al., 2017; Tsai et al., 2017). Mediator makes multiple contacts to RNAPII including at the stalk and jaw domains (Figure 1.4) (Schilbach et al., 2017). As one Mediator complex can make several contacts with multiple RNAPII and TF molecules it would allow for bridging between regions of DNA that are far away genomically but close together in 3D topological space (Figure 1.9).

Recently, super-resolution *in vivo* imaging has proposed the idea of transcriptional hubs which help link RNAPII to initiation sites. These hubs or clusters often contain multiple

RNAPII molecules tethered together and associate at gene promoters to more efficiently link RNAPII to the gene (Kimura and Sato, 2022). This has been seen in multiple instances and it is now known that these transcriptional hubs are linked together by other factors such as myosin VI to tether multiple RNAPII molecules into one hub (Hari-Gupta et al., 2022).

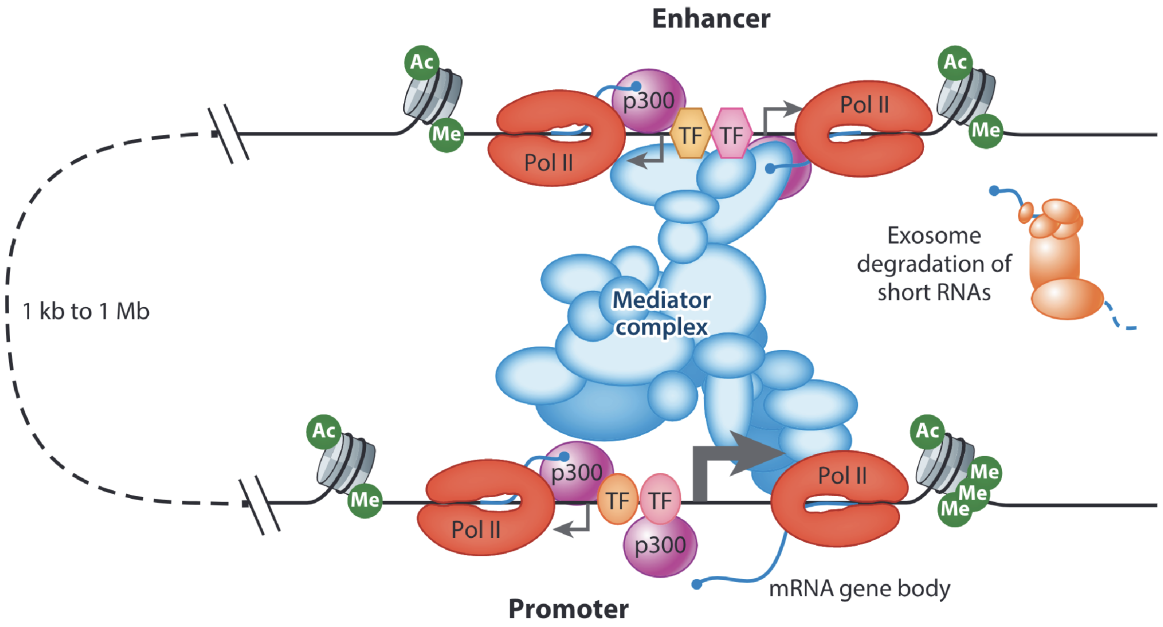


Figure 1.9: **Role of Mediator complex at promoters and enhancers.** Relationship between promoter regions and their regulatory enhancers located up to 1Mb away from each other in genomic location. Mediator complex (blue) shown can physically link these regions together in 3D space coupling their functions allowing RNAPII (red) activation at TSS (grey arrow). Other features of enhancers such as nucleosomes and TF also shown. Figure adapted from (Field and Adelman, 2020)

DNA melting and opening

TFIIH contains both ATPase (XPB and XPD) and kinase (CDK7) modules both of which are important for the function of DNA opening (Grünberg et al., 2012). Due to the size and complexity of TFIIH it took many years to get high resolution structures capable of providing detail on TFIIH contacts within the PIC (Schilbach et al., 2017; Greber et al., 2019). New studies have now shed light on this interaction, which shows that while the XPB translocase module binds to the downstream DNA region, the XPD translocase

is located far away from this region, backing up older biochemical data that XPB and not XPD is vital for the DNA melting (Tirode et al., 1999; Greber et al., 2019). This process is ATP dependent and upon ATP hydrolysis via TFIIH, RNAPII will attempt to translocate along the DNA. However as it is bound stably to the PIC via TFIIE, rather than translocating along DNA it reels DNA towards the PIC and threads the DNA through the RNAPII active site (Fishburn et al., 2015; Osman and Cramer, 2020). Although the kinase module does not directly induce DNA melting, it is responsible for phosphorylation of the CTD which as discussed above is vital for regulating RNAPII and inducing escape from the promoter region (Svejstrup et al., 1996).

Abortive initiation

However, before RNAPII can escape the promoter region, it will often undergo several rounds of 'abortive initiation'. Abortive initiation occurs when RNAPII synthesises a small amount of RNA (6-11 nucleotides) before pausing, releasing the RNA and reverting back to the original state (Duchi et al., 2016). smFRET based assays showed that in bacteria this abortive initiation occurs via a 'DNA scrunching' mechanism (Kapanidis et al., 2006). In this model RNAP is bound to the promoter and cannot move along DNA, so instead reels DNA into its active site. As the DNA and RNA builds up in the active site it causes increased stress on the polymerase which can lead to one of two outcomes. Either the RNA is released leading to abortive initiation or the upstream DNA is pushed out of the active site leading to productive initiation and promoter escape (Kapanidis et al., 2006). It is hypothesised that initial transcription via eukaryotic RNAPII is carried out by a similar mechanism, but with some distinct changes. In the case for eukaryotic RNAPII, rather than actively 'scrunching' the DNA it is thought that TFIIH acts as a 'molecular winch' reeling in downstream DNA which in turn builds up stress causing the

same outcome as in bacteria (Shino and Takada, 2021).

1.3.3 Elongation

Once RNAPII has escaped the promoter it can now enter active elongation and transcribe template DNA synthesising nascent RNA. RNAPII will undergo several key structural changes when transitioning from the PIC to the elongation complex (EC) which permits transcription (Osman and Cramer, 2020). Firstly, the GTFs are released from RNAPII as their initiation roles are no longer required. Interestingly, TFIIF and TFIIH do have functions within the elongation pathway and can associate with RNAPII during pausing and repair (Svejstrup et al., 1996; Saponaro et al., 2014). Secondly,the DNA around the EC also undergoes structural changes as the unwound downstream DNA enters the active site of RNAPII. The DNA duplex is unwound via the bridge helix, with the template strand (TS) being directed into the active site while the non-template strand (NTS) is passed out of RNAPII (Osman and Cramer, 2020; Bernecky et al., 2016) (Figure 1.10). The two strands will then re-anneal upstream of RNAPII reforming the duplex, exiting at right angles from the entering DNA. NTPs are then added into the active site, synthesising the RNA chain with the 3' stably associated to the active site forming the transcription bubble. Once the RNA chain reaches a critical length (9 nucleotides) the 5' end of the RNA is separated from the DNA via the RNAPII lid loop causing the 5' RNA to exit via the RNAPII exit channel (Figure 1.10). The addition of these NTPs into the growing RNA chain is tightly regulated and occurs via a conserved NTP addition cycle (NAC) (Brueckner et al., 2009).

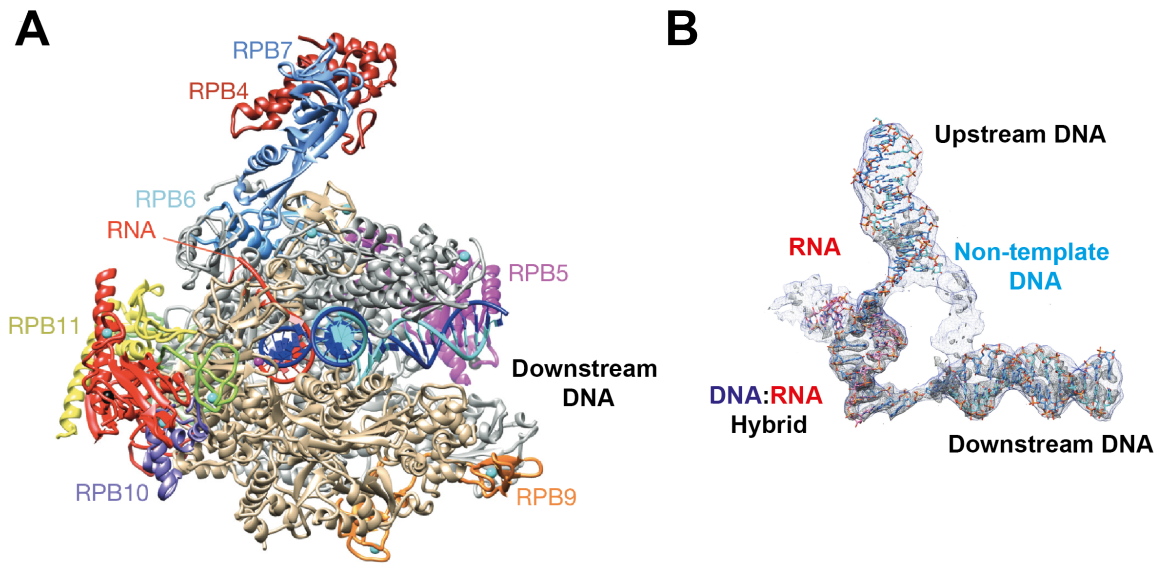


Figure 1.10: **Structure of RNAPII elongation complex.** (A) Cryo-EM structure of mammalian RNAPII EC showing location of RNAPII subunits (various colours) and the bound nucleic acids (Blue for DNA, Red for RNA). (B) EM density for downstream, upstream and DNA:RNA hybrid within the EC structure showing a 9nt DNA:RNA hybrid within the active site. Figure adapted from (Bernecky et al., 2016)

NTP addition cycle

In general, RNAPII selects the cognate NTP based on the TS sequence and adds this NTP in between the 3' end of the growing RNA chain and the RNAPII bridge helix and trigger loop (Osman and Cramer, 2020). The NTP is added via a phosphodiester bond formation causing the release of pyrophosphate and will translocate 1 nucleotide downstream to the next TS base freeing up the NTP binding site for the next NTP to bind. Based on bacterial and eukaryotic work it is now known that this mechanism of NTP addition is conserved in all multi-subunit polymerases (Vassylyev et al., 2007; Murakami and Darst, 2003; Osman and Cramer, 2020).

An NTP will first bind randomly and transiently to the open active site where the trigger loop is mobile forming an inactive pre-insertion state. If this NTP is the cognate base then the trigger loop will fold, closing the active site and move the NTP into the insertion site (Osman and Cramer, 2020). Binding of the non-cognate NTP causes the trigger loop to

remain transitioning between the open and closed state and will not effectively move the NTP into the insertion site (Kireeva et al., 2008; Mazumder et al., 2020) (Figure 1.11).

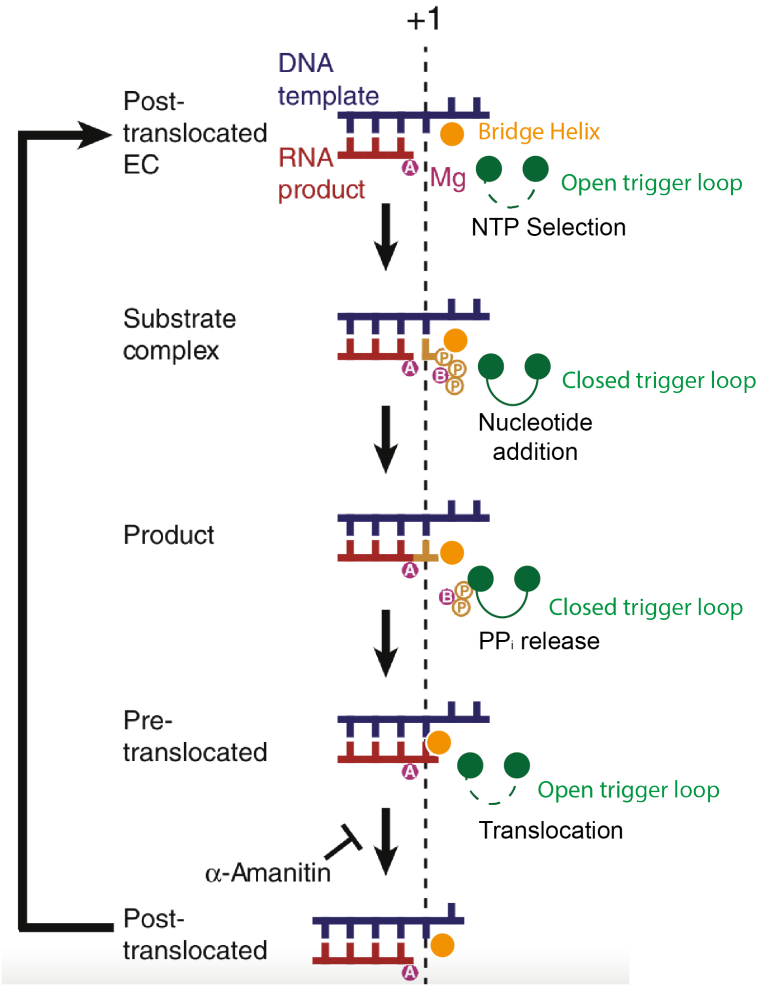


Figure 1.11: **Nucleotide Addition Cycle.** (Schematic of the steps involved in the NAC showing key components such as the bridge helix (orange circle) and trigger loop (green circle). Vertical dashed line represented the +1 position where NTP catalysis occurs. Figure adapted from (Martinez-Rucobo and Cramer, 2013) and (Brueckner et al., 2007)

Now in this position, the NTP is able to undergo phosphodiester bond formation which is carried out via a two metal ion mechanism (Sosunov et al., 2003). The RNAPII active centre contains two metal ions (usually Mg^{2+}), metal A and metal B. Metal A is permanently bound and contacts the RNA 3'OH group as well as Rpb1 aspartate loop. Metal B is far more mobile and will often enter with the incoming NTP forming multiple contacts. Once in place, the 3' RNA OH group attacks the γ-phosphate group of the incoming NTP

via a nucleophillic substitution mechanism (Brueckner et al., 2009) (Figure 1.11). After catalysis, the pyrophosphate group is released inducing the trigger loop to revert back to the initial open conformation and the active site is then ready to bind the next incoming cognate NTP restarting the NAC (Martinez-Rucobo and Cramer, 2013) (Figure 1.11). Within the nucleus, there is a high concentration of both NTPs and dNTPs but RNAPII is able to discriminate between them via contacts with the 2' OH moiety on the ribose sugar, showing a 5000-fold selectivity for NTPs (Cheung et al., 2011).

Once the NTP is added, RNAPII will translocate along the DNA by 1 base pair and repeat this NAC. The mechanism of translocation was heavily debated in the past decade with different models being proposed, the 'power-stroke' model and the 'Brownian ratchet' model (Martinez-Rucobo and Cramer, 2013). In the 'power-stroke' model, it was hypothesised that translocation occurred due to conformational changes in the polymerase that drove RNAPII forward, coupled to the hydrolysis of the cognate NTP. The evidence for this was mainly based on structural data from single-subunit polymerases (Yin and Steitz, 2004) but more recent structural and single-molecule evidence proved that this model was unlikely to account for RNAPII translocation (Kireeva et al., 2010; Abbondanzieri et al., 2005; Osman and Cramer, 2020). Instead, the Brownian ratchet model has been favoured where thermal fluctuations between a pre- and post-translocated state of RNAPII are driven by other steps that release energy (such as NTP binding and catalysis) which in turn cause a net movement forward by RNAPII (Martinez-Rucobo and Cramer, 2013). It is thought that in this model, the trigger loop and bridge helix form key components of this Brownian ratchet where the EC transitions in equilibrium between the pre- and post-translocated state (Kireeva et al., 2010). This model was first elucidated by incubating RNAPII with a toxin,α-amanitin, that has been shown to inhibit RNAPII elongation

(Brueckner et al., 2009). These structures showed that α-amanitin made several key contacts with RNAPII, in particular with the trigger loop trapping it in an intermediate state between the pre- and post-translocated state and was also backed up by kinetic data (Gong et al., 2004). In this state, the trigger loop is wedged between the bridge helix blocking the active site preventing movement into the post-translocated state. Based on this structural and kinetic data it was hypothesised that RNAPII translocates via a two-step Brownian ratchet mechanism (Martinez-Rucobo and Cramer, 2013).

Pausing and backtracking

During transcription RNAPII will often undergo transient pausing varying in duration depending on the cause of this pause. DNA roadblocks such as DNA damage, nucleosomes or certain DNA sequences have all been shown to have an affect on RNAPII (discussed more in section 1.5 below) (Osman and Cramer, 2020). One characteristic of RNAPII in a paused state is the fraying of the 3' RNA chain within the active site. Once frayed, this can lead to interaction with the Rpb2 subunit effectively blocking any further NTPs to be inserted into the active site (Landick, 2009). Another characteristic of paused RNAPII is a tilt in the DNA:RNA hybrid within the active site also preventing NTP insertion (Vos et al., 2018a). Paused RNAPII will often correct itself as in the case of a incorrectly incorporated NTPs. RNAPII has an intrinsic proofreading and endonuclease capability and can detect structural defects in the DNA:RNA hybrid when a mismatch occurs (Sydow et al., 2009; Cheung and Cramer, 2011). In this scenario, the 3' RNA frays from the DNA hybrid inducing the paused state as discussed above (Sydow et al., 2009). RNAPII will then backtrack 1 nucleotide moving the frayed end into a proofreading site, aligning the phosphodiester bond within the catalytic site. RNAPII will cleave a dinucleotide containing the mismatched base and return the new RNA chain back into

the active site, allowing the next NTP to be readily inserted before continuing active elongation (Martinez-Rucobo and Cramer, 2013).

However, if the pause cannot be rectified by RNAPII alone then it will backtrack causing larger rearrangements of the DNA:RNA hybrid (Gregersen and Svejstrup, 2018). In this instance, the 3' RNA is again disengaged from the DNA hybrid but now extruded through a pore on the underside of the active site. In most cases the RNA will thread only a few nucleotides through the pore before other factors will bind to rectify the blockage, but it has been shown that RNAPII can backtrack up to 10 nucleotides which can result in transcriptional arrest (Osman and Cramer, 2020). In these situations, RNAPII cannot usually recover by itself and instead requires other factors to bind such as TFIIS (Sigurdsson et al., 2010). TFIIS is a TF that contains three domains, two of which (domain II and III) are essential for RNAPII binding and RNA cleavage activity (Martinez-Rucobo and Cramer, 2013). Structural studies have elucidated the mechanism as to how TFIIS acts on RNAPII, showing large interactions with the RNAPII pore where backtracked RNA is located (Kettenberger et al., 2004). An acidic loop in domain II inserts itself through the pore into the active site and will stimulate RNAPII catalytic activity inducing the cleavage of RNA chain also via a two metal ion mechanism. TFIIS will then release itself from RNAPII and in turn allow the now free RNA to exit the pore and dissociate from RNAPII, allowing transcription to be resumed (Cheung and Cramer, 2011).

Not all pauses are induced by a blockage in the transcription pathway, in fact recent studies have shown that many RNAPII molecules will undergo a process called promoter-proximal pausing. DNA sequencing and structural studies have helped to elucidate this phenomenon which shows RNAPII pausing often occurs at a site 50-80 base pairs downstream of the TSS (Core and Adelman, 2019; Core et al., 2008). The cause of this pausing

could be due to many factors including DNA sequence, transcription factors (TFs) or even the chromatin state (Core and Adelman, 2019). Once in this paused state, RNAPII will associate with regulatory TFs such as DRB-sensitivity-inducing factor (DSIF) and negative elongation factor (NELF) which stabilise this state (Osman and Cramer, 2020). Although it may seem counter-intuitive RNAPII pausing is an essential regulatory process of transcription and allows for any errors to be detected or for the chromatin state to be re-arranged before entering active elongation (Core and Adelman, 2019). Phosphorylation of these factors via positive transcription elongation factor b (pTEFb) will cause NELF to release from the complex and RNAPII will continue active elongation (Adelman and Lis, 2012).

Regulation of transcription elongation

RNAPII transcribes at an average rate of between 1-4 kb/min synthesising protein coding genes (Maiuri et al., 2011). To achieve this fast and accurate transcription RNAPII is tightly regulated both by modification of RNAPII itself and the binding of external factors.

RNAPII associates with a plethora of TFs that control RNAPII function depending on what stage of transcription RNAPII is in. As mentioned above, the TFs DSIF and NELF associate with paused RNAPII and will stabilise this state (Figure 1.12). In fact, structural studies have shown that NELF blocks the RNAPII pore preventing TFIIS to associate with promoter-proximal paused RNAPII and therefore preventing RNA cleavage (Vos et al., 2018b). p-TEFb is a vital TF that controls RNAPII processivity as it not only shifts paused RNAPII into active elongation by dissociation of NELF but it can also control the phosphorylation state of DSIF. Once phosphorylated, it will associate and track with elongating RNAPII often until the end of a gene (Vos et al., 2018a).

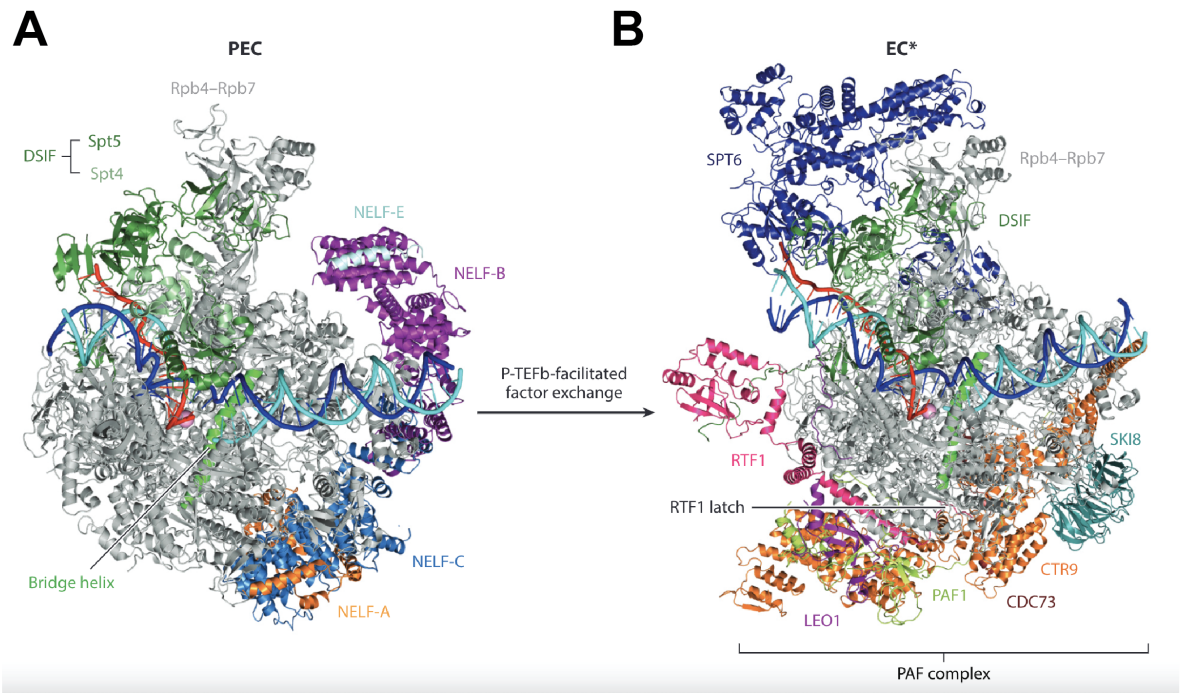


Figure 1.12: **Structure of RNAPII EC bound to elongation factors.** Structures of RNAPII (grey) with the elongation factors DSIF (green) and NELF (blue, orange and magenta) (A) and SPT6 (blue) and PAF complex (multiple colours) (B) showing their interaction with RNAPII during pausing and active elongation. PDB: 6GML (A) and 6GMH (B). Figure adapted from (Osman and Cramer, 2020)

While some TFs such as p-TEFb and TFIIS act directly to induce RNAPII function at the active site, other TFs bind and control RNAPII allosterically. SPT6 and the polymerase associated factor (PAF) complex have been shown to associate with RNAPII during active elongation and alter its function, but are not actively involved in the polymerisation process (Vos et al., 2018a, 2020) (Figure 1.12). Instead it is thought that these factors bind to RNAPII and induce structural changes which in turn will alter RNAPII function. SPT6 for example binds to the phosphorylated CTD of RNAPII and loosens the DSIF binding to RNA preventing the paused state. PAF contains at least 5 subunits and also contacts the outer surface of Rpb2, with one of its subunits (RTF1) also contacting the bridge helix of the active site inducing structural changes (Xu et al., 2017b; Yu et al., 2015) (Figure 1.12). Other factors such as ELL, elongin and the super elongation complex also bind allosterically to RNAPII and will regulate its function (Zhou et al., 2012).

1.3.4 Termination

Overview

The termination phase of RNAPII transcription is possibly the least understood stage of the transcription cycle, partly due to the complex nature of this stage (Eaton and West, 2020). RNAPII transcribes a variety of gene types and therefore the mechanism of termination is not the same throughout. Despite this complexity, the best understood pathway of termination is the polyadenylation signalling (PAS) pathway (Proudfoot, 2016).

In this pathway the DNA contains a specific motif (AAUAAA) flanked by a U-rich region upstream and GU-rich regions downstream, which signals the 3' end of a gene and directs RNAPII termination (Proudfoot, 2016). RNAPII will transcribe beyond this region but in doing so the cleavage and polyadenylation (CPA) complex is associated to the gene. The CPA complex contains the cleavage and polyadenylation specificity factor (CPSF), a cleavage stimulatory factor and cleavage factors I and II (Eaton and West, 2020). The CPA complex recognises the PAS motif via its CPSF30 subunit while the WDR33 endonuclease will cleave at a specific site after the AAUAAA motif (Eaton and West, 2020). This would then allow release of the mature RNA which can be targeted for downstream processes.

However, even within the PAS pathway there are conflicting models that have emerged and only recently has a unified model been hypothesised, these will be discussed below (Logan and al., 1987; Connelly and Manley, 1988) (Figure 1.13). It should be noted however that other termination models do exist such as in the case of intergenic transcription and RNAPIII termination, which utilises RNA hairpins, a model that is more akin to bacterial transcription (Ray-Soni and al., 2016; Arimbasseri and Maraia, 2015).

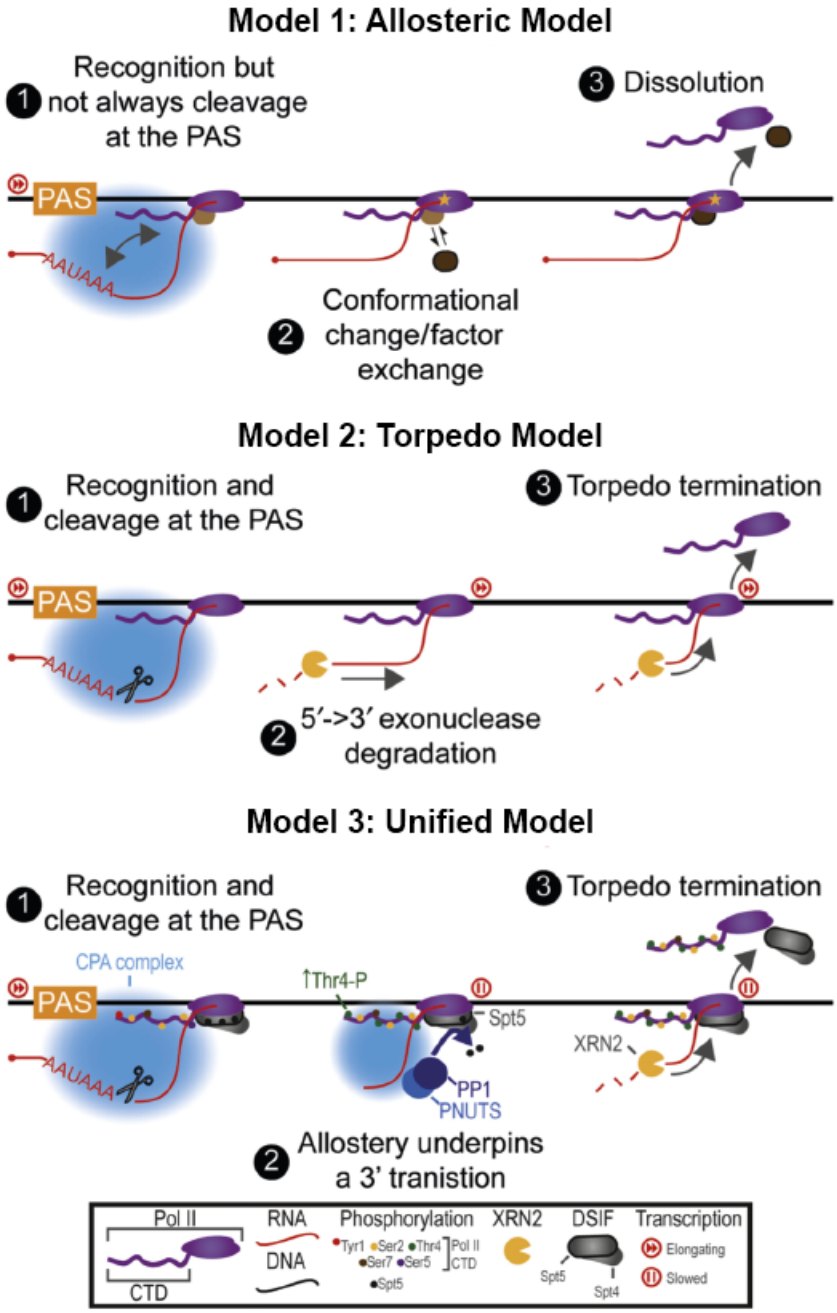


Figure 1.13: **Models of RNAPII transcription termination.** The three different models of transcription termination proposed for RNAPII. 1. Allosteric model where conformational changes in RNAPII drive termination. 2. Torpedo model where factors such as Xrn2 exonuclease drive RNAPII dissociation. 3. Unified model where both models 1 and 2 are combined and work in concert to terminate transcription. Key at bottom shows key factors in this process. Figure adapted from (Eaton and West, 2020)

Allosteric Model

The allosteric model was first hypothesised in 1987 and suggests that as RNAPII transcribes past the PAS region, key elongation factors will dissociate from RNAPII, inducing

a structural change within the polymerase allowing dissociation of RNAPII from the gene (Logan and al., 1987) (Figure 1.13). Anti-termination factors such as SCAF4 and SCAF8 are hypothesised to inhibit termination until the PAS is reached (Gregersen and al., 2019). Other factors such as CDK12 and PABPN1 (a poly(A) binding protein) have also been implicated in dissociation from RNAPII at PAS sites (Dubbury and al., 2018; Jenal and al., 2012). There has also been evidence for conformational changes within RNAPII that drive termination without cleavage of the PAS within the RNA. However, this evidence has been show to have caveats and it is unlikely this model would solely account for termination within the cell (Osheim and al., 2002; Zhang and al., 2015).

Torpedo Model

The torpedo model, hypothesised a year later in 1988, consists of a 5'-3' exonuclease which degrades the PAS cleaved RNA product still bound to RNAPII and will translocate along the RNA until reaching the RNAPII molecule (Connelly and Manley, 1988) (Figure 1.13). This will in turn cause dissociation of RNAPII as the exonuclease disrupts its binding, acting as a 'torpedo' (Eaton and West, 2020). In this scenario, PAS cleavage must precede exonuclease binding as to prevent degradation of the gene product, and this has been shown to be true in previous studies (Nojima and al., 2018; Eaton and al., 2020). One of the major exonucleases responsible for this in humans is Xrn2 and Rat2 in yeast (Eaton and al., 2018; West et al., 2004; Kim and al., 2004). However, although they have been implicated in this pathway, termination has been shown to continue without their presence and so cannot fully account for RNAPII termination either (Eaton and West, 2020).

Unified Model

Since neither of these models can account for efficient termination by themselves, a unified model has been proposed where the two work in concert together (Eaton and al., 2020) (Figure 1.13). In this unified model, it is proposed that the allosteric effect starts this process. Recent evidence in yeast has showed that RNAPII decelerates after the PAS region, possibly due to the protein phosphatase 1 enzyme (PP1) (Parua and al., 2018; Kecman and al., 2018). PP1 is part of the human CPA complex and targets two components of the RNAPII EC which act as allosteric signals (Shi and al., 2009). The first of these is de-phosphorylation of SPT5 (a major subunit in DSIF), which as discussed earlier is an essential factor in converting RNAPII from its initiation state into the EC. It may have a similar role here where upon its de-phosphorylation, RNAPII undergoes a conformational change in turn slowing down transcription after the PAS. A second target of PP1 is Tyr1, part of the CTD, which in turn signals for the recruitment of the CPA complex (Schreieck and al., 2014). Furthermore, other allosteric factors such as R-loops have been proposed to slow down RNAPII at the end of genes (Zatreanu et al., 2019; Crossley and al., 2019). More evidence of this was shown a year later, where RNAPII elongation speeds dropped dramatically after the PAS, from $\sim$ 2 kb/min to almost $\sim$ 0.1 kb/min (Cortazar and al., 2019).

For the torpedo model to be valid, it has to at least match the velocity of RNAPII after the PAS region, otherwise it will never catch up to the polymerase (Eaton and West, 2020). Although XRN2 velocities are still unknown, a similar exonuclease has been shown to translocate at speeds similar to RNAPII ($\sim$ 2.3kb/min) (Hoek and al., 2019). It is therefore hypothesised that the deceleration achieved by allosteric methods would allow for Xrn2 to easily catch RNAPII which is effectively stalled after the PAS and

coupling of the two models together. In this unified model of PAS-dependent termination, PAS cleavage by CPA would promote de-phosphorylation of DSIF which in turn slows down RNAPII to allow Xrn2 to bind PAS cleaved RNA. Xrn2 is now able to catch up to RNAPII and act as a torpedo dissociating it from the gene, leading to efficient termination of transcription (Eaton and West, 2020).

1.4 Roadblocks to Transcription

1.4.1 Overview

Throughout transcription RNAPII will encounter several roadblocks such as nucleosomes or DNA lesions that must be navigated to continue processive and accurate transcription. Obstacles along the DNA pathway within the template strand can lead to premature termination of a gene which in turn can cause several severe genetic disorders and cancers (Lans et al., 2019; Noe Gonzalez et al., 2021). Many of these roadblocks are due to intrinsic sources (e.g nucleosomes) but DNA lesions can occur due to a variety of external sources such as chemotherapeutics, ionising radiation and UV-light (Lans et al., 2019). Depending on their severity, these obstacles have been shown to cause detrimental effects on RNAPII such as pausing, backtracking or even transcriptional arrest (Figure 1.14).

Eukaryotic DNA is heavily coated in nucleosomes and histone proteins have also been shown to be present in archaea (Teves et al., 2014; Stevens et al., 2020). Nucleosomes consist of a histone octamer (2X H2A-H2B dimers and a H3-H4 tetramer) wrapped around 147 bp of dsDNA (Willhoft et al., 2018) (Figure 1.14A). Nucleosomes are quite large (∼40Å) and passage through them carries a big challenge for RNAPII (Cutter and Hayes, 2015). Not only must it be bypassed, but the nucleosome contains important epigenetic

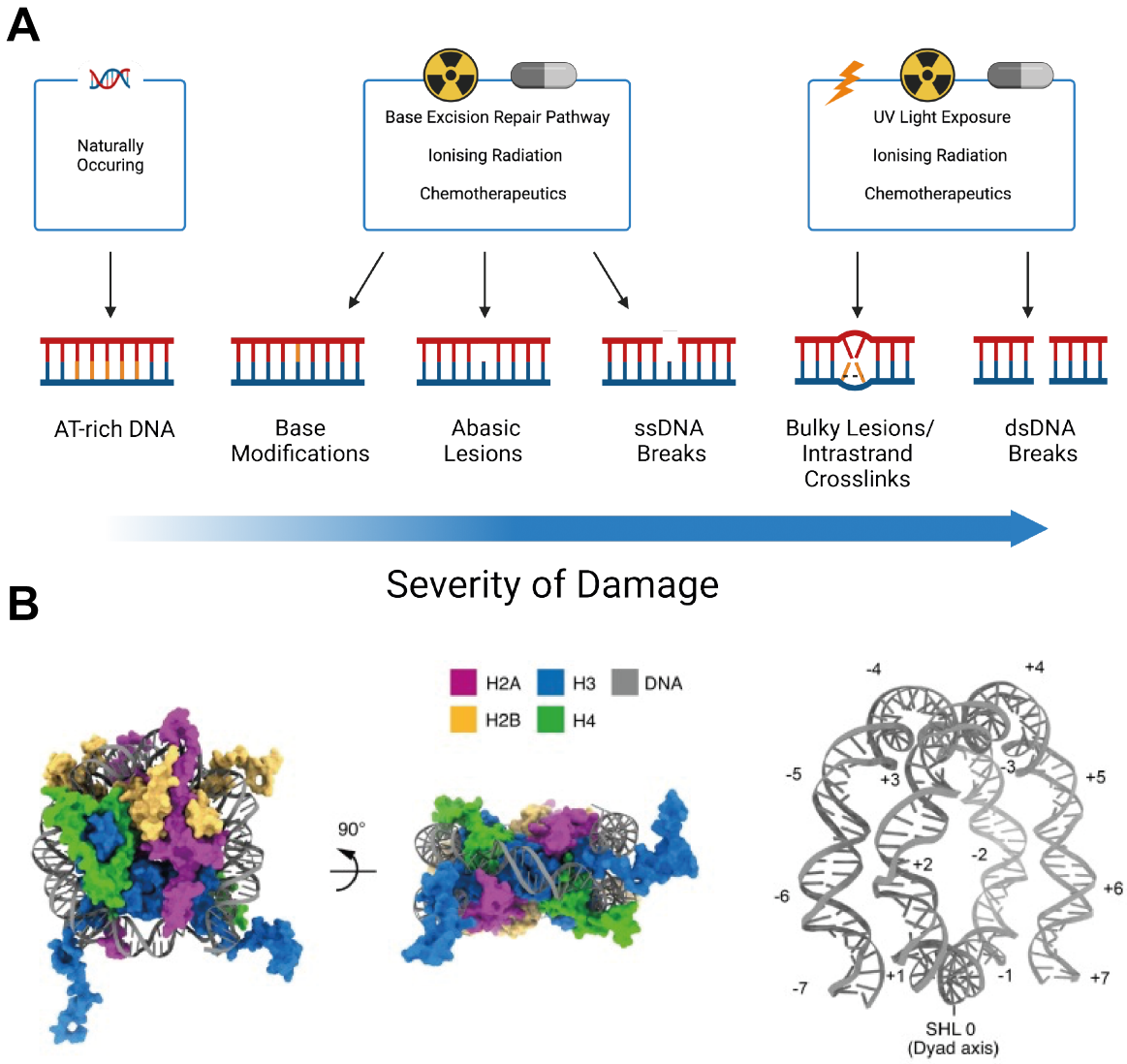


Figure 1.14: **Effect of DNA roadblocks on transcription.** (A) Overview of different DNA lesions found within eukaryotic cells that cause transcriptional stress. Causes of damages shown above and their relative severity to transcription. (B) Nucleosome structure showing core histones (H2A, H2B, H3 and H4, see colour map) (PDB 1KX5) and the (SHLs) within the nucleosome (dyad marked SHL 0). Figure (A) made in Biorender and (B) adapted from (Kobayashi and Kurumizaka, 2019)

information and therefore must be preserved after bypass (Chen et al., 2017). Recent structural studies have shown that RNAPII pauses at 4 main sites within the nucleosome. Superhelical locations (SHL) -6 and -5, which are at the entry site to the nucleosome and SHL -2 and -1 which are close to the nucleosome dyad (centre) (Farnung et al., 2018; Kujirai et al., 2018). In all cases these locations correspond to major sites of DNA-histone contact which pose a barrier to RNAPII forward translocation. Recently, it has been shown that RNAPII displaces the H2A-H2B dimer during passage and can also

backtrack along a nucleosome, aided by TFIIS (Farnung et al., 2022). Bypass of this roadblock involves a multitude of factors including Elf1, FACT and histone remodellers (Ehara et al., 2019; Formosa and Winston, 2020).

The DNA sequence itself has also been shown to be a barrier to transcription inducing transcriptional stress (Xu et al., 2017a). For example, it has been shown that DNA contains regions that are highly rich in either Adenine and Thymine (AT-rich) or Guanine and Cytosine (CpG islands) (Hughes et al., 2020; Rajewska et al., 2012). The cell actually makes use of these regions and AT-rich regions can often be found in promoter sites within the eukaryotic genome to promote DNA opening (Rajewska et al., 2012). Despite their prevalence, AT-rich regions have been shown to induce transcriptional pausing, backtracking and premature termination (Xu et al., 2017a). Although this is likely due to the weak base pairing they exhibit, exactly how they induce RNAPII stalling and premature termination has yet to be defined fully.

DNA lesions have also been shown to be a severe blockage to processive transcription and this is the main focus of this thesis. Much of our understanding has come from biochemical and structural studies, however much of the kinetic mechanism and dynamics of RNAPII at these damages has yet to be fully probed. Described below are three main damages that this thesis will deal with (Figure 1.15).

1.4.2 Abasic lesions

Abasic lesions are one of the most abundant forms of DNA damage found within eukaryotic cells, where up 10,000 lesions can occur per cell per day (Lindahl, 1993). Abasic lesions can occur spontaneously where a base is depurinated or as part of the base-excision repair pathway (BER). During BER, DNA glycosylases recognise oxidized and alkylated

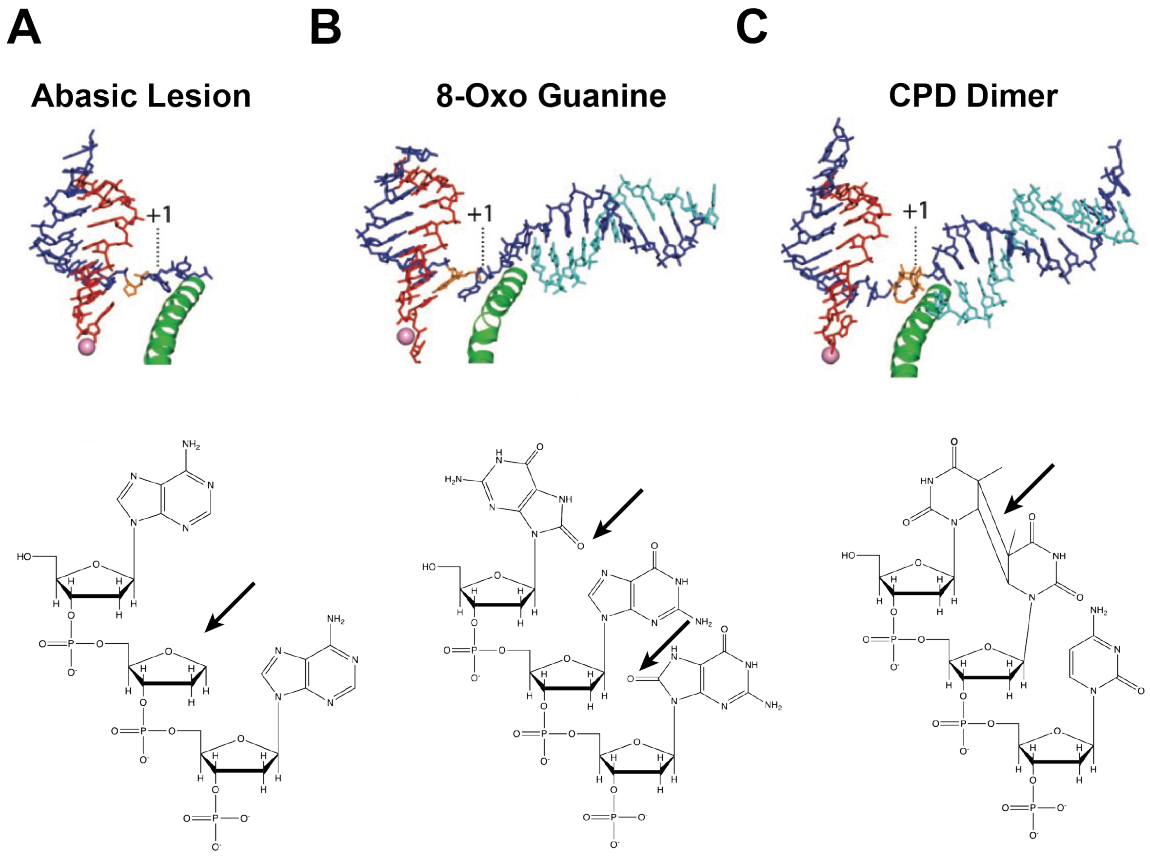


Figure 1.15: **Structures of DNA damage.** (top) Structures from X-ray crystallography or Cryo-EM of RNAPII active sites at varying DNA damages (orange) showing how the bridge helix (green) and +1 site are affected. (bottom) chemical structure of each damage. Black arrow shoes point of damage. (A) Abasic lesion. (B) 8 Oxo-guanine. (C) cyclopyrimidine (CPD) dimer. Figure adapted from (Osman and Cramer, 2020)

DNA and removes the base from the DNA chain to prevent genetic mutation (Pagès et al., 2008) (Figure 1.15A). Abasic lesions have been shown to be mutagenic as their presence within the DNA template can lead to mis-incorporation of non-cognate NTPs and also the formation of DNA-protein crosslinks which can block multiple other cellular processes (Thompson and Cortez, 2020) . It has been shown in multiple biochemical studies that RNAPII follows the A-rule and will readily insert AMP opposite an abasic site (Wang et al., 2018a). A combination of biochemical and structural studies have highlighted the impact abasic lesions have on RNAPII (Wang et al., 2018a). Abasic lesions cause complete blockage of transcription at the site of damage, where RNAPII cannot bypass. Interestingly, there are two stages of stalling at this damage. Firstly, a

fast step where RNAPII stalls rapidly at the abasic lesion followed by a slower extension of one base where RNAPII will insert ATP opposite the abasic lesion (Wang et al., 2018a). A recent Cryo-EM structure of RNAPII at an abasic lesion has shed further insight into this stalling effect. This structure showed RNAPII at both positions, in the first stalling position the abasic site is located above the bridge helix of the RNAPII active site, in turn blocking its critical translocation function (Wang et al., 2018a). The next step of ATP insertion showed that neither the ATP nor the template strand were in the natural position that allows efficient NTP incorporation hence why the efficiency is reduced compared to undamaged DNA (Wang et al., 2018a). Furthermore, upon insertion of ATP opposite the abasic lesion both the 3' RNA and the template strand showed an increased flexibility due to their inability to base pair with each other which is hypothesised to also block the RNAPII from moving forwards beyond the damage.

In order to fix this damage, RNAPII has been shown to backtrack by 1-2 nucleotides where the 3' base will be cleaved via the action of TFIIS (Wang et al., 2018a). As mentioned earlier, TFIIS is a TF that actively works on backtracked RNAPII to remove potential genetic mutations (Martinez-Rucobo and Cramer, 2013). Once backtracked, abasic lesions are targeted by the BER pathway where DNA glycosylases target the abasic lesion for excision via apyrimidine (AP) endonucleases such as APE-1 and will be replaced by the new undamaged base permitting transcription to continue (Boiteux and Guillet, 2004).

1.4.3 Oxidative damage

8-Oxo guanine is one of the most common oxidative lesions along the DNA template caused by reactive oxygen species attacking a Guanine base (Swenberg et al., 2011). Once oxidised, 8-Oxo guanine can actually form two conformations, the usual anti- conformation

and a more rare syn- conformation (Damsma and Cramer, 2009) (Figure 5.4A). Although the normal anti- conformation allows insertion of the cognate CMP opposite, the syn- conformation induces mis-incorporation of AMP into the growing RNA (Damsma and Cramer, 2009; Oh et al., 2020). This mis-incorporation is likely due to the rotation of 8-Oxo guanine as it enters the active site and forming Hoogsteen base pairing with AMP leading to a 21,000 fold increase in AMP incorporation (Damsma and Cramer, 2009; Oh et al., 2020). Unlike for abasic lesions, RNAPII has been shown to bypass 8-Oxo guanine lesions in both conformations readily with only a moderate decrease in transcription rate, suggesting minimal pausing at this lesion (Oh et al., 2020). Due to this rapid bypass, mutations within the RNA chain are common causing C to A substitutions in the protein coding gene. In fact, presence of C-A mutations due to oxidative damage like 8-Oxo guanine has been linked to several cancers including lung, pancreas and ovarian adenocarcinomas (Kucab et al., 2019). There is some biochemical evidence that shows RNAPII can backtrack at the site of DNA damage when in the presence of TFIIS but there is no cleavage of the 3' RNA to remove the mis-incorporated AMP (Oh et al., 2020). Furthermore, early structural work on this complex showed minimal changes in the RNAPII active as was the case for the abasic lesion. Both the template strand and 3' RNA occupied the usual sites permitting AMP incorporation and extension (Damsma and Cramer, 2009) (Figure 1.15). Interestingly, 8-Oxo guanine can be oxidised further into other oxidative lesions including 5-guanidinohydantoin (Gh) and spiroiminodihydantoin (R-Sp and S-Sp). Although these are both small DNA lesions they have been shown to have a larger stalling affect on RNAPII and inducing larger structural changes within the RNAPII active site acting more akin to a bulky DNA lesion (Oh et al., 2020).

Although they are readily bypassed by RNAPII, 8-Oxo guanine lesions are targeted by the

same repair pathway as abasic sites, BER. The mechanism of action is similar to above except different glycosylases are used to detect 8-Oxo guanine, OCG1 and MUTYH in eukaryotic cells (Banda et al., 2017). Strikingly, a new role for nucleotide excision repair (NER) proteins has been suggested to aid in the repair of 8-Oxo guanine lesions (Kumar et al., 2022). NER factors DDB2, XPC and XPA have been shown to accumulate at sites of these lesions and promote recruitment of OCG1, highlighting a possible role of both NER and BER in this repair pathway (Kumar et al., 2022).

1.4.4 UV-induced damage

Ultraviolet-light (UV-light) is a common source of DNA damage causing the formation of intra-strand crosslinks (Brueckner et al., 2007). Adjacent pyrimidines (often thymine) become covalently linked to form cyclobutane pyrimidine dimers (CPDs) within the template strand. Bulky lesions like CPDs have been shown to have a more detrimental effect on transcription compared to smaller lesions like abasic sites or 8-Oxo guanine groups (Osman and Cramer, 2020). CPDs formed by UV-light have been heavily linked to multiple diseases and cancers including skin carcinomas and melanomas (Hayward et al., 2017; Nik-Zainal et al., 2015).

RNAPII behaves in a similar manner at CPDs as it does with abasic lesions, where stalling at the site of damage is detected (Brueckner et al., 2007; Xu et al., 2017a). Upon stalling, RNAPII rate of transcription is readily decreased and slow incorporation of UMP opposite the 5' base of the CPD occurs (Brueckner et al., 2007). Once inserted RNAPII becomes stalled again in between the CPD. Unlike the 8-Oxo guanine group almost no bypass of this damage has been seen *in vitro* or *in vivo* (Osman and Cramer, 2020). Although it behaves in a similar manner biochemically to abasic lesions, structural studies have

highlighted that despite this bulky lesion, there is not a significantly large structural change in RNAPII on a CPD containing template (Brueckner et al., 2007; Lahiri et al., 2019). Instead it is hypothesised that it is the mis-incorporation of an NTP opposite the CPD that drives the stalling (Brueckner et al., 2007). It should be noted that this is in contrast to purine based dimers and older structural data of CPDs within a DNA duplex which both suggest these dimers do form severe structural changes within a DNA complex (Walmacq et al., 2015; Park et al., 2002). Therefore, further study of this damage is needed to better understand this stalling mechanism.

Although RNAPII cannot bypass this damage there is evidence for RNAPII exhibiting backtracking upon encountering a CPD but this has not been fully characterised (Brueckner et al., 2007; Lahiri et al., 2019). To repair this damage, eukaryotic cells will not use the BER pathway but instead uses the Nucleotide Excision Repair (NER) pathway to remove the damage in question and repair the template strand. Factors such as TFIIS, CSB and TFIIH have been shown to be important here (Xu et al., 2017a).

1.4.5 Damage response pathways

As mentioned earlier, DNA damage causes severe detrimental effects on transcription. In response to this, eukaryotic cells have developed repair pathways to target, remove and fix the damage to maintain genome integrity (Gregersen and Svejstrup, 2018). There are two main pathways within eukaryotic cells, BER and NER. While BER deals with small lesions, such as abasic and 8-Oxo guanine sites, NER deals with much larger adducts within the DNA that requires a multitude of proteins to help repair (Beard et al., 2019; Gregersen and Svejstrup, 2018). One sub-pathway of NER is transcription coupled-NER (TC-NER) which is triggered by RNAPII stalling at a specific damage and not the damage

itself (Gregersen and Svejstrup, 2018). If these pathways are not enough to fix the damaged lesion then RNAPII itself will become degraded upon Rpb1 ubiquitylation, in turn removing it from the DNA template to allow complete repair of the damage (Gregersen and Svejstrup, 2018; Herlihy et al., 2022).

Base Excision Repair

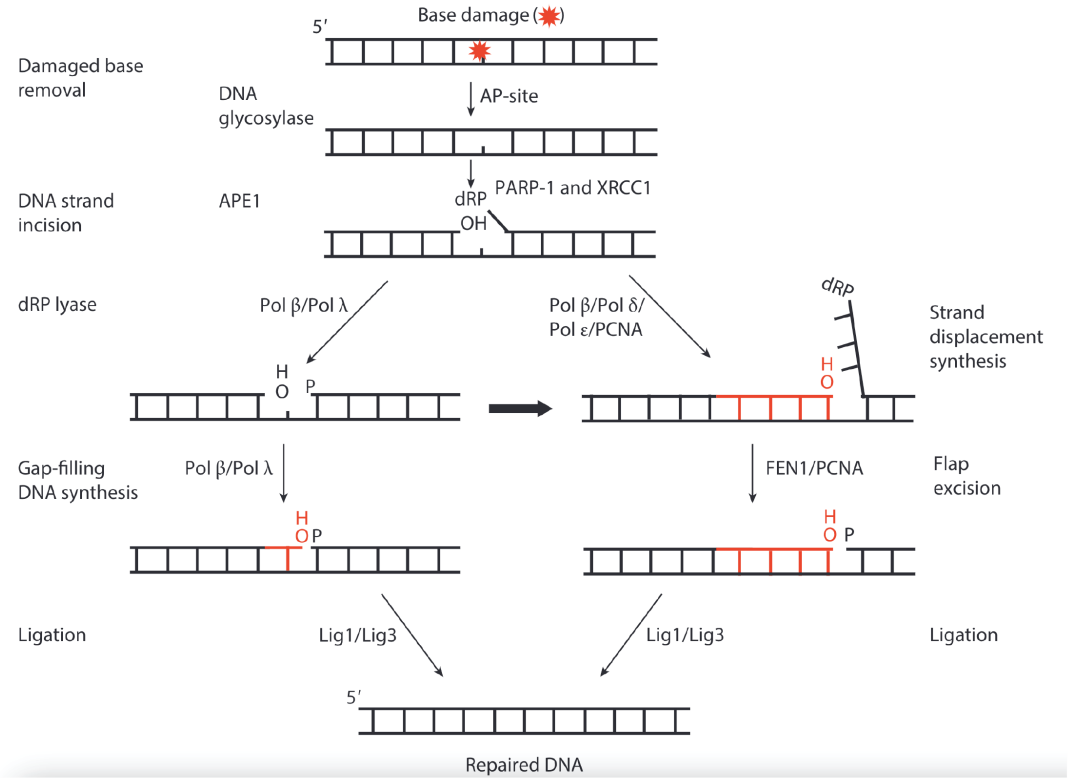


Figure 1.16: **Schematic of the base excision repair pathway.** Schematic of BER pathway and the associated factors involved in the process. DNA glycosylases recognise and cleave the damage base, AP endonucleases (APE1) nick the abasic site which is filled in the DNA polymerases and ligated together. Figure adapted from (Beard et al., 2019)

Although small lesions like 8-Oxo guanine and abasic sites do not cause significant structural changes to the DNA backbone, they often lead to mutations within the genome and must be fixed. In this instance they are fixed by the BER pathway which is initiated upon recognition of the DNA lesion by DNA glycosylases (Beard et al., 2019) (Figure 1.16). Specific glycosylases recognise different lesions, for example OCG1 and MYH target 8-Oxo

guanine lesions either paired with the cognate or non-cognate NTP respectively (Beard et al., 2019; Banda et al., 2017). Once detected, the glycosylase will distort the DNA around the lesion and 'flip out' the base from the DNA backbone cleaving the base from the phosphate backbone (Mol et al., 1995; Parikh et al., 2000). This flip-out and cleavage mechanism has been investigated for some time and the consensus is that it has a two stage recognition mechanism (Beard et al., 2019). This mechanism consists of an initial fast step which is termed the burst phase, linked to the catalytic rate constant of the enzyme, followed by a slower steady-state rate linked with the dissociation rate constant for the DNA (Beard et al., 2019).

Once the base is removed an abasic site remains within the DNA backbone, which itself is a form of damage as mentioned above. Apurinic/apyrimidine endonucleases must now act on the abasic site inducing a nick in the DNA backbone. This nick is essential as it now forms a 5' flap which can be targeted by the downstream BER factors (Figure 1.16). In mammalian cells, APE1 is the major endonuclease that targets most abasic site (Beard et al., 2019). It has also been reported that like DNA glycosylases, APE1 exhibits a similar kinetic pattern where there is an initial burst phase followed by a slower phase, although these rates vary considerably depending on the damage being targeted (Maher and Bloom, 2007; Schermerhorn and Delaney, 2013). This mechanism involves a deprotonated water molecule which interacts with APE1, acting as a nucleophile attacking the 5' phosphate group of the abasic lesion, in turn leaving a 5' phosphate and a 3' OH group on the adjacent base (Mol et al., 2000). The final two steps of this process are filling in the missing base and re-ligating it to the DNA backbone to maintain a fully dsDNA genome (Beard et al., 2019). DNA polymerases β and λ fill in the single nucleotide gap via a fill-in process while DNA polymerase β, ϵ and δ will replace not only the single excised base but also undergo

strand displacement synthesises replacing several nucleotides within the DNA backbone (Beard et al., 2019). Once the gap is filled, DNA ligases including Lig1/Lig3 will re-ligate the nick within the DNA backbone reforming the important phosphodiester bond within the DNA backbone (Beard et al., 2019) (Figure 1.16).

This pathway is the general BER pathway used for all small lesions within the nucleus and will also occur within the transcription pathway, however new evidence has shed light on a potential sub-pathway, transcription coupled-BER (TC-BER) (Chakraborty et al., 2021). Endonucleases including NEIL2 have been shown to be active specifically on DNA lesions within the template strand at stalled RNAPII sites (Chakraborty et al., 2021). Our understanding of this pathway is still in its early days but it is likely that transcription coupled repair pathways are highly inter-connected, with NER factors also being shown to interact with the BER repair pathway (Kumar et al., 2022).

Transcription Coupled-Nucleotide Excision Repair

Unlike other repair pathways, TC-NER is triggered by RNAPII stalling at a site of damage, rather than the damage lesion itself as in the case of BER. Once stalled, RNAPII effectively blocks the damage site from being recognised by DNA damage binding factors and hence why it is the focus for TC-NER instead (Brueckner et al., 2007) (Figure 1.16). Although originally discovered in mammalian cells it has now been shown to occur in bacteria and yeast (Mellon and Hanawalt, 1989; Smerdon and Thoma, 1990). One of the first steps of TC-NER is the stable association of the ATPase Cockayne syndrome group B (CSB) with RNAPII, which has been shown to transiently associate with elongating RNAPII but only becomes stably associated upon RNAPII stalling (Van Den Boom et al., 2004).

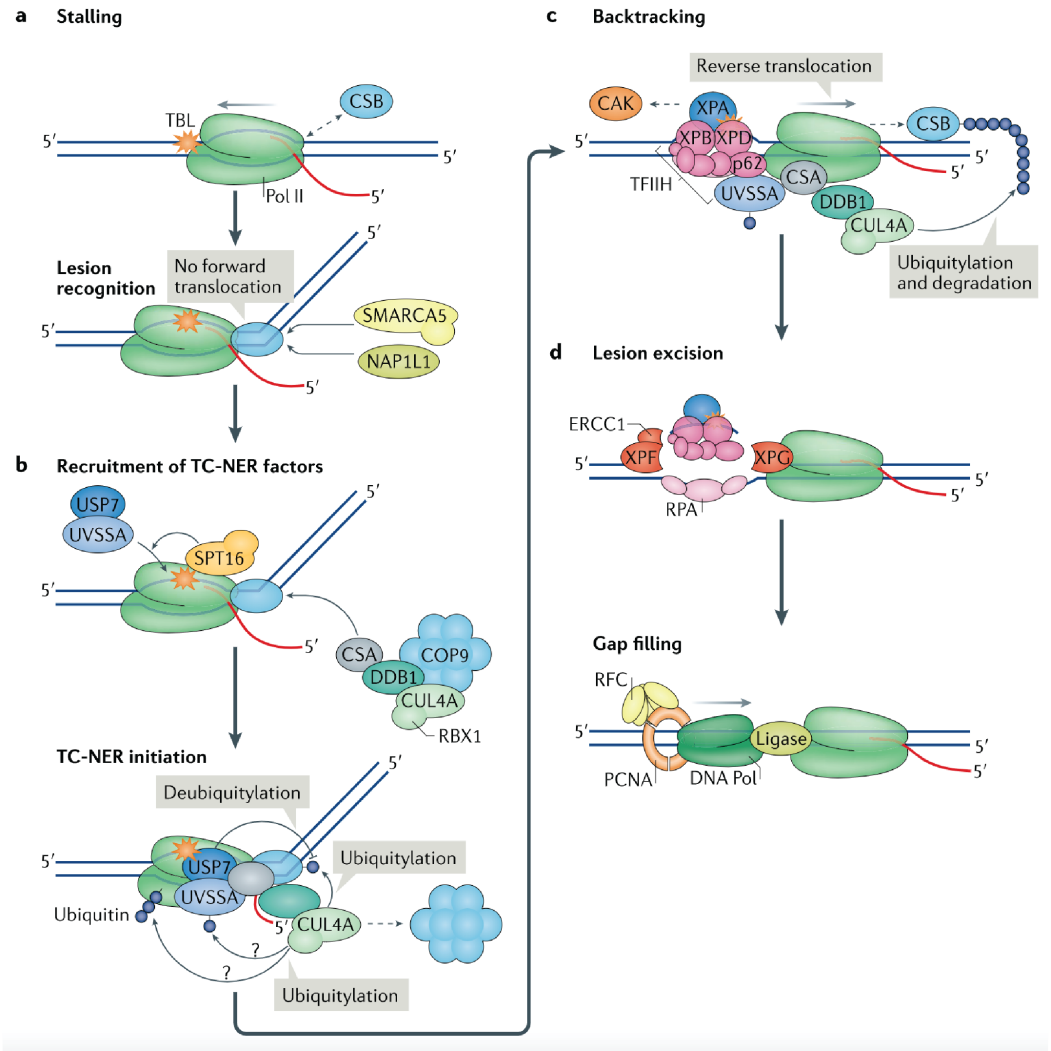


Figure 1.17: **Schematic of the TC-NER repair pathway.** Schematic of TC-NER pathway and the associated factors involved in the process. RNAPII becomes stalled at the DNA lesion and recognised by CSB. Downstream TC-NER factors will bind to RNAPII inducing lesion excision and eventual gap filling of the DNA backbone. Figure adapted from (Lans et al., 2019)

Recent structural work on the yeast orthologue of this protein, Rad26, has highlighted how it interacts with RNAPII stalled at sites of DNA damage (Xu et al., 2017a; Wang et al., 2018b). CSB binds upstream of RNAPII making several key contacts with the polymerase and effectively pushes RNAPII forward past the DNA lesion (Xu et al., 2017a). If the lesion is small enough CSB should push RNAPII beyond the damage and allow continuation of transcription. However, if the lesion is too large to be bypassed, such as CPDs, then CSB becomes associated with RNAPII stably and is recognised by downstream factors that initiate TC-NER (Lans et al., 2019). Bacteria have a similar protein

(Mfd) which has shown to have similar properties as CSB but has an additional feature not identified in eukaryotes, where it will translocate along DNA itself behind RNAPII and 'ram' into RNAPII inducing the pushing action (Le et al., 2018). Once TC-NER is initiated, other downstream factors will begin to associate upon CSB recognition such as UV-stimulated scaffold protein A (UVSSA) (Lans et al., 2019). Before the damage can be excised, RNAPII must be moved away from the stalled position to free up space for endonucleases to bind. This is achieved by backtracking of RNAPII away from the damage, an effect that has been seen in CPD stalled RNAPII (Xu et al., 2017a). This backtracking has been shown to be stimulated by TFIIS but also another TF, TFIIH (Lans et al., 2019).

As discussed in Section 1.3.2, TFIIH contains helicase modules that induced DNA opening in the initiation stage of transcription (Grünberg et al., 2012). These same helicase modules have been implicated in the NER pathway and it is hypothesised that this could provide enough force to push RNAPII backwards away from the damage site (Lans et al., 2019). Upon backtracking of RNAPII and the reveal of the damage, XPA will bind to the DNA in a single-stranded manner promoting the binding of the single-stranded binding protein RPA (Li et al., 2015). This in turn promotes recruitment of TC-NER endonucleases ERCC1-XPF and XPG, excising the damage within a 20-30 nucleotide region which is then filled and re-ligated by DNA synthesis machinery (PCNA, DNA polymerase δ, ϵ and κ) (Lans et al., 2019) (Figure 1.16). Once fixed, RNAPII is then free to transcribe the undamaged template as normal. It should be noted that the regulation of this pathway is also very important as mis-regulation can also lead to severe genetic disorders such as Cockayne Syndrome (CS) or Xeroderma pigmentosum (XP) (Gregersen and Svejstrup, 2018).

Although the cell has these repair pathways in place, RNAPII will still encounter these damages before they are fixed and so understanding how RNAPII interacts with them is of great importance. Furthermore, the kinetic mechanism of repair factors such as Rad26 interaction with RNAPII is poorly understood and a single-molecule approach will yield further information on these dynamic processes.

1.5 Single-molecule studies of transcription

A detailed knowledge of RNAPII transcription mechanisms is vital to understanding this complex and dynamic process. A large amount of our current understanding has come from a plethora of structural and biochemical studies that have shed light on key snapshots of RNAPII at several intermediates along the transcription pathway allowing better understanding of RNAPII interaction with DNA damage and repair pathways. Single-molecule assays have often been utilised alongside structural work to investigate the kinetics and conformational dynamics of RNAPII binding, elongation and stalling behaviour.

Some of the first single-molecule studies that probed the transcription pathway was done on bacterial RNAP. Using fluorescently labelled RNAP Kapanidis *et al.* probed the initial initiation mechanism of bacterial RNAP. Exactly how RNAP opened up DNA and began transcription was not fully understood and three main models were originally proposed (Kapanidis et al., 2006). The transient excursions model where RNAP transiently cycles between forward and backwards movement until it reaches a critical length of RNA would involve RNAP moving as a whole unit. The second model, inchworming, would involve a flexible element of RNAP where the mobile element would move backwards and forwards eventually pulling in the DNA into the active site. The final model, scrunching, would

also involve a flexible element but within the DNA strand not RNAP itself. In this model, RNAP pulls downstream DNA into the active site allowing RNA synthesis until a critical length is reached and RNAP escapes the promoter (Kapanidis et al., 2006). To discriminate between these models, both RNAP and the DNA were labelled with FRET pairs to monitor the distance between them upon addition of NTPs (Kapanidis et al., 2006). Using a series of experiments, these models were whittled down until only one possible mechanism could explain all of the data. One of the first experiments tested if the leading and trailing edge of RNAP moved closer or further to the DNA downstream/upstream of it respectively. This revealed that while the leading edge showed movement relative to the DNA, the trailing edge did not essentially eliminating the transient excursion model. The second experiment probed the potential expansion within RNAP, and the FRET remained unchanged upon addition of NTPs suggesting the inchworming model would not account for these results (Kapanidis et al., 2006). This left the scrunching model which was validated by FRET experiments showing the DNA downstream of RNAP was pulled closer towards the DNA upstream of it, without any movement in RNAP (Kapanidis et al., 2006). At the same time, a similar result which showed RNAP used a scrunching mechanism to initiate transcription was carried out using magnetic tweezers further validating this hypothesis (Revyakin et al., 2006).

Unlike bacterial RNA Polymerase, which can bind DNA with aid of a few factors, RNAPII initiation requires the binding of a multitude of GTFs (Section 1.3.2). Due to this complexity studying eukaryotic initiation has been difficult with most of our understanding of this process has come from biochemical reconstituted systems and structural assays where the PIC has been solved (Hantsche and Cramer, 2017). *In vivo* single-molecule assays have shed light on this process by fluorescently labelling of endogenous RNAPII and mon-

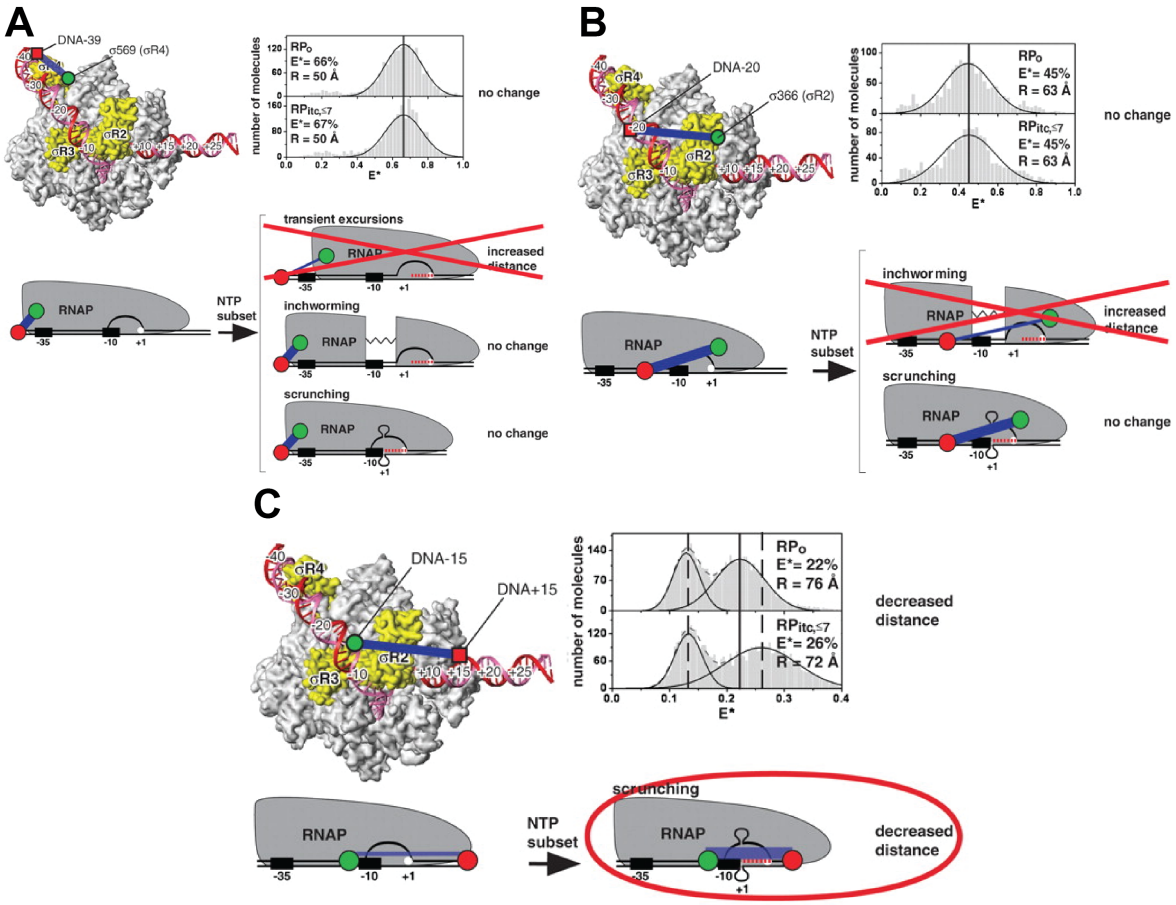


Figure 1.18: **smFRET assay revealed scrunching mechanism of RNAP initiation.** smFRET experiment to probe the different RNAP initiation mechanism models. (A) labelling of RNAP (green circle) on the trailing edge with a FRET pair (red circle) on the DNA showed no change in FRET (0.5) upon addition of NTPs which eliminated the transient excursion model (red cross). (B) labelling of RNAP within the leading edge (green circle) and DNA upstream of this position (red circle) also showed no change in FRET (0.63) upon addition of NTPs eliminating the inchworming model. (C) validation of the scrunching model by labelling of the DNA with two FRET pairs (green and red circle) which showed a decrease in distance between the dyes upon addition of NTPs with no change in RNAP movement (0.76 to 0.72). Structure of RNAP with DNA and labelling position shown for all assays. FRET histograms show FRET observed in the different conditions tested and fit to a Gaussian fit. Figure adapted from (Kapanidis et al., 2006)

.

itoring it at the gene level within live cells, in turn highlighting the highly dynamic nature of RNAPII within a cellular context (Li et al., 2019). More recently, *in vitro* studies have been used to tackle the formation of the PIC which has challenged the canonical model of PIC formation (Baek et al., 2021; Revyakin et al., 2012). The first of these studies from Revyakin *et al.* utilised a surface based TIRF assay where DNA was immobilised onto a surface and GTF and RNAPII binding monitored via fluorescence. This showed

that RNAPII initiation and re-initiation are non-cooperative reactions and independent of each other, although this is in contrast to other evidence which suggests the transcription scaffold helps to promote multiple round of transcription initiation together (Yudkovsky et al., 2000). Work form Stephen Buratwoski has also probed the interaction of RNAPII at enhancers and activators bound to a surface. RNAPII and other GTFs were fluorescently labelled and their binding to a synthetic DNA construct containing enhancers and promoters was monitored using TIRF microscopy (Baek et al., 2021). This highlighted that contrary to the canonical model, which shows RNAPII and TFIIF arriving together, TFIIF can often ($\sim 50\%$) bind to the promoter first and then interact with RNAPII on the DNA (Baek et al., 2021). Furthermore, RNAPII often bound to enhancer regions upstream of the promoter first before association with GTFs at the promoter, a mechanism which has been suggested to occur *in vivo* where RNAPII at enhancers and promoters are tethered together via factors such as mediator (Osman and Cramer, 2020).

Monitoring of the elongation phase of RNAPII transcription posed several issues, mainly due to the complexity of the initiation stage. In order to monitor RNAPII at the elongation phase specifically, transcription elongation complexes (TECs) were generated consisting of a synthetic TS, RNA primer, RNAPII and NTS. The TS and RNA primer are first annealed together, followed by binding of RNAPII and then annealing of the NTS to form a full TEC (Oh et al., 2019). The majority of these studies have used optical tweezers to monitor RNAPII elongation (Dangkulwanich et al., 2013; Lee and Myong, 2021). Optical tweezers involves using optically trapped beads which hold a DNA molecule between them (explained in more detail in Section 2.3) (Figure 1.19). In this instance, RNAPII is often biotinylated and upon forming a TEC is ligated into a larger construct (of several kilo-bases) where one end of the DNA is attached to one bead, and RNAPII to the other

(Dangkulwanich et al., 2013) (Figure 1.19). Observation of RNAPII was achieved by monitoring the movement of the beads, one of which is directly linked to RNAPII. One of the benefits of optical tweezers is the application of force onto the construct, which has been used to monitor the stalling force of RNAPII (Zhou et al., 2013). The use of this assay helped to validate the Brownian ratchet model of transcription elongation, as well as highlighting that transcription elongation contains two major phases, active elongation and pausing (Davenport et al., 2000; Galburt et al., 2007). These optical tweezer assays showed that RNAPII will naturally pause along the DNA template exhibiting small but rapid backtracking events, even on DNA templates that contained no roadblocks to transcription (Shundrovsky et al., 2004). Instead inherent DNA sequences could lead to pausing of RNAPII, and in some cases lead to backtracking.

A similar strategy was used to monitor RNAPII elongation through DNA roadblocks including nucleosomes (Zhou et al., 2013). Early studies of this process showed that when a nucleosome was placed downstream of RNAPII, it would induce pausing at the nucleosome as had been seen in biochemical assays (Kulaeva et al., 2013). This experiment showed that passage through the nucleosome depended on RNAPII dynamics and natural fluctuations within the nucleosome itself. RNAPII pausing increased upon encountering nucleosomes which cannot bypass the barrier, but instead must wait for natural fluctuations within the DNA-nucleosome region (Kulaeva et al., 2013). This was backed up by separate assays that showed RNAPII can generate a force of around 7-8pN, but a nucleosome requires at least 10-20pN to become fully unwrapped (Galburt et al., 2007; Chen et al., 2019). More recently, RNAPII stalling within a canonical and H2A variant nucleosome has been investigated with higher spatial resolution to allow precise monitoring of stalling positions within the nucleosome by use of a 'molecular ruler' (Chen et al.,

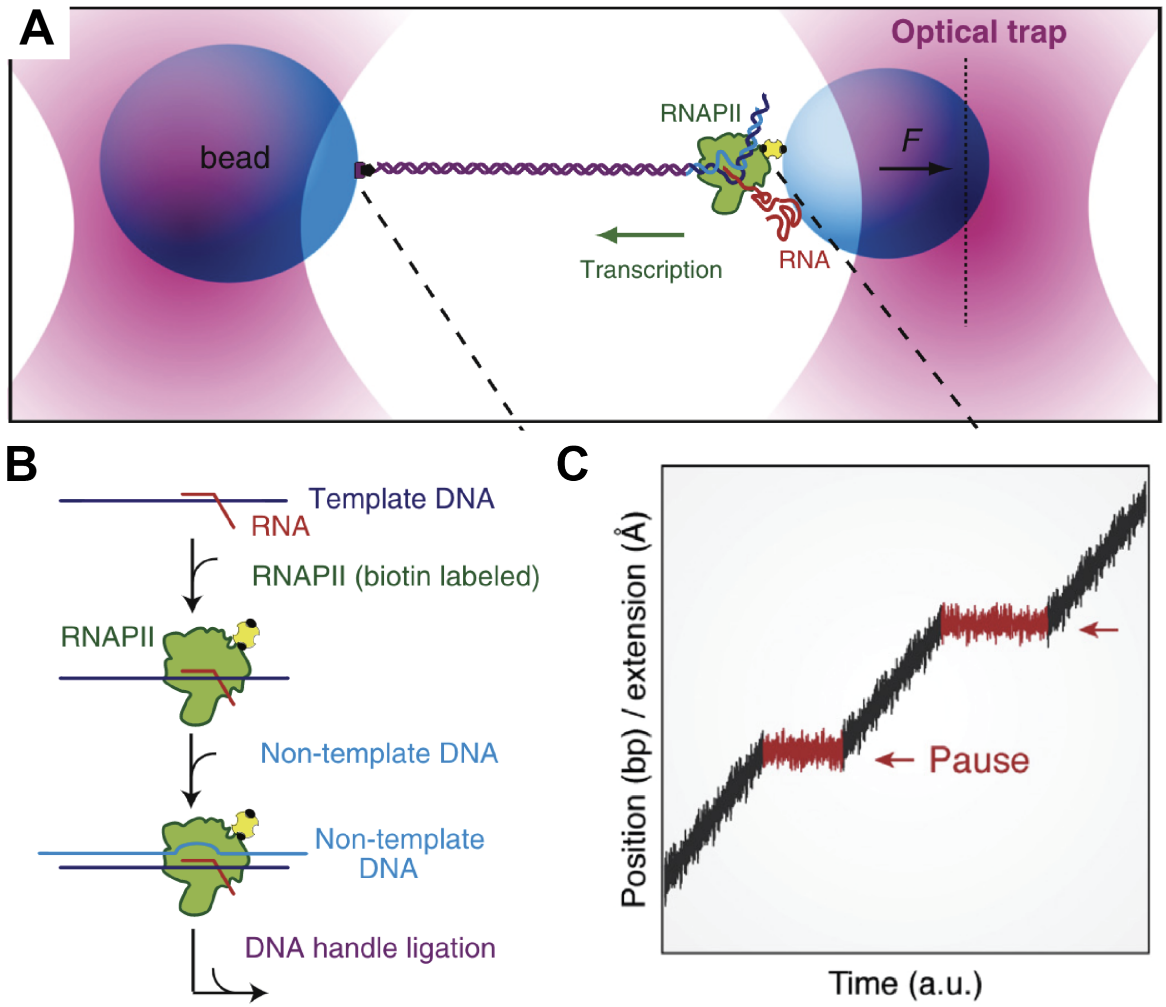


Figure 1.19: **Optical tweezer assay to monitor RNAPII elongation.** (A) Schematic of optical tweezer set up with beads (blue circles) being trapped within optical traps. DNA is tethered between them with RNAPII (green) immobilised to one bead via biotin (black circle)-streptavidin (yellow) linkage. Synthesised RNA is in red but is not directly visualised. (B) Formation of the TEC with RNAPII bound to DNA:RNA hybrid before ligation to DNA handle for optical tweezer assay. (C) Example of a teal time extension trace over time showing RNAPII transcribing and pausing (red arrow) at specific locations. Figure adapted from (Lee and Myong, 2021; Zhou et al., 2013)

.

2019). This ruler consisted of 64 bp of known DNA and by comparing the measured distance against this region they were able to identify where along the template RNAPII was, even within the nucleosome (Chen et al., 2019). The pausing positions within the nucleosome matched what has been solved by high-resolution structural data of RNAPII within nucleosomes (Farnung et al., 2018; Kujirai et al., 2018).

1.6 Thesis Objectives

Transcription within eukaryotes is a highly complex process involving multiple factors including RNA Polymerase II, the core processive enzyme synthesising protein coding genes. The elongation phase of this process is made even more complex due to the nature of its DNA template, which is full of roadblocks preventing efficient and processive transcription. These roadblocks pose a severe threat to transcription and have been shown to cause transcriptional stalling, backtracking and arrest. If not dealt with, many of these lesions can cause genetic mutations eventually leading to severe diseases and cancers. Although pathways within eukaryotic cells can address these damages, they themselves can become mis-regulated also leading to genetic disorders, such as Xeroderma pigmentosum. Understanding the key mechanism of RNAPII at these lesions and how repair factors interact with RNAPII is of vital importance to enable us to better address these health issues. Key questions still remain though as to exactly how different damages can cause such different outcomes within the transcription pathway. And how do repair factors like CSB interact with RNAPII and trigger the repair pathway required to fix the damage.

Biochemical and structural studies have provided invaluable insight into these processes elucidating interactions and mechanisms with a plethora of elongation factors, repair factors and DNA roadblocks that have developed our knowledge of this complex process. Alongside these structural studies, single-molecule assays have also been key in highlighting the dynamics and intermediates that RNAPII transitions between during this process. Although single-molecules assays to study eukaryotic transcription elongation do exist, these have primarily been carried out with optical tweezers and modified RNAPII. One of the major issues with measuring elongation dynamics using this method is the spatial

resolution of optical tweezers, which is not as accurate as one can achieve using TIRF or FRET based assays. Although recent work has been able to achieve high spatial resolution within nucleosomes this requires heavy modification to RNAPII itself (biotinylation) and attachment to large polystyrene beads (1-4 µm) which is not always possible or desirable. Furthermore, placement of smaller DNA lesions like abasic sites, 8-Oxo guanine and CPDs within a large piece of DNA can be complex and non-efficient, making the use of smaller synthetic DNA complexes more suitable. Moreover, monitoring elongation using a FRET based system would allow for more direct measurement of transcription dynamics at a higher resolution than most optical tweezer assays will generate, especially at small lesions that I aim to study.

Therefore the aim of this study was to generate a novel single-molecule FRET based assay that would require no modification to RNAPII itself, allowing the use of eukaryotic RNAPII from mammalian sources, and would allow direct visualisation of RNAPII dynamics at small DNA lesions placed specifically within a DNA template. Throughout this thesis I will lay out how the FRET assay was developed using both mammalian and yeast RNAPII to study the dynamics at different DNA lesions to address key questions that are still unanswered (Chapter 4). Furthermore, I use this assay to investigate how repair factors including Rad26 interact with RNAPII at sites of damage in order to elucidate the mechanism of triggering TC-NER (Chapter 5). Due to the power that optical tweezers and fluorescence can achieve when combined, in the final chapter (Chapter 6) I will discuss my attempts to also generate constructs that would allow direct visualisation of RNAPII that also do not require genetic modification of RNAPII allowing this powerful technique to be adapted to eukaryotic sample such as mammalian RNAPII.

Bibliography

E. A. Abbondanzieri, W. J. Greenleaf, J. W. Shaevitz, R. Landick, and S. M. Block. Direct observation of base-pair stepping by RNA polymerase. *Nature*, 438(7067):460–465, 11 2005. ISSN 1476-4687. doi: 10.1038/NATURE04268. URL https://pubmed.ncbi.nlm.nih.gov/16284617/.

K. Adelman and J. T. Lis. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nature reviews. Genetics*, 13(10):720, 2012. ISSN 14710064. doi: 10.1038/NRG3293. URL /pmc/articles/PMC3552498//pmc/articles/PMC3552498/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC3552498/.

P. Allepuz-fuster, M. J. O. Brien, N. González-polo, and B. Pereira. RNA polymerase II plays an active role in the formation of gene loops through the Rpb4 subunit. (34), 2019.

G. Andrey and S. Mundlos. The three-dimensional genome: regulating gene expression during pluripotency and development. *Development (Cambridge, England)*, 144(20):3646–3658, 10 2017. ISSN 1477-9129. doi: 10.1242/DEV.148304. URL https://pubmed.ncbi.nlm.nih.gov/29042476/.

A. G. Arimbasseri and R. J. Maraia. Mechanism of transcription termination by RNA polymerase III utilizes a nontemplate-strand sequence-specific signal element. *Molecular cell*, 58(6):1124, 6 2015. ISSN 10974164. doi: 10.1016/J.MOLCEL.2015.04.002. URL /pmc/articles/PMC4475470//pmc/articles/PMC4475470/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4475470/.

K. J. Armache, H. Kettenberger, and P. Cramer. Architecture of initiation-competent 12-subunit RNA polymerase II. *Proceedings of the National Academy of Sciences of the United States of America*, 100(12):6964–6968, 6 2003. ISSN 0027-8424. doi: 10.1073/PNAS.1030608100. URL https://pubmed.ncbi.nlm.nih.gov/12746495/.

I. Baek, L. J. Friedman, J. Gelles, and S. Buratowski. Single-molecule studies reveal branched pathways for activator-dependent assembly of RNA polymerase II pre-initiation complexes. *Molecular Cell*, 81(17):3576–3588, 2021. ISSN 10974164. doi: 10.1016/j.molcel.2021.07.025. URL https://doi.org/10.1016/j.molcel.2021.07.025.

D. M. Banda, N. N. Nuñez, M. A. Burnside, K. M. Bradshaw, and S. S. David. Repair of 8-oxoG:A Mismatches by the MUTYH Glycosylase: Mechanism, Metals & Medicine. *Free radical biology & medicine*, 107:202, 6 2017. ISSN 18734596. doi: 10.1016/J.FREERADBIOMED.2017.01.008. URL /pmc/articles/PMC5457711//pmc/articles/PMC5457711/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC5457711/.

W. A. Beard, J. K. Horton, R. Prasad, and S. H. Wilson. Eukaryotic Base Excision Repair: New Approaches Shine Light on Mechanism. *Annual Review of Biochemistry*, 88:137–162, 6 2019. ISSN 15454509. doi: 10.1146/ANNUREV-BIOCHEM-013118-111315. URL https://www.annualreviews.org/doi/abs/10.1146/annurev-biochem-013118-111315.

C. Bernecky, F. Herzog, W. Baumeister, J. M. Plitzko, and P. Cramer. Structure of transcribing mammalian RNA polymerase II. *Nature*, 529(7587):551–554, 2016. ISSN 14764687. doi: 10.1038/nature16482. URL http://dx.doi.org/10.1038/nature16482.

S. Boiteux and M. Guillet. Abasic sites in DNA: repair and biological consequences in Saccharomyces cerevisiae. *DNA repair*, 3(1):1–12, 1 2004. ISSN 1568-7864. doi: 10.1016/J.DNAREP.2003.10.002. URL https://pubmed.ncbi.nlm.nih.gov/14697754/.

S. Borkotoky and A. Murali. The highly efficient T7 RNA polymerase: A wonder macromolecule in biological realm. *International Journal of Biological Macromolecules*, 118:49–56, 10 2018. ISSN 18790003. doi: 10.1016/J.IJBIOMAC.2018.05.198.

J. Brosius and C. A. Raabe. What is an RNA? A top layer for RNA classification. *RNA Biology*, 13(2):140, 2 2016. ISSN 15558584. doi: 10.1080/15476286.2015.1128064. URL /pmc/articles/PMC4829331//pmc/articles/PMC4829331/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4829331/.

F. Brueckner, U. Hennecke, T. Carell, and P. Cramer. CPD Damage Recognition by Transcribing RNA Polymerase II. *Science*, 315(February):859–862, 2007.

F. Brueckner, J. Ortiz, and P. Cramer. A movie of the RNA polymerase nucleotide addition cycle. *Current Opinion in Structural Biology*, 19(3):294–299, 6 2009. ISSN 0959-440X. doi: 10.1016/J.SBI.2009.04.005.

M. Buck, M. T. Gallegos, D. J. Studholme, Y. Guo, and J. D. Gralla. The bacterial enhancer-dependent sigma(54) (sigma(N)) transcription factor. *Journal of bacteriology*, 182(15):4129–4136, 8 2000. ISSN 0021-9193. doi: 10.1128/JB.182.15.4129-4136.2000. URL https://pubmed.ncbi.nlm.nih.gov/10894718/.

R. R. Burgess. RNA POLYMERASE. *Annual Review of Biochemistry*, 40:711–740, 1971. URL www.annualreviews.org.

R. R. Burgess, A. A. Travers, J. J. Dunn, and E. K. Bautz. Factor Stimulating Transcription by RNA Polymerase. *Nature 1969 221:5175*, 221(5175):43–46, 1969. ISSN 1476-4687. doi: 10.1038/221043a0. URL https://www.nature.com/articles/221043a0.

D. A. Bushnell and R. D. Kornberg. Complete, 12-subunit RNA polymerase II at 4.1-A resolution: implications for the initiation of transcription. *Proceedings of the National Academy of Sciences of the United States of America*, 100(12):6969–6973, 6 2003. ISSN 0027-8424. doi: 10.1073/PNAS.1130601100. URL https://pubmed.ncbi.nlm.nih.gov/12746498/.

A. Chakraborty, N. Tapryal, A. Islam, S. Mitra, and T. Hazra. Transcription coupled base excision repair in mammalian cells: So little is known and so much to uncover. *DNA Repair*, 107:103204, 11 2021. ISSN 1568-7864. doi: 10.1016/J.DNAREP.2021.103204.

G. M. Cheetham and T. A. Steitz. Insights into transcription: structure and function of single-subunit DNA-dependent RNA polymerases. *Current opinion in structural biology*, 10(1):117–123, 2 2000. ISSN 0959-440X. doi: 10.1016/S0959-440X(99)00058-5. URL https://pubmed.ncbi.nlm.nih.gov/10679468/https://pubmed.ncbi.nlm.nih.gov/10679468/?dopt=Abstract.

G. M. Cheetham, D. Jeruzalmi, and T. A. Steltz. Structural basis for initiation of transcription from an RNA polymerase–promoter complex. *Nature 1999 399:6731*, 399(6731):80–83, 5 1999. ISSN 1476-4687. doi: 10.1038/19999. URL https://www.nature.com/articles/19999.

H. T. Chen, L. Warfield, and S. Hahn. The positions of TFIIF and TFIIE in the RNA polymerase II transcription preinitiation complex. *Nature Structural & Molecular Biology 2007 14:8*, 14(8):696–703, 7 2007. ISSN 1545-9985. doi: 10.

1038/nsmb1272. URL https://www.nature.com/articles/nsmb1272https://www.nature.com/articles/nsmb1272/.

Z. Chen, S. Li, S. Subramaniam, J. Y.-J. Shyy, and S. Chien. Epigenetic Regulation: A New Frontier for Biomedical Engineers. *Annual Review of Biomedical Engineering*, 19 (1):195–219, 2017. ISSN 1523-9829. doi: 10.1146/annurev-bioeng-071516-044720. URL http://www.annualreviews.org/doi/10.1146/annurev-bioeng-071516-044720.

Z. Chen, R. Gabizon, A. I. Brown, A. Lee, A. Song, C. Díaz-Celis, C. D. Kaplan, E. F. Koslover, T. Yao, and C. Bustamante. High-resolution and high-accuracy topographic and transcriptional maps of the nucleosome barrier. *eLife*, 8:1–38, 2019. ISSN 2050-084X. doi: 10.7554/elife.48281.

Z. A. Chen, A. Jawhari, L. Fischer, C. Buchen, S. Tahir, T. Kamenski, M. Rasmussen, L. Lariviere, J. C. Bukowski-Wills, M. Nilges, P. Cramer, and J. Rappsilber. Architecture of the RNA polymerase II–TFIIF complex revealed by cross-linking and mass spectrometry. *The EMBO Journal*, 29(4):717, 2 2010. ISSN 02614189. doi: 10.1038/EMBOJ.2009.401. URL /pmc/articles/PMC2810376//pmc/articles/PMC2810376/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC2810376/.

A. C. Cheung and P. Cramer. Structural basis of RNA polymerase II backtracking, arrest and reactivation. *Nature*, 471(7337):249–253, 3 2011. ISSN 1476-4687. doi: 10.1038/NATURE09785. URL https://pubmed.ncbi.nlm.nih.gov/21346759/.

A. C. Cheung, S. Sainsbury, and P. Cramer. Structural basis of initial RNA polymerase II transcription. *The EMBO journal*, 30(23):4755–4763, 11 2011. ISSN 1460-2075. doi: 10.1038/EMBOJ.2011.396. URL https://pubmed.ncbi.nlm.nih.gov/22056778/.

E. J. Cho, T. Takagi, C. R. Moore, and S. Buratowski. mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminaldomain. *Genes & Development*, 11(24):3319, 12 1997. ISSN 08909369. doi: 10.1101/GAD.11.24.3319. URL /pmc/articles/PMC316800//pmc/articles/PMC316800/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC316800/.

E. J. Cho, M. S. Kobor, M. Kim, J. Greenblatt, and S. Buratowski. Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal

domain. *Genes & development*, 15(24):3319–3329, 12 2001. ISSN 0890-9369. doi: 10.1101/GAD.935901. URL https://pubmed.ncbi.nlm.nih.gov/11751637/.

S. Connelly and J. Manley. A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. *Genes Dev.*, 2:440–452, 1988.

L. Core and K. Adelman. Promoter-proximal pausing of RNA polymerase II: A nexus of gene regulation. *Genes and Development*, 33(15-16):960–982, 2019. ISSN 15495477. doi: 10.1101/gad.325142.119.

L. J. Core, J. J. Waterfall, and J. T. Lis. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science (New York, N.Y.)*, 322 (5909):1845–1848, 12 2008. ISSN 1095-9203. doi: 10.1126/SCIENCE.1162228. URL https://pubmed.ncbi.nlm.nih.gov/19056941/.

M. Cortazar and e. al. Control of RNA Pol II speed by PNUTS-PP1 and Spt5 dephosphorylation facilitates termination by a "sitting duck torpedo" mechanism. *Mol. Cell*, 76:896–908, 2019.

P. Cramer. Organization and regulation of gene transcription, 9 2019a. ISSN 14764687.

P. Cramer. Eukaryotic Transcription Turns 50. *Cell*, 179(4):808–812, 2019b. ISSN 10974172. doi: 10.1016/j.cell.2019.09.018. URL https://doi.org/10.1016/j.cell.2019.09.018.

P. Cramer, D. A. Bushnell, J. Fu, A. L. Gnatt, B. Maier-Davis, N. E. Thompson, R. R. Burgess, A. M. Edwards, P. R. David, and R. D. Kornberg. Architecture of RNA polymerase II and implications for the transcription mechanism. *Science (New York, N.Y.)*, 288(5466):640–649, 4 2000. ISSN 0036-8075. doi: 10.1126/SCIENCE.288.5466.640. URL https://pubmed.ncbi.nlm.nih.gov/10784442/.

P. Cramer, D. A. Bushnell, and R. D. Kornberg. Structural Basis of Transcription: RNA Polymerase II at 2.8 Ångstrom Resolution. *Science*, 292(5523):1863–1876, 6 2001. ISSN 00368075. doi: 10.1126/SCIENCE.1059493. URL https://www.science.org/doi/abs/10.1126/science.1059493.

F. Crick. Central Dogma of Molecular Biology. *Nature 1970 227:5258*, 227(5258):561–563, 1970. ISSN 1476-4687. doi: 10.1038/227561a0. URL https://www.nature.com/articles/227561a0.

M. Crossley and e. al. R-loops as cellular regulators and genomic threats. *Mol. Cell*, 73: 398–411, 2019.

A. R. Cutter and J. J. Hayes. A Brief Review of Nucleosome Structure. *FEBS letters*, 589(20 0 0):2914, 10 2015. ISSN 18733468. doi: 10.1016/J.FEBSLET.2015.05.016. URL /pmc/articles/PMC4598263//pmc/articles/PMC4598263/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4598263/.

G. E. Damsma and P. Cramer. Molecular basis of transcriptional mutagenesis at 8-oxoguanine. *Journal of Biological Chemistry*, 284(46):31658–31663, 2009. ISSN 00219258. doi: 10.1074/jbc.M109.022764.

M. Dangkulwanich, T. Ishibashi, S. Liu, M. L. Kireeva, L. Lubkowska, M. Kashlev, and C. J. Bustamante. Complete dissection of transcription elongation reveals slow translocation of RNA polymerase II in a linear ratchet mechanism. *eLife*, 2013(2):1–22, 2013. ISSN 2050084X. doi: 10.7554/eLife.00971.

R. J. Davenport, G. J. Wuite, R. Landick, and C. Bustamante. Single-molecule study of transcriptional pausing and arrest by E. coli RNA polymerase. *Science (New York, N.Y.)*, 287(5462):2497–2500, 3 2000. ISSN 0036-8075. doi: 10.1126/SCIENCE.287.5462.2497. URL https://pubmed.ncbi.nlm.nih.gov/10741971/.

S. Dubbury and e. al. CDK12 regulates DNA repair genes by suppressing intronic polyadenylation. *Nature*, 564:141–145, 2018.

D. Duchi, D. L. V. Bauer, L. Fernandez, Z. Morichaud, K. Brodolin, and A. N. Kapanidis Correspondence. RNA Polymerase Pausing during Initial Transcription. *Molecular Cell*, 63:939–950, 2016. doi: 10.1016/j.molcel.2016.08.011. URL http://dx.doi.org/10.1016/j.molcel.2016.08.011.

J. Eaton and e. al. Xrn2 accelerates termination by RNA polymerase II, which is underpinned by CPSF73 activity. *Genes Dev.*, 32:127–139, 2018.

J. Eaton and e. al. A unified allosteric/torpedo mechanism for transcriptional termination on human protein-coding genes. *Genes Dev.*, 34:132–145, 2020.

J. D. Eaton and S. West. Termination of Transcription by RNA Polymerase II: BOOM! *Trends in Genetics*, 36(9):664–675, 9 2020. ISSN 0168-9525. doi: 10.1016/J.TIG.2020.05.008. URL http://www.cell.com/article/S0168952520301256/fulltexthttp://www.cell.com/article/S0168952520301256/abstracthttps://www.cell.com/trends/genetics/abstract/S0168-9525(20)30125-6.

H. Ehara, T. Kujirai, Y. Fujino, M. Shirouzu, H. Kurumizaka, and S.-i. Sekine. Structural insight into nucleosome transcription by RNA polymerase II with elongation factors. *Science*, 747(February):eaav8912, 2019. ISSN 0036-8075. doi: 10.1126/science.aav8912.

J. Eichner, H. T. Chen, L. Warfield, and S. Hahn. Position of the general transcription factor TFIIF within the RNA polymerase II transcription preinitiation complex. *The EMBO Journal*, 29(4):706, 2 2010. ISSN 02614189. doi: 10.1038/EMBOJ.2009.386. URL /pmc/articles/PMC2829161//pmc/articles/PMC2829161/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC2829161/.

L. Farnung, S. M. Vos, and P. Cramer. Structure of transcribing RNA polymerase II-nucleosome complex. *Nature Communications*, 9(1):1–6, 2018. ISSN 20411723. doi: 10.1038/s41467-018-07870-y. URL http://dx.doi.org/10.1038/s41467-018-07870-y.

L. Farnung, M. Ochmann, G. Garg, S. M. Vos, and P. Cramer. Structure of a backtracked hexasomal intermediate of nucleosome transcription. *Molecular Cell*, 7 2022. ISSN 1097-2765. doi: 10.1016/J.MOLCEL.2022.06.027. URL http://www.ncbi.nlm.nih.gov/pubmed/35858621.

A. Feklístov, B. D. Sharon, S. A. Darst, and C. A. Gross. Bacterial Sigma Factors: A Historical, Structural, and Genomic Perspective. *http://dx.doi.org/10.1146/annurev-micro-092412-155737*, 68:357–376, 9 2014. ISSN 15453251. doi: 10.1146/ANNUREV-MICRO-092412-155737. URL https://www.annualreviews.org/doi/abs/10.1146/annurev-micro-092412-155737.

A. Field and K. Adelman. Evaluating Enhancer Function and Transcription. *Annual Review of Biochemistry*, 89:213–234, 2020. ISSN 15454509. doi: 10.1146/annurev-biochem-011420-095916.

J. Fishburn, E. Tomko, E. Galburt, and S. Hahn. Double-stranded DNA translocase activity of transcription factor TFIIH and the mechanism of RNA polymerase II open complex formation. *Proceedings of the National Academy of Sciences*, 112(13):3961–3966, 2015. ISSN 0027-8424. doi: 10.1073/pnas.1417709112. URL http://www.pnas.org/lookup/doi/10.1073/pnas.1417709112.

T. Formosa and F. Winston. The role of FACT in managing chromatin: disruption, assembly, or repair? *Nucleic Acids Research*, 48(21):11929, 12 2020. ISSN 13624962. doi: 10.1093/NAR/GKAA912. URL /pmc/articles/PMC7708052//pmc/articles/PMC7708052/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC7708052/.

J. J. FURTH, J. HURWITZ, R. KRUG, and M. ALEXANDER. The Incorporation of Adenylic and Cytidylic Acids into Ribonucleic Acid. *Journal of Biological Chemistry*, 236(12):3317–3322, 12 1961. ISSN 0021-9258. doi: 10.1016/S0021-9258(18)94017-2.

E. A. Galburt, S. W. Grill, A. Wiedmann, L. Lubkowska, J. Choy, E. Nogales, M. Kashlev, and C. Bustamante. Backtracking determines the force sensitivity of RNAP II in a factor-dependent manner. *Nature*, 446(7137):820–823, 4 2007. ISSN 1476-4687. doi: 10.1038/NATURE05701. URL https://pubmed.ncbi.nlm.nih.gov/17361130/.

L. R. Ganser, M. L. Kelly, D. Herschlag, and H. M. Al-Hashimi. The roles of structural dynamics in the cellular functions of RNAs. *Nature Reviews Molecular Cell Biology 2019 20:8*, 20(8):474–489, 6 2019. ISSN 1471-0080. doi: 10.1038/s41580-019-0136-0. URL https://www.nature.com/articles/s41580-019-0136-0.

Y. Gholamalipour, A. Karunanayake Mudiyanselage, and C. T. Martin. 3' end additions by T7 RNA polymerase are RNA self-templated, distributive and diverse in character—RNA-Seq analyses. *Nucleic Acids Research*, 46(18):9253–9263, 10 2018. ISSN 0305-1048. doi: 10.1093/NAR/GKY796. URL https://academic.oup.com/nar/article/46/18/9253/5096072.

T. Ghosh, D. Bose, and X. Zhang. Mechanisms for activating bacterial RNA polymerase. *FEMS Microbiology Reviews*, 34(5):611–627, 9 2010. ISSN 0168-6445. doi: 10.1111/J.1574-6976.2010.00239.X. URL https://academic.oup.com/femsre/article/34/5/611/795499.

A. L. Gnatt, P. Cramer, J. Fu, D. A. Bushnell, and R. D. Kornberg. Structural Basis of Transcription: An RNA Polymerase II Elongation Complex at 3.3 Å Resolution. *Science*, 292(5523):1876–1882, 6 2001. ISSN 00368075. doi: 10.1126/SCIENCE.1059495. URL https://www.science.org/doi/abs/10.1126/science.1059495.

X. Q. Gong, Y. A. Nedialkov, and Z. F. Burton. Alpha-amanitin blocks translocation by human RNA polymerase II. *The Journal of biological chemistry*, 279(26):27422–27427, 6 2004. ISSN 0021-9258. doi: 10.1074/JBC.M402163200. URL https://pubmed.ncbi.nlm.nih.gov/15096519/.

R. L. Gourse, W. Ross, and T. Gaal. UPs and downs in bacterial transcription initiation: the role of the alpha subunit of RNA polymerase in promoter recognition. *Molecular microbiology*, 37(4):687–695, 2000. ISSN 0950-382X. doi: 10.1046/J.1365-2958.2000.01972.X. URL https://pubmed.ncbi.nlm.nih.gov/10972792/.

B. J. Greber, D. B. Toso, J. Fang, and E. Nogales. The complete structure of the human TFIIH core complex. *eLife*, 8, 3 2019. ISSN 2050-084X. doi: 10.7554/ELIFE.44771. URL https://pubmed.ncbi.nlm.nih.gov/30860024/.

L. Gregersen and e. al. SCAF4 and SCAF8, mRNA anti-terminator proteins. *Cell*, 177: 1797–1813, 2019.

L. H. Gregersen and J. Q. Svejstrup. The Cellular Response to Transcription-Blocking DNA Damage. *Trends in Biochemical Sciences*, 43(5):327–341, 2018. ISSN 13624326. doi: 10.1016/j.tibs.2018.02.010. URL http://dx.doi.org/10.1016/j.tibs.2018.02.010.

D. Grohmann, A. Hirtreiter, and F. Werner. RNAP subunits F/E (RPB4/7) are stably associated with archaeal RNA polymerase: using fluorescence anisotropy to monitor RNAP assembly in vitro. *The Biochemical journal*, 421(3):339–343, 8 2009. ISSN 1470-8728. doi: 10.1042/BJ20090782. URL https://pubmed.ncbi.nlm.nih.gov/19492989/.

T. M. Gruber and C. A. Gross. Multiple sigma subunits and the partitioning of bacterial transcription space. *Annual review of microbiology*, 57:441–466, 2003. ISSN 0066-4227. doi: 10.1146/ANNUREV.MICRO.57.030502.090913. URL https://pubmed.ncbi.nlm.nih.gov/14527287/.

S. Grünberg, C. Reich, M. E. Zeller, M. S. Bartlett, and M. Thomm. Rearrangement of the RNA polymerase subunit H and the lower jaw in archaeal elongation complexes. *Nucleic Acids Research*, 38(6):1950–1963, 4 2010. ISSN 0305-1048. doi: 10.1093/NAR/GKP1190. URL https://academic.oup.com/nar/article/38/6/1950/3112562.

S. Grünberg, L. Warfield, and S. Hahn. Architecture of the RNA polymerase II preinitiation complex and mechanism of ATP-dependent promoter opening. *Nature Structural & Molecular Biology 2012 19:8*, 19(8):788–796, 7 2012. ISSN 1545-9985. doi: 10.1038/nsmb.2334. URL https://www.nature.com/articles/nsmb.2334.

M. Hantsche and P. Cramer. Conserved RNA polymerase II initiation complex structure. *Current Opinion in Structural Biology*, 47:17–22, 2017. ISSN 1879033X. doi: 10.1016/j.sbi.2017.03.013. URL http://dx.doi.org/10.1016/j.sbi.2017.03.013.

Y. Hari-Gupta, N. Fili, dos Santos, A. W. Cook, R. E. Gough, H. C. Reed, L. Wang, J. Aaron, T. Venit, E. Wait, A. Grosse-Berkenbusch, J. C. M. Gebhardt, P. Percipalle, T. L. Chew, M. Martin-Fernandez, and C. P. Toseland. Myosin VI regulates the spatial organisation of mammalian transcription initiation. *Nature Communications 2022 13:1*, 13(1):1–18, 3 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-28962-w. URL https://www.nature.com/articles/s41467-022-28962-w.

S. P. Haugen, W. Ross, and R. L. Gourse. Advances in bacterial promoter recognition and its control by factors that do not bind DNA. *Nature reviews. Microbiology*, 6 (7):507–519, 7 2008. ISSN 1740-1534. doi: 10.1038/NRMICRO1912. URL https://pubmed.ncbi.nlm.nih.gov/18521075/.

N. K. Hayward, J. S. Wilmott, N. Waddell, P. A. Johansson, M. A. Field, K. Nones, A. M. Patch, H. Kakavand, L. B. Alexandrov, H. Burke, V. Jakrot, S. Kazakoff, O. Holmes, C. Leonard, R. Sabarinathan, L. Mularoni, S. Wood, Q. Xu, N. Waddell, V. Tembe, G. M. Pupo, R. De Paoli-Iseppi, R. E. Vilain, P. Shang, L. M. Lau, R. A. Dagg, S. J. Schramm, A. Pritchard, K. Dutton-Regester, F. Newell, A. Fitzgerald, C. A. Shang, S. M. Grimmond, H. A. Pickett, J. Y. Yang, J. R. Stretch, A. Behren, R. F. Kefford, P. Hersey, G. V. Long, J. Cebon, M. Shackleton, A. J. Spillane, R. P. Saw, N. López-Bigas, J. V. Pearson, J. F. Thompson, R. A. Scolyer, and G. J. Mann. Whole-genome landscapes of major melanoma subtypes. *Nature 2017 545:7653*, 545(7653):175–180, 5

2017. ISSN 1476-4687. doi: 10.1038/nature22071. URL https://www.nature.com/articles/nature22071.

Y. He, J. Fang, D. J. Taatjes, and E. Nogales. Structural visualization of key steps in human transcription initiation. *Nature*, 495(7442):481–486, 3 2013. ISSN 1476-4687. doi: 10.1038/NATURE11991. URL https://pubmed.ncbi.nlm.nih.gov/23446344/.

A. E. Herlihy, S. Boeing, J. C. Weems, J. Walker, A. B. Dirac-Svejstrup, M. H. Lehner, R. C. Conaway, J. W. Conaway, and J. Q. Svejstrup. UBAP2/UBAP2L regulate UV-induced ubiquitylation of RNA polymerase II and are the human orthologues of yeast Def1. *DNA repair*, 115:103343, 7 2022. ISSN 1568-7856. doi: 10.1016/J.DNAREP.2022.103343. URL https://pubmed.ncbi.nlm.nih.gov/35633597/.

A. Hirtreiter, D. Grohmann, and F. Werner. Molecular mechanisms of RNA polymerase—the F/E (RPB4/7) complex is required for high processivity in vitro. *Nucleic Acids Research*, 38(2):585–596, 1 2010. ISSN 0305-1048. doi: 10.1093/NAR/GKP928. URL https://academic.oup.com/nar/article/38/2/585/2409603.

D. Hnisz, B. J. Abraham, T. I. Lee, A. Lau, V. Saint-André, A. A. Sigova, H. A. Hoke, and R. A. Young. Super-enhancers in the control of cell identity and disease. *Cell*, 155(4):934, 11 2013. ISSN 1097-4172. doi: 10.1016/J.CELL.2013.09.053. URL https://pubmed.ncbi.nlm.nih.gov/24119843/.

T. Hoek and e. al. Single-molecule imaging uncovers rules governing nonsense-mediated mRNA decay. *Mol. Cell*, 75:324–339, 2019.

J. P. Hsin and J. L. Manley. The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes & Development*, 26(19):2119, 10 2012. ISSN 08909369. doi: 10.1101/GAD.200303.112. URL /pmc/articles/PMC3465734//pmc/articles/PMC3465734/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC3465734/.

A. L. Hughes, J. R. Kelley, and R. J. Klose. Understanding the interplay between CpG island-associated gene promoters and H3K4 methylation. *Biochimica et Biophysica Acta. Gene Regulatory Mechanisms*, 1863(8), 8 2020. ISSN 18764320. doi: 10.1016/J.BBAGRM.2020.194567. URL /pmc/

articles/PMC7294231//pmc/articles/PMC7294231/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC7294231/.

M. Jenal and e. al. The poly(A)-binding protein nuclear 1 suppresses alternative cleavage and polyadenylation sites. *Cell*, 149:538–553, 2012.

J. T. Kadonaga. Perspectives on the RNA polymerase II core promoter. *Wiley interdisciplinary reviews. Developmental biology*, 1(1):40–51, 1 2012. ISSN 1759-7692. doi: 10.1002/WDEV.21. URL https://pubmed.ncbi.nlm.nih.gov/23801666/.

X. Kang, Y. Hu, Y. Li, X. Guo, X. Jiang, L. Lai, B. Xia, and C. Jin. Structural, biochemical, and dynamic characterizations of the hRPB8 subunit of human RNA polymerases. *The Journal of biological chemistry*, 281(26):18216–18226, 6 2006. ISSN 0021-9258. doi: 10.1074/JBC.M513241200. URL https://pubmed.ncbi.nlm.nih.gov/16632472/.

A. N. Kapanidis, E. Margeat, S. O. Ho, E. Kortkhonjia, S. Weiss, and R. H. Ebright. Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science*, 314(5802):1144–1147, 11 2006. ISSN 00368075. doi: 10.1126/SCIENCE.1131399/SUPPL{_}FILE/KAPANIDIS.SOM.PDF. URL https://www.science.org/doi/abs/10.1126/science.1131399.

T. Kecman and e. al. Elongation/termination factor exchange mediated by PP1 phosphatase orchestrates transcription termination. *Cell Rep.*, 25:259–269, 2018.

H. Kettenberger, K.-j. Armache, and P. Cramer. Structure and Its Interactions with NTP and TFIIS. *Molecular Cell*, 16(6):955–965, 2004.

M. Kim and e. al. The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature*, 432:517–522, 2004.

T. K. Kim, R. H. Ebright, and D. Reinberg. Mechanism of ATP-dependent promoter melting by transcription factor IIH. *Science*, 288(5470):1418–1421, 5 2000. ISSN 00368075. doi: 10.1126/SCIENCE.288.5470.1418/SUPPL{_}FILE/1047530S2{_}THUMB.GIF. URL https://www.science.org/doi/full/10.1126/science.288.5470.1418.

H. Kimura and Y. Sato. Imaging transcription elongation dynamics by new technologies unveils the organization of initiation and elongation in transcription factories. *Current*

Opinion in Cell Biology, 74:71–79, 2022. ISSN 09550674. doi: 10.1016/j.ceb.2022.01.002. URL https://doi.org/10.1016/j.ceb.2022.01.002.

M. Kireeva, M. Kashlev, and Z. F. Burton. Translocation by multi-subunit RNA polymerases. *Biochimica et biophysica acta*, 1799(5-6):389–401, 2010. ISSN 0006-3002. doi: 10.1016/J.BBAGRM.2010.01.007. URL https://pubmed.ncbi.nlm.nih.gov/20097318/.

M. L. Kireeva, Y. A. Nedialkov, G. H. Cremona, Y. A. Purtov, L. Lubkowska, F. Malagon, Z. F. Burton, J. N. Strathern, and M. Kashlev. Transient reversal of RNA polymerase II active site closing controls fidelity of transcription elongation. *Molecular cell*, 30 (5):557–566, 6 2008. ISSN 1097-4164. doi: 10.1016/J.MOLCEL.2008.04.017. URL https://pubmed.ncbi.nlm.nih.gov/18538654/.

W. Kobayashi and H. Kurumizaka. Structural transition of the nucleosome during chromatin remodeling and transcription. *Current Opinion in Structural Biology*, 59:107–114, 12 2019. ISSN 0959-440X. doi: 10.1016/J.SBI.2019.07.011.

E. V. Koonin. Does the central dogma still stand? *Biology Direct*, 7(1):1–7, 8 2012. ISSN 17456150. doi: 10.1186/1745-6150-7-27/COMMENTS. URL https://biologydirect.biomedcentral.com/articles/10.1186/1745-6150-7-27.

Y. Korkhin, U. M. Unligil, O. Littlefield, P. J. Nelson, D. I. Stuart, P. B. Sigler, S. D. Bell, and N. G. Abrescia. Evolution of Complex RNA Polymerases: The Complete Archaeal RNA Polymerase Structure. *PLOS Biology*, 7(5):e1000102, 5 2009. ISSN 1545-7885. doi: 10.1371/JOURNAL.PBIO.1000102. URL https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1000102.

D. Kostrewa, M. E. Zeller, K. J. Armache, M. Seizl, K. Leike, M. Thomm, and P. Cramer. RNA polymerase II–TFIIB structure and mechanism of transcription initiation. *Nature 2009 462:7271*, 462(7271):323–330, 10 2009. ISSN 1476-4687. doi: 10.1038/nature08548. URL https://www.nature.com/articles/nature08548.

J. E. Kucab, X. Zou, S. Morganella, M. Joel, A. S. Nanda, E. Nagy, C. Gomez, A. Degasperi, R. Harris, S. P. Jackson, V. M. Arlt, D. H. Phillips, and S. Nik-Zainal. A Compendium of Mutational Signatures of Environmental Agents.

Cell, 177(4):821–836, 5 2019. ISSN 10974172. doi: 10.1016/J.CELL.2019.03.001/ATTACHMENT/89036770-6861-4BE2-B402-049DE838B1A7/MMC6.PDF. URL http://www.cell.com/article/S0092867419302636/fulltexthttp://www.cell.com/article/S0092867419302636/abstracthttps://www.cell.com/cell/abstract/S0092-8674(19)30263-6.

T. Kujirai, H. Ehara, Y. Fujino, M. Shirouzu, S.-i. Sekine, and H. Kurumizaka. Structural basis of the nucleosome transition during RNA polymerase II passage. *Science*, 9904 (October):eaau9904, 2018. ISSN 0036-8075. doi: 10.1126/science.aau9904. URL http://www.sciencemag.org/lookup/doi/10.1126/science.aau9904.

O. I. Kulaeva, F.-K. Hsieh, H.-W. Chang, D. S. Luse, and V. M. Studitsky. Mechanism of transcription through a Nucleosome by RNA polymerase II. *Biochimica et Biophysica Acta*, 1829(1):76–83, 2013. ISSN 15378276. doi: 10.1038/jid.2014.371.

N. Kumar, A. F. Theil, V. Roginskaya, Y. Ali, M. Calderon, S. C. Watkins, R. P. Barnes, P. L. Opresko, A. Pines, H. Lans, W. Vermeulen, and B. Van Houten. Global and transcription-coupled repair of 8-oxoG is initiated by nucleotide excision repair proteins. *Nature Communications 2022 13:1*, 13(1):1–16, 2 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-28642-9. URL https://www.nature.com/articles/s41467-022-28642-9.

I. Lahiri, J. Xu, B. G. Han, J. Oh, D. Wang, F. DiMaio, and A. E. Leschziner. 3.1 Å structure of yeast RNA polymerase II elongation complex stalled at a cyclobutane pyrimidine dimer lesion solved using streptavidin affinity grids. *Journal of structural biology*, 207(3):270–278, 9 2019. ISSN 1095-8657. doi: 10.1016/J.JSB.2019.06.004. URL https://pubmed.ncbi.nlm.nih.gov/31200019/.

R. Landick. Transcriptional pausing without backtracking. *Proceedings of the National Academy of Sciences of the United States of America*, 106 (22):8797–8798, 6 2009. ISSN 00278424. doi: 10.1073/PNAS.0904373106/ASSET/5F772AA6-3276-4CC5-BC00-310F991865FD/ASSETS/GRAPHIC/ZPQ9990984070001.JPEG. URL www.pnas.orgcgidoi10.1073pnas.0904373106.

H. Lans, J. H. Hoeijmakers, W. Vermeulen, and J. A. Marteijn. The DNA damage response to transcription stress. *Nature Reviews Molecular Cell Biology*, 20(12):766–

784, 2019. ISSN 14710080. doi: 10.1038/s41580-019-0169-4. URL http://dx.doi.org/10.1038/s41580-019-0169-4.

T. T. Le, Y. Yang, C. Tan, M. M. Suhanovsky, R. M. Fulbright, J. T. Inman, M. Li, J. Lee, S. Perelman, J. W. Roberts, A. M. Deaconescu, and M. D. Wang. Mfd Dynamically Regulates Transcription via a Release and Catch-Up Mechanism. *Cell*, 172(1-2):344–357, 1 2018. ISSN 10974172. doi: 10.1016/J.CELL.2017.11.017/ATTACHMENT/F59041D7-6C3D-450F-A8E2-F2ECC5D3CDF1/MMC1.PDF. URL http://www.cell.com/article/S0092867417313648/fulltexthttp://www.cell.com/article/S0092867417313648/abstracthttps://www.cell.com/cell/abstract/S0092-8674(17)31364-8.

C.-Y. Lee and S. Myong. Probing steps in DNA transcription using single-molecule methods. *J. Biol. Chem*, pages 297–300, 2021. doi: 10.1016/j.jbc.2021.101086. URL http://creativecommons.org/licenses/by/4.0/.

C. L. Li, F. M. Golebiowski, Y. Onishi, N. L. Samara, K. Sugasawa, and W. Yang. Tripartite DNA Lesion Recognition and Verification by XPC, TFIIH, and XPA in Nucleotide Excision Repair. *Molecular cell*, 59(6):1025, 9 2015. ISSN 10974164. doi: 10.1016/J.MOLCEL.2015.08.012. URL /pmc/articles/PMC4617536//pmc/articles/PMC4617536/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4617536/.

J. Li, A. Dong, K. Saydaminova, H. Chang, G. Wang, H. Ochiai, T. Yamamoto, and A. Pertsinidis. Single-Molecule Nanoscopy Elucidates RNA Polymerase II Transcription at Single Genes in Live Cells. *Cell*, pages 1–16, 2019. ISSN 00928674. doi: 10.1016/j.cell.2019.05.029. URL https://linkinghub.elsevier.com/retrieve/pii/S0092867419305574.

T. Lindahl. Instability and decay of the primary structure of DNA. *Nature*, 362(6422):709–715, 1993. ISSN 1476-4687. doi: 10.1038/362709a0. URL https://www.nature.com/articles/362709a0.

J. Logan and e. al. A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse beta maj-globin gene. *Proc. Natl. Acad. Sci. U. S. A.*, 84:8306–8310, 1987.

R. L. Maher and L. B. Bloom. Pre-steady-state kinetic characterization of the AP endonuclease activity of human AP endonuclease 1. *The Journal of biological chemistry*, 282(42):30577–30585, 10 2007. ISSN 0021-9258. doi: 10.1074/JBC.M704341200. URL https://pubmed.ncbi.nlm.nih.gov/17724035/.

P. Maiuri, A. Knezevich, A. De Marco, D. Mazza, A. Kula, J. G. McNally, and A. Marcello. Fast transcription rates of RNA polymerase II in human cells. *EMBO Reports*, 12 (12):1280–1285, 2011. ISSN 1469221X. doi: 10.1038/embor.2011.196. URL http://dx.doi.org/10.1038/embor.2011.196.

F. W. Martinez-Rucobo and P. Cramer. Structural basis of transcription elongation. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1829(1):9–19, 2013. ISSN 18749399. doi: 10.1016/j.bbagrm.2012.09.002. URL http://dx.doi.org/10.1016/j.bbagrm.2012.09.002.

A. G. Matera, R. M. Terns, and M. P. Terns. Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nature Reviews Molecular Cell Biology 2007 8:3*, 8(3):209–220, 3 2007. ISSN 1471-0080. doi: 10.1038/nrm2124. URL https://www.nature.com/articles/nrm2124.

R. Mathew and D. Chatterji. The evolving story of the omega subunit of bacterial RNA polymerase. *Trends in microbiology*, 14(10):450–455, 10 2006. ISSN 0966-842X. doi: 10.1016/J.TIM.2006.08.002. URL https://pubmed.ncbi.nlm.nih.gov/16908155/.

A. Mazumder, M. Lin, A. N. Kapanidis, and R. H. Ebright. Closing and opening of the RNA polymerase trigger loop. *Proceedings of the National Academy of Sciences of the United States of America*, 117(27):15642–15649, 7 2020. ISSN 10916490. doi: 10.1073/PNAS.1920427117/SUPPL{_}FILE/PNAS.1920427117.SAPP.PDF. URL https://www.pnas.org/doi/abs/10.1073/pnas.1920427117.

I. Mellon and P. C. Hanawalt. Induction of the Escherichia coli lactose operon selectively increases repair of its transcribed DNA strand. *Nature 1989 342:6245*, 342(6245):95–98, 1989. ISSN 1476-4687. doi: 10.1038/342095a0. URL https://www.nature.com/articles/342095a0.

C. D. Mol, A. S. Arvai, G. Slupphaug, B. Kavli, I. Alseth, H. E. Krokan, and J. A. Tainer. Crystal structure and mutational analysis of human uracil-DNA glycosylase: structural

basis for specificity and catalysis. *Cell*, 80(6):869–878, 3 1995. ISSN 0092-8674. doi: 10.1016/0092-8674(95)90290-2. URL https://pubmed.ncbi.nlm.nih.gov/7697717/.

C. D. Mol, T. Izumi, S. Mitra, and J. A. Talner. DNA-bound structures and mutants reveal abasic DNA binding by APE1 DNA repair and coordination. *Nature 2000 403:6768*, 403(6768):451–456, 1 2000. ISSN 1476-4687. doi: 10.1038/35000249. URL https://www.nature.com/articles/35000249.

K. S. Murakami. Structural Biology of Bacterial RNA Polymerase. *Biomolecules*, 5(2):848, 5 2015. ISSN 2218273X. doi: 10.3390/BIOM5020848. URL /pmc/articles/PMC4496699//pmc/articles/PMC4496699/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4496699/.

K. S. Murakami and S. A. Darst. Bacterial RNA polymerases: the wholo story. *Current opinion in structural biology*, 13(1):31–39, 2 2003. ISSN 0959-440X. doi: 10.1016/S0959-440X(02)00005-2. URL https://pubmed.ncbi.nlm.nih.gov/12581657/.

S. Nik-Zainal, J. E. Kucab, S. Morganella, D. Glodzik, L. B. Alexandrov, V. M. Arlt, A. Weninger, M. Hollstein, M. R. Stratton, and D. H. Phillips. The genome as a record of environmental exposure. *Mutagenesis*, 30(6):763–770, 11 2015. ISSN 0267-8357. doi: 10.1093/MUTAGE/GEV073. URL https://academic.oup.com/mutage/article/30/6/763/2622786.

M. Noe Gonzalez, D. Blears, and J. Q. Svejstrup. Causes and consequences of RNA polymerase II stalling during transcript elongation. *Nature Reviews Molecular Cell Biology*, 22, 2021. doi: 10.1038/s41580-020-00308-8. URL www.nature.com/nrm.

E. Nogales, R. K. Louder, and Y. He. Structural Insights into the Eukaryotic Transcription Initiation Machinery. *Annual Review of Biophysics*, 46(1):59–83, 2017. ISSN 1936-122X. doi: 10.1146/annurev-biophys-070816-033751. URL http://www.annualreviews.org/doi/10.1146/annurev-biophys-070816-033751.

T. Nojima and e. al. Deregulated expression of mammalian lncRNA through loss of SPT6 induces R-loop formation, replication stress, and cellular senescence. *Mol. Cell*, 72:970–984, 2018.

J. Oh, J. Xu, J. Chong, and D. Wang. Structural and biochemical analysis of DNA lesion-induced RNA polymerase II arrest. *Methods*, (October 2018):0–1, 2019. ISSN 10462023. doi: 10.1016/j.ymeth.2019.02.019. URL `https://linkinghub.elsevier.com/retrieve/pii/S1046202318303001`.

J. Oh, A. M. Fleming, J. Xu, J. Chong, C. J. Burrows, and D. Wang. RNA polymerase II stalls on oxidative DNA damage via a torsion-latch mechanism involving lone pair–π and CH–π interactions. *Proceedings of the National Academy of Sciences of the United States of America*, 117(17):9338–9348, 2020. ISSN 10916490. doi: 10.1073/pnas.1919904117.

S. M. Orlicky, P. T. Tran, M. H. Sayre, and A. M. Edwards. Dissociable Rpb4-Rpb7 Subassembly of RNA Polymerase II Binds to Single-strand Nucleic Acid and Mediates a Post-recruitment Step in Transcription Initiation*. *THE JOURNAL OF BIOLOGICAL CHEMISTRY*, 276(13):10097–10102, 2000. doi: 10.1074/jbc.M003165200. URL `http://www.jbc.org`.

Y. Osheim and e. al. EM visualization of Pol II genes in Drosophila: most genes terminate without prior 3' end cleavage of nascent transcripts. *Chromosoma*, 111:1–12, 2002.

S. Osman and P. Cramer. Structural Biology of RNA Polymerase II Transcription: 20 Years on. *Annual Review of Cell and Developmental Biology*, 36:1–34, 2020. ISSN 15308995. doi: 10.1146/annurev-cellbio-042020-021954.

V. Pagès, R. E. Johnson, L. Prakash, and S. Prakash. Mutational specificity and genetic control of replicative bypass of an abasic site in yeast. *Proceedings of the National Academy of Sciences of the United States of America*, 105(4):1170–1175, 2008. ISSN 10916490. doi: 10.1073/pnas.0711227105.

S. S. Parikh, G. Walcher, G. D. Jones, G. Slupphaug, H. E. Krokan, G. M. Blackburn, and J. A. Tainer. Uracil-DNA glycosylase-DNA substrate and product structures: conformational strain promotes catalytic efficiency by coupled stereoelectronic effects. *Proceedings of the National Academy of Sciences of the United States of America*, 97(10):5083–5088, 5 2000. ISSN 0027-8424. doi: 10.1073/PNAS.97.10.5083. URL `https://pubmed.ncbi.nlm.nih.gov/10805771/`.

H. J. Park, K. Zhang, Y. Ren, S. Nadji, N. Sinha, J. S. Taylor, and C. H. Kang. Crystal structure of a DNA decamer containing a cis-syn thymine dimer. *Proceedings of the National Academy of Sciences of the United States of America*, 99(25):15965, 12 2002. ISSN 00278424. doi: 10.1073/PNAS.242422699. URL /pmc/articles/PMC138548//pmc/articles/PMC138548/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC138548/.

S. C. Parker, M. L. Stitzel, D. L. Taylor, J. M. Orozco, M. R. Erdos, J. A. Akiyama, K. L. Van Bueren, P. S. Chines, N. Narisu, B. L. Black, V. Axel, L. A. Pennacchio, and F. S. Collins. Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proceedings of the National Academy of Sciences of the United States of America*, 110(44):17921–17926, 10 2013. ISSN 00278424. doi: 10.1073/PNAS.1317023110.

P. Parua and e. al. A Cdk9-PP1 switch regulates the elongation-termination transition of RNA polymerase II. *Nature*, 558:460–464, 2018.

N. J. Proudfoot. Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. *Science (New York, N.Y.)*, 352(6291):aad9926, 6 2016. ISSN 10959203. doi: 10.1126/SCIENCE.AAD9926. URL /pmc/articles/PMC5144996//pmc/articles/PMC5144996/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC5144996/.

M. Rajewska, K. Wegrzyn, and I. Konieczny. AT-rich region and repeated sequences – the essential elements of replication origins of bacterial replicons. *FEMS Microbiology Reviews*, 36(2):408–434, 3 2012. ISSN 0168-6445. doi: 10.1111/J.1574-6976.2011.00300.X. URL https://academic.oup.com/femsre/article/36/2/408/565738.

A. Ray-Soni and e. al. Mechanisms of bacterial transcription termination: all good things must end. *Annu. Rev. Biochem.*, 85:319–347, 2016.

A. Revyakin, C. Liu, R. H. Ebright, and T. R. Strick. Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science (New York, N.Y.)*, 314 (5802):1139–1143, 11 2006. ISSN 1095-9203. doi: 10.1126/SCIENCE.1131398. URL https://pubmed.ncbi.nlm.nih.gov/17110577/.

A. Revyakin, Z. Zhang, R. A. Coleman, Y. Li, C. Inouye, J. K. Lucas, S. R. Park, S. Chu, and R. Tjian. Transcription initiation by human RNA polymerase II visualized at single-molecule resolution. *Genes & development*, 26(15):1691–1702, 8 2012. ISSN 1549-5477. doi: 10.1101/GAD.194936.112. URL https://pubmed.ncbi.nlm.nih.gov/22810624/.

R. G. Roeder and W. J. Rutter. Multiple Forms of DNA-dependent RNA Polymerase in Eukaryotic Organisms. *Nature 1969 224:5216*, 224(5216):234–237, 1969. ISSN 1476-4687. doi: 10.1038/224234a0. URL https://www.nature.com/articles/224234a0.

S. Sainsbury, C. Bernecky, and P. Cramer. Structural basis of transcription initiation by RNA polymerase II. *Nature Reviews Molecular Cell Biology*, 16(3):129–143, 2015. ISSN 14710080. doi: 10.1038/nrm3952. URL http://dx.doi.org/10.1038/nrm3952.

M. Saponaro, T. Kantidakis, R. Mitter, G. P. Kelly, M. Heron, H. Williams, J. Söding, A. Stewart, and J. Q. Svejstrup. RECQL5 controls transcript elongation and suppresses genome instability associated with transcription stress. *Cell*, 157(5):1037–1049, 2014. ISSN 10974172. doi: 10.1016/j.cell.2014.03.048.

K. M. Schermerhorn and S. Delaney. Transient-state kinetics of apurinic/apyrimidinic (AP) endonuclease 1 acting on an authentic AP site and commonly used substrate analogs: The effect of diverse metal ions and base mismatches. *Biochemistry*, 52 (43):7669–7677, 10 2013. ISSN 00062960. doi: 10.1021/BI401218R/SUPPL{_}FILE/BI401218R{_}SI{_}001.PDF. URL https://pubs.acs.org/doi/full/10.1021/bi401218r.

S. Schilbach, M. Hantsche, D. Tegunov, C. Dienemann, C. Wigge, H. Urlaub, and P. Cramer. Structures of transcription pre-initiation complex with TFIIH and Mediator. *Nature 2017 551:7679*, 551(7679):204–209, 11 2017. ISSN 1476-4687. doi: 10.1038/nature24282. URL https://www.nature.com/articles/nature24282.

S. Schoenfelder and P. Fraser. Long-range enhancer–promoter contacts in gene expression control. *Nature Reviews Genetics 2019 20:8*, 20(8):437–455, 5 2019. ISSN 1471-0064. doi: 10.1038/s41576-019-0128-0. URL https://www.nature.com/articles/s41576-019-0128-0.

A. Schreieck and e. al. RNA polymerase II termination involves C-terminal-domain tyrosine dephosphorylation by CPF subunit Glc7. *Nat. Struct. Mol. Biol.*, 21:175–179, 2014.

E. Serfling, M. Jasin, and W. Schaffner. Enhancers and eukaryotic gene transcription. *Trends in Genetics*, 1(C):224–230, 1 1985. ISSN 0168-9525. doi: 10.1016/0168-9525(85)90088-5.

Y. Shi and e. al. Molecular architecture of the human pre-mRNA 3' processing complex. *Mol. Cell*, 33:365–376, 2009.

G. Shino and S. Takada. Modeling DNA Opening in the Eukaryotic Transcription Initiation Complexes via Coarse-Grained Models. *Frontiers in Molecular Biosciences*, 8:1021, 11 2021. ISSN 2296889X. doi: 10.3389/FMOLB.2021.772486/BIBTEX.

A. Shundrovsky, T. J. Santangelo, J. W. Roberts, and M. D. Wang. A Single-Molecule Technique to Study Sequence-Dependent Transcription Pausing. *Biophysical Journal*, 87(6):3945, 2004. ISSN 00063495. doi: 10.1529/BIOPHYSJ.104.044081. URL `/pmc/articles/PMC1304905//pmc/articles/PMC1304905/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC1304905/`.

S. Sigurdsson, A. B. Dirac-Svejstrup, and J. Q. Svejstrup. Evidence that Transcript Cleavage Is Essential for RNA Polymerase II Transcription and Cell Viability. *Molecular Cell*, 38(2):202–210, 2010. ISSN 10972765. doi: 10.1016/j.molcel.2010.02.026. URL `http://dx.doi.org/10.1016/j.molcel.2010.02.026`.

S. T. Smale and J. T. Kadonaga. The RNA Polymerase II Core Promoter. *Annual Review of Biochemistry*, 72(1):449–479, 2003. ISSN 0066-4154. doi: 10.1146/annurev.biochem.72.121801.161520. URL `http://www.annualreviews.org/doi/10.1146/annurev.biochem.72.121801.161520`.

M. J. Smerdon and F. Thoma. Site-specific DNA repair at the nucleosome level in a yeast minichromosome. *Cell*, 61(4):675–684, 5 1990. ISSN 0092-8674. doi: 10.1016/0092-8674(90)90479-X. URL `https://pubmed.ncbi.nlm.nih.gov/2188732/`.

V. Sosunov, E. Sosunova, A. Mustaev, I. Bass, V. Nikiforov, and A. Goldfarb. Unified two-metal mechanism of RNA synthesis and degradation by RNA polymerase.

The EMBO Journal, 22(9):2234, 5 2003. ISSN 02614189. doi: 10.1093/EMBOJ/CDG193. URL /pmc/articles/PMC156065//pmc/articles/PMC156065/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC156065/.

L. Statello, C. J. Guo, L. L. Chen, and M. Huarte. Gene regulation by long non-coding RNAs and its biological functions. *Nature Reviews Molecular Cell Biology 2020 22:2*, 22 (2):96–118, 12 2020. ISSN 1471-0080. doi: 10.1038/s41580-020-00315-9. URL https://www.nature.com/articles/s41580-020-00315-9.

T. A. Steitz. The structural changes of T7 RNA polymerase from transcription initiation to elongation. *Current Opinion in Structural Biology*, 19(6):683–690, 12 2009. ISSN 0959-440X. doi: 10.1016/J.SBI.2009.09.001.

K. M. Stevens, J. B. Swadling, A. Hocher, C. Bang, S. Gribaldo, R. A. Schmitz, and T. Warnecke. Histone variants in archaea and the evolution of combinatorial chromatin complexity. *Proceedings of the National Academy of Sciences of the United States of America*, 117(52):33384–33395, 12 2020. ISSN 10916490. doi: 10.1073/PNAS.2007056117. URL www.pnas.org/cgi/doi/10.1073/pnas.2007056117.

J. Svejstrup, P. Vichi, and J. Egly. The multiple roles of transcription/repair factor TFIIH. *Trends in Biochemical Sciences*, 21(9):346–350, 9 1996. ISSN 0968-0004. doi: 10.1016/S0968-0004(96)10046-3.

J. A. Swenberg, K. Lu, B. C. Moeller, L. Gao, P. B. Upton, J. Nakamura, and T. B. Starr. Endogenous versus Exogenous DNA Adducts: Their Role in Carcinogenesis, Epidemiology, and Risk Assessment. *Toxicological Sciences*, 120(Suppl 1):S130, 2011. ISSN 10966080. doi: 10.1093/TOXSCI/KFQ371. URL /pmc/articles/PMC3043087//pmc/articles/PMC3043087/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC3043087/.

J. F. Sydow, F. Brueckner, A. C. Cheung, G. E. Damsma, S. Dengl, E. Lehmann, D. Vassylyev, and P. Cramer. Structural Basis of Transcription: Mismatch-Specific Fidelity Mechanisms and Paused RNA Polymerase II with Frayed RNA. *Molecular Cell*, 34(6):710–721, 6 2009. ISSN 10972765. doi: 10.1016/J.MOLCEL.2009.06.002/ATTACHMENT/B6DE46F1-51FD-4ABF-B63B-E94C056D6CEB/MMC1.PDF. URL http://www.cell.com/article/S1097276509003931/fulltexthttp:

//www.cell.com/article/S1097276509003931/abstracthttps://www.cell.com/molecular-cell/abstract/S1097-2765(09)00393-1.

S. S. Teves, C. M. Weber, and S. Henikoff. Transcribing through the nucleosome. *Trends in Biochemical Sciences*, 39(12):577–586, 2014. ISSN 13624326. doi: 10.1016/j.tibs.2014.10.004. URL http://dx.doi.org/10.1016/j.tibs.2014.10.004.

N. Thompson, D. B. Aronson, and R. Burgess. Purification of Eukaryotic Chromatography II by Immunoaffinity. *Journal of Biological Chemistry*, 265(12):7069–7077, 1990.

P. S. Thompson and D. Cortez. New Insights into Abasic Site Repair and Tolerance. *DNA repair*, 90:102866, 6 2020. ISSN 15687856. doi: 10.1016/J.DNAREP.2020.102866. URL /pmc/articles/PMC7299775//pmc/articles/PMC7299775/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC7299775/.

F. Tirode, D. Busso, F. Coin, and J. M. Egly. Reconstitution of the transcription factor TFIIH: assignment of functions for the three enzymatic subunits, XPB, XPD, and cdk7. *Molecular cell*, 3(1):87–95, 1999. ISSN 1097-2765. doi: 10.1016/S1097-2765(00)80177-X. URL https://pubmed.ncbi.nlm.nih.gov/10024882/.

K. L. Tsai, X. Yu, S. Gopalan, T. C. Chao, Y. Zhang, L. Florens, M. P. Washburn, K. Murakami, R. C. Conaway, J. W. Conaway, and F. J. Asturias. Mediator structure and rearrangements required for holoenzyme formation. *Nature*, 544(7649):196–201, 3 2017. ISSN 1476-4687. doi: 10.1038/nature21393. URL https://www.nature.com/articles/nature21393.

V. Van Den Boom, E. Citterio, D. Hoogstraten, A. Zotter, J. M. Egly, W. A. Van Cappellen, J. H. Hoeijmakers, A. B. Houtsmuller, and W. Vermeulen. DNA damage stabilizes interaction of CSB with the transcription elongation machinery. *The Journal of cell biology*, 166(1):27–36, 7 2004. ISSN 0021-9525. doi: 10.1083/JCB.200401056. URL https://pubmed.ncbi.nlm.nih.gov/15226310/.

A. Vannini and P. Cramer. Conservation between the RNA Polymerase I, II, and III Transcription Initiation Machineries. *Molecular Cell*, 45(4):439–446, 2 2012. ISSN 1097-2765. doi: 10.1016/J.MOLCEL.2012.01.023. URL http://www.cell.com/article/S1097276512000895/fulltexthttp:

//www.cell.com/article/S1097276512000895/abstracthttps://www.cell.com/molecular-cell/abstract/S1097-2765(12)00089-5.

D. G. Vassylyev, M. N. Vassylyeva, J. Zhang, M. Palangat, I. Artsimovitch, and R. Landick. Structural basis for substrate loading in bacterial RNA polymerase. *Nature*, 448(7150):163–168, 7 2007. ISSN 1476-4687. doi: 10.1038/NATURE05931. URL https://pubmed.ncbi.nlm.nih.gov/17581591/.

S. M. Vos, L. Farnung, M. Boehning, C. Wigge, A. Linden, H. Urlaub, and P. Cramer. Structure of activated transcription complex Pol II–DSIF–PAF–SPT6. *Nature*, 560 (607-612), 2018a. ISSN 0028-0836. doi: 10.1038/s41586-018-0440-4. URL http://www.nature.com/articles/s41586-018-0440-4.

S. M. Vos, L. Farnung, H. Urlaub, and P. Cramer. Structure of paused transcription complex Pol II–DSIF–NELF. *Nature*, 560:601–606, 2018b. ISSN 0028-0836. doi: 10.1038/s41586-018-0442-2. URL http://www.nature.com/articles/s41586-018-0442-2.

S. M. Vos, L. Farnung, A. Linden, H. Urlaub, and P. Cramer. Structure of complete Pol II–DSIF–PAF–SPT6 transcription complex reveals RTF1 allosteric activation. *Nature Structural & Molecular Biology 2020 27:7*, 27(7):668–677, 6 2020. ISSN 1545-9985. doi: 10.1038/s41594-020-0437-1. URL https://www.nature.com/articles/s41594-020-0437-1.

C. Walmacq, L. Wang, J. Chong, K. Scibelli, L. Lubkowska, A. Gnatt, P. J. Brooks, D. Wang, and M. Kashlev. Mechanism of RNA polymerase II bypass of oxidative cyclopurine DNA lesions. *Proceedings of the National Academy of Sciences of the United States of America*, 112(5):E410–E419, 2015. ISSN 10916490. doi: 10.1073/pnas.1415186112.

N. G. Walter and D. R. Engelke. Ribozymes: Catalytic RNAs that cut things, make things, and do odd and useful jobs. *Biologist (London, England)*, 49(5):199, 10 2002. ISSN 00063347. URL /pmc/articles/PMC3770912//pmc/articles/PMC3770912/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC3770912/.

D. Wang, D. A. Bushnell, K. D. Westover, C. D. Kaplan, and R. D. Kornberg. Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. *Cell*, 127(5):941, 12 2006. ISSN 00928674. doi: 10.1016/J.CELL.2006.

11.023. URL /pmc/articles/PMC1876690//pmc/articles/PMC1876690/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC1876690/.

W. Wang, C. Walmacq, J. Chong, M. Kashlev, and D. Wang. Structural basis of transcriptional stalling and bypass of abasic DNA lesion by RNA polymerase II. *Proceedings of the National Academy of Sciences of the United States of America*, 115(11):E2538–E2545, 2018a. ISSN 1091-6490. doi: 10.1073/pnas.1722050115. URL http://www.ncbi.nlm.nih.gov/pubmed/29487211http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5856558.

W. Wang, J. Xu, J. Chong, and D. Wang. Structural basis of DNA lesion recognition for eukaryotic transcription-coupled nucleotide excision repair. *DNA repair*, 71:43–55, 11 2018b. ISSN 1568-7856. doi: 10.1016/J.DNAREP.2018.08.006. URL https://pubmed.ncbi.nlm.nih.gov/30174298/.

A. Weiss, B. D. Moore, M. H. Tremblay, D. Chaput, A. Kremer, and L. N. Shaw. The ω subunit governs RNA polymerase stability and transcriptional specificity in Staphylococcus aureus. *Journal of Bacteriology*, 199(2), 10 2017. ISSN 10985530. doi: 10.1128/JB.00459-16/SUPPL{_}FILE/ZJB999094298S2.PDF. URL https://journals.asm.org/doi/full/10.1128/JB.00459-16.

S. B. Weiss and L. Gladstone. A Mammalian System for The Incorporation of Cytidine Triphosphate into Ribonucleic Acid. *Journal of the American Chemical Society*, 81(15):4118–4119, 1959. ISSN 15205126. doi: 10.1021/JA01524A087/ASSET/JA01524A087.FP.PNG{_}V03. URL https://pubs.acs.org/doi/abs/10.1021/ja01524a087.

F. Werner. Structure and function of archaeal RNA polymerases. *Molecular microbiology*, 65(6):1395–1404, 9 2007. ISSN 0950-382X. doi: 10.1111/J.1365-2958.2007.05876.X. URL https://pubmed.ncbi.nlm.nih.gov/17697097/https://pubmed.ncbi.nlm.nih.gov/17697097/?dopt=Abstract.

F. Werner and D. Grohmann. Evolution of multisubunit RNA polymerases in the three domains of life. *Nature Reviews Microbiology 2011 9:2*, 9(2):85–98, 1 2011. ISSN 1740-1534. doi: 10.1038/nrmicro2507. URL https://www.nature.com/articles/nrmicro2507.

S. West, N. Gromak, and N. J. Proudfoot. Human 5' → 3' exonuclease Xm2 promotes transcription termination at co-transcriptional cleavage sites. *Nature*, 432(7016):522–525, 2004. ISSN 00280836. doi: 10.1038/nature03035.

O. Willhoft, M. Ghoneim, C. L. Lin, E. Y. Chua, M. Wilkinson, Y. Chaban, R. Ayala, E. A. McCormack, L. Ocloo, D. S. Rueda, and D. B. Wigley. Structure and dynamics of the yeast SWR1-nucleosome complex. *Science*, 362(6411), 10 2018. ISSN 10959203. doi: 10.1126/SCIENCE.AAT7716/SUPPL{_}FILE/AAT7716-WILLHOFT-SM.PDF. URL https://www.science.org/doi/abs/10.1126/science.aat7716.

J. Xu, I. Lahiri, W. Wang, A. Wier, M. A. Cianfrocco, J. Chong, A. A. Hare, P. B. Dervan, F. DiMaio, A. E. Leschziner, and D. Wang. Structural basis for the initiation of eukaryotic transcription-coupled DNA repair. *Nature*, 551(7682):653–657, 2017a. ISSN 14764687. doi: 10.1038/nature24658.

Y. Xu, C. Bernecky, C. T. Lee, K. C. Maier, B. Schwalb, D. Tegunov, J. M. Plitzko, H. Urlaub, and P. Cramer. Architecture of the RNA polymerase II-Paf1C-TFIIS transcription elongation complex. *Nature communications*, 8, 6 2017b. ISSN 2041-1723. doi: 10.1038/NCOMMS15741. URL https://pubmed.ncbi.nlm.nih.gov/28585565/.

Y. W. Yin and T. A. Steitz. The structural mechanism of translocation and helicase activity in T7 RNA polymerase. *Cell*, 116(3):393–404, 2 2004. ISSN 0092-8674. doi: 10.1016/S0092-8674(04)00120-5. URL https://pubmed.ncbi.nlm.nih.gov/15016374/.

M. Yu, W. Yang, T. Ni, Z. Tang, T. Nakadai, J. Zhu, and R. G. Roeder. RNA polymerase II-associated factor 1 regulates the release and phosphorylation of paused RNA polymerase II. *Science (New York, N.Y.)*, 350(6266):1383–1386, 12 2015. ISSN 1095-9203. doi: 10.1126/SCIENCE.AAD2338. URL https://pubmed.ncbi.nlm.nih.gov/26659056/.

N. Yudkovsky, J. A. Ranish, and S. Hahn. A transcription reinitiation intermediate that is stabilized by activator. *Nature*, 408(6809):225–229, 11 2000. ISSN 0028-0836. doi: 10.1038/35041603. URL https://pubmed.ncbi.nlm.nih.gov/11089979/.

D. Zatreanu, Z. Han, R. Mitter, E. Tumini, H. Williams, L. Gregersen, A. B. Diracsvejstrup, S. Roma, A. Stewart, A. Aguilera, and J. Svejstrup. Elongation Factor

TFIIS Prevents Transcription Stress and R-Loop Accumulation to Maintain Genome Stability. *Molecular Cell*, 76:1–13, 2019. doi: 10.2139/ssrn.3283710.

G. Zhang, E. A. Campbell, L. Minakhin, C. Richter, K. Severinov, and S. A. Darst. Crystal Structure of Thermus aquaticus Core RNA Polymerase at 3.3 Å Resolution. *Cell*, 98(6):811–824, 9 1999. ISSN 0092-8674. doi: 10.1016/S0092-8674(00)81515-9.

H. Zhang and e. al. Poly(A) Signal-dependent transcription termination occurs through a conformational change mechanism that does not require cleavage at the Poly(A) site. *Mol. Cell*, 59:437–448, 2015.

J. Zhou, V. Schweikhard, and S. M. Block. Single-molecule studies of RNAPII elongation. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1829(1):29–38, 2013. ISSN 18749399. doi: 10.1016/j.bbagrm.2012.08.006. URL `http://dx.doi.org/10.1016/j.bbagrm.2012.08.006`.

Q. Zhou, T. Li, and D. H. Price. RNA Polymerase II Elongation Control. *Annual Review of Biochemistry*, 81(1):119–143, 2012. ISSN 0066-4154. doi: 10.1146/annurev-biochem-052610-095910. URL `http://www.annualreviews.org/doi/10.1146/annurev-biochem-052610-095910`.

Chapter 2

Background Theory

2.1 Single-Molecule Imaging

Cellular processes are highly dynamic and complex, often involving multiple proteins and pathways. Studying such dynamics can be difficult to monitor, and experimental approaches have often utilised a combination of *in vitro* and *in vivo* approaches. One such technique is fluorescence microscopy. Fluorescence microscopy allows the visualisation of biomolecules and large cellular structures, including DNA and proteins, in either a cellular or *in vitro* setting by using fluorescently tagged or stained molecules. Single-molecule imaging is a subset of this technique where single biomolecules are monitored, often simultaneously, to provide greater detail of their dynamic behaviour (Valeur and Berberan-Santos, 2013). A key benefit of single-molecule imaging over other ensemble biochemical *in vitro* techniques, including electrophoresis and fluorescence polarisation, is the ability to capture the heterogeneity of a sample rather than obtaining an averaged response which can often be misinterpreted. For example, a population of molecules with a single conformation will give rise to a similar result to a population with at least two different conformations where some are more active than others. Where ensemble methods cannot identify these differences, single-molecule imaging can. Single-molecule imaging relies on the principle of monitoring a single molecule (often fluorescently tagged) over a time period with a high signal-to-noise ratio (SNR). In order to capture and track single-molecule dynamics with high accuracy *in vitro*, fluorescence-based techniques are often utilised (Section 2.2). By utilising a combination of TIRFM (Total Internal Reflection Fluorescence Microscopy) (Section 2.2.1) and smFRET (single-molecule Forster Resonance Energy Transfer) (2.2.2) we are able to achieve this.

Other single-molecule techniques that are often used are force-based, utilising optical and magnetic tweezers (Section 2.3). Biomolecules such as DNA, RNA and proteins can be

trapped between microspheres and manipulated by stretching the molecule to study its conformation. Furthermore, in combination with fluorescent techniques (2.3.1) molecules translocating along DNA can be tracked. The structure of DNA can also be modified easily with magnetic tweezers, often inducing supercoiling to create a more biologically relevant substrates.

Finally, cryogenic electron microscopy (Cryo-EM) can be used to obtain high resolution structures of single-particles. Although this does often involve averaging of tens of thousands of single-particles, it can often lead to identification of different conformations. However, in order to study the transitions between the identified conformations, smFRET is often utilised in combination (Willhoft et al., 2018).

In this project I predominately focus on fluorescence based techniques including TIRF and smFRET which I will discuss in more detail below. In Chapter 6, I will discuss some other assays I developed which utilise optical trapping, as described briefly below.

2.2 Single-Molecule TIRF Microscopy

Single-Molecule imaging requires a high signal-to-noise ratio in order to detect the emission from a single fluorophore. Possibly the simplest way to achieve this is by increasing the intensity of the fluorescent signal. One of the first attempts at achieving this was labelling a single antibody with multiple fluorophores in order to increase the fluorescence signal above the background noise (Hirschfeld, 1976). The second way to achieve a high SNR is by lowering the background noise significantly, which can be obtained by total internal reflection fluorescence microscopy (TIRFM) (Section 2.2.1). An example of this was the monitoring of ATP turnover by myosin molecules (Funatsu et al., 1995).

Furthermore, by combining TIRF with smFRET (section 2.2.2), it was shown that the interaction between two molecules could be observed, and later used to monitor conformational changes within biomolecules including ribozymes (Ha et al., 1999a,b; Zhuang et al., 2000, 2002). The combination of these two techniques have been shown to be extremely powerful since its inception, being used to study multiple biological processes such as RNA secondary structure (Rueda and Walter, 2005; Zhao and Rueda, 2009), Splicing(Steiner et al., 2008), DNA replication(Christian et al., 2009; Hohlbein et al., 2013; Gahlon et al., 2017), nucleosome chaperones(Willhoft et al., 2018), CRISPR activity (Newton et al., 2019), translation (Munro et al., 2007; Chen et al., 2011) and transcription which is the focus of my studies (Kapanidis et al., 2006; Andrecka et al., 2008; Robb et al., 2013; Tomescu et al., 2014; Mazumder et al., 2021; Malinen et al., 2022).

2.2.1 Principles of Total Internal Reflection Microscopy

Visualisation of fluorescence requires the generation of a fluorescent signal via excitation with a light source and the mode of excitation can be altered depending on what you want to visualise. Widefield, or brightfield microscopy, is one of the simplest forms of microscopy where a sample is exposed to a white light source, illuminating the entire sample (Valeur and Berberan-Santos, 2013). Confocal excitation is similar but rather than using a white light, a laser of a specific wavelength is used and hence a narrower beam of light is passed on to the sample, exciting a smaller region than traditional widefield (Valeur and Berberan-Santos, 2013). However, both of these methods traditionally cause a low SNR, and observation of single-molecules can be challenging without the use of super-resolution techniques. TIRFM is a method of fluorescence excitation that allows for reduced background fluorescence, while maintaining the high signal that can be achieved

using other microscopy techniques (Axelrod et al., 1984). This is achieved by immobilising fluorescently labelled molecules of interest to a microscope slide surface and selectively exciting them. This method had been widely used to image a variety of cellular processes, but more recently has been coupled with smFRET to monitor these processes on a surface due to the enhanced signal-to-noise ratio achieved (Figure 2.1).

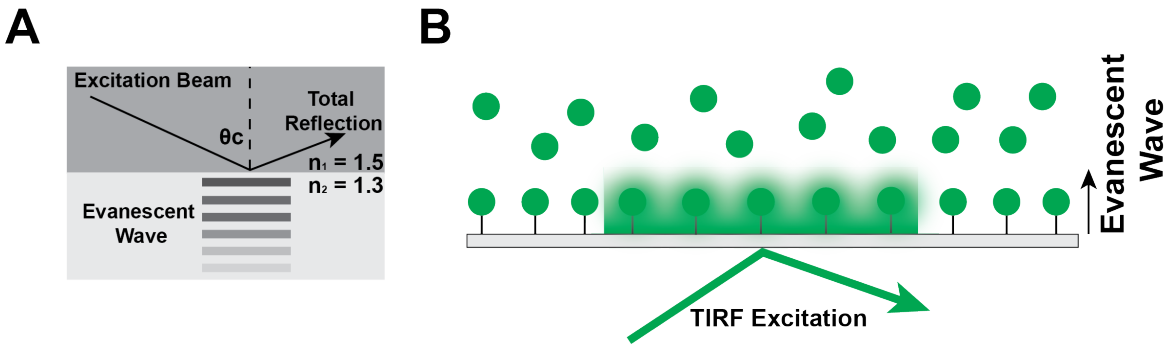


Figure 2.1: **Principles of TIRFM.** (A) Excitation beam is refracted when passing from a high refractive index medium (n_1) to a lower refractive index medium (n_2). The critical angle (θ_c)can be seen where an evanescent wave is produced entering the lower refractive index medium. (B) The evanescent wave penetrates a short distance into the second medium selectively exciting bound fluorophores while note exciting fluorophores far from the surface.

In order to achieve this selective surface excitation, TIRFM relies on the use of an evanescent wave. An evanescent wave is produced when a light source passes through a medium of a particular refractive index (n_1) and encounters a boundary of a second medium of a lower refractive index (n_2) causing it to refract. This angle of refraction (θ_r) is dependent on the two refractive indexes (n_1 and n_2) and the angle on incidence (θ_i) as defined by Snell's Law (Axelrod et al., 1984) (Figure 2.1A):

$$n_1 * \theta_i = n_2 * \theta_r \tag{2.1}$$

The evanescent wave is only produced when θ_i is greater than the 'critical angle', where

(θ_c) is defined as:

$$\theta_c = sin^{-1}\frac{n_2}{n_1} \tag{2.2}$$

This evanescent field will penetrate into the second medium by only a short distance, where the intensity begins to decrease, dependent upon n_1, n_2, θ_i and the wavelength of the incident light (λ_i) used:

$$d = \frac{\lambda_i}{4\pi}(n_1^2 * sin^2\theta_i - n_2^2)^{-\frac{1}{2}} \tag{2.3}$$

In this project, a 532nm laser was predominately used as λ_i, with n_1 and n_2 being 1.46 and 1.33 respectively (quartz slide and water), the penetration depth would vary between 800 - 77 nm as the θ_i increases past the θ_c.

Due to this phenomenon, the evanescent wave selectively excites the fluorophores within the penetration distance, while keeping the fluorophores outside of the wave in their dark state. This in turn reduces the background noise and increases the SNR that is required for smFRET (Figure 2.1B).

2.2.2 Single-Molecule Forster Resonance Energy Transfer (FRET)

Fluorescence occurs when a fluorophore is excited to a higher energy level and upon decay returns to the lower energy level, in turn releasing a photon at a specific wavelength (Figure 2.2A) (Lakowicz, 2006). In traditional confocal microscopy, the resolution of this fluorescence is limited to ∼250 nm, due to the diffraction limit of light (Lakowicz, 2006). To monitor small-scale dynamic movements which often occur in the low nm range within proteins, this is too large. In order to achieve such resolution, several techniques have been utilised including super-resolution imaging (STORM, PALM, DNA-PAINT) and FRET.

FRET is the principle of energy transfer between two fluorophores (donor and acceptor) due to a transfer of energy via dipole-dipole interactions (Ha et al., 1996). In order for this transfer to occur, the dyes must have spectral overlap and be within close proximity to each other (10-100Å). If the FRET pair meet these criteria (e.g. Cy3 and Cy5), then excitation of the donor fluorophore will cause it to be excited to a higher energy state as normal, but upon its decay some of the energy released will be transferred to the acceptor fluorophore, in turn exciting it (Ha et al., 1996). The acceptor fluorophore will now also reach a higher energy level and upon its own decay emit photons at its given wavelength (Figure 2.2A). This principle essentially circumvents the diffraction limit and the user is able to monitor dynamic movements of diffraction limited single-molecules between 1 - 10 nm.

As mentioned, the efficiency of this transfer (E) is dependent on both the spectral overlap of the dyes and the distance between them (Figure 2.2B). One important factor to consider is R_0, a term which defines the distance at which 50% energy transfer takes place between FRET pairs and is calculated as such:

$$R_0 = 9.725 * 10^3 (Q_D * J(\lambda) * n^{-4} * k^2)^{\frac{1}{6}} \tag{2.4}$$

Where Q_D is the donor fluorophore quantum yield, $J(\lambda)$ is the overlap between the donor emission spectra and acceptor excitation spectra, n is the refractive index of the medium used, k is the orientation of the dyes. k^2 can be reduced to approx. $\frac{2}{3}$ for Cy3 and Cy5 which are the two fluorophores used in this study giving a R_0 of approx 58Å.

Once the R_0 is known, the E can be calculated with the following equation, where R is

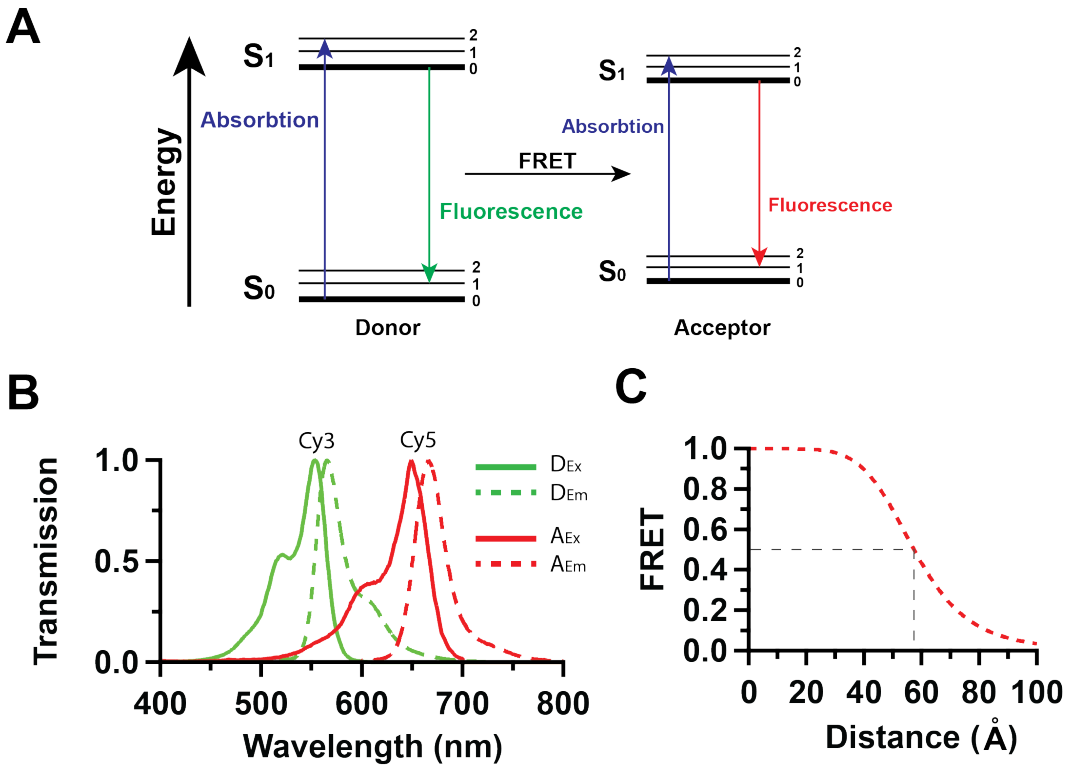


Figure 2.2: **Principles of smFRET.** (A) Jablonski diagram showing fluorescence excitation and emission of donor fluorophore (D) and FRET transfer to the respective acceptor fluorophore (A) causing its own emission. (B) Excitation (Ex) and emission (Em) spectra of Cy3 (D) and Cy5 (A) fluorophores. (C) FRET efficiency (E) for Cy3 and Cy5 fluorophores used in this study showing the response of FRET against the distance between dyes. Black dashed line represents R_0 which in this case is 58Å. Figures adapted from (Lakowicz, 2006)

the distance between the two fluorophores:

$$E = \frac{R_0^6}{R^6 + R_0^6} \tag{2.5}$$

However, during a normal single-molecule experiment E is usually calculated from measuring the intensity of both the donor (I_D) and acceptor (I_A) fluorophores upon donor excitation. These values can then be inserted into the FRET equation (Figure 2.2C):

$$E = \frac{I_A}{I_A + I_D} \tag{2.6}$$

The E value gives precise information about the proximity of the two dyes and can be used by itself, or converted back into a distance measurement using Equation 2.6. Both are commonly quoted, but in many instances the E alone is sufficient and I will use this throughout this thesis.

FRET can be used in a variety of ways to estimate the conformation of an entire population (via fluorometers or native electrophoresis) or at the scale of single-molecules. Average FRET measurements can be useful for many experiments, such as looking at binding of proteins or other nucleic acids, but it cannot provide insight into possible heterogeneity of a sample. The Weiss lab showed that FRET can be used to measure single molecules (smFRET) by utilising confocal scanning where a dilute sample of molecules will diffuse through the excitation volume, where E can then be calculated (Ha et al., 1996). This provided greater insight into the heterogeneity of the sample and helped to observe different conformational identities. One limitation of confocal smFRET however, is that the observed time scale is relatively quick (µs range), meaning observation of the transitions between the observed FRET states be difficult can. This is because the molecule will diffuse through the small excitation volume rapidly (quicker than most dynamics occur) and so although you will obtain information about the different conformations it adopts, the transitions between them will not be captured.

In order to observe these transitions, smFRET can be combined with TIRF where molecules are immobilised onto a surface and excited via TIRFM. Multiple molecules bound to the surface will now be excited at once, while the background fluorescence will remain low due to the TIRF excitation wave, leading to the high SNR that is required. The same molecule can also be tracked over time limited mainly by its fluorescence lifetime. Therefore, unlike confocal smFRET, the transitions between conformations can be tracked over

several minutes. This combination provides several benefits that allows the user to monitor the conformational transitions in more detail providing greater information as to the mechanism of various proteins, such as RNA Polymerase II.

In this thesis I will use a TIRF setup to accommodate smFRET that predominantly uses a 532nm and, a prism to refract the light source and produce the TIRF wave, a 60X objective with emission filters in place, a dual beam splitter to separate the donor and acceptor emission paths and an EMCCD camera to detect the fluorescence emission and determine the FRET ratio of single molecules (Figure 2.3).

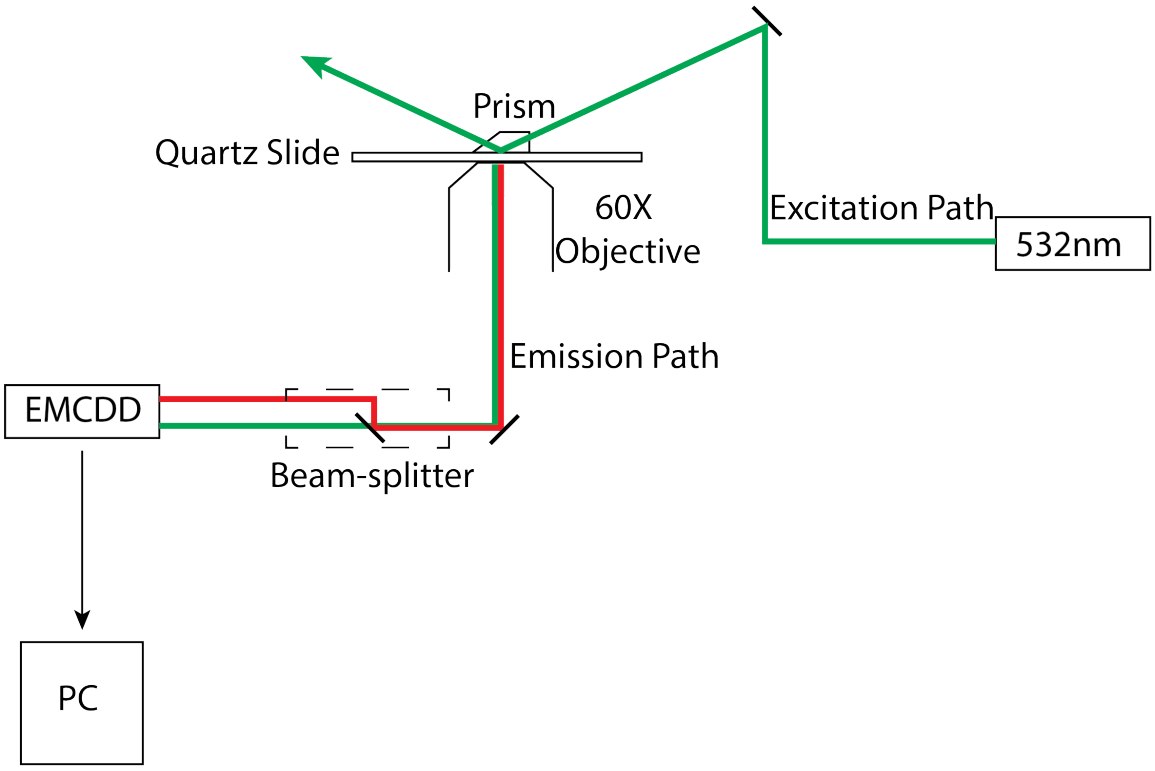


Figure 2.3: **Schematic of smFRET TIRF setup used in this thesis.** Fluorophores were excited via 532nm laser and TIRF field generated by refraction through a prism. The Cy3 and Cy5 emissions were both collected, split via a beam splitter and detected via an EMCCD. The intensity is then transferred to a PC to generate an image and calculate FRET. See Section 3.3 for more details.

2.3 Optical Trapping Microscopy

Optical trapping or optical tweezers is a single-molecule method used to study the biophysical properties of DNA, proteins and their interaction together. It is based on the principle that focused light (such as a laser) can be used to trap small microspheres. This was first shown by Arthur Ashkin in both an *in vitro* and *in vivo* setting, which later won him the 2018 Nobel Prize (Ashkin, 1970; Ashkin et al., 1990).

Manipulation of DNA and RNA require immobilisation to microspheres, as DNA cannot be trapped in the trapping beam itself, and three main geometries of optically trapped microspheres are often utilised. The first being surface based, where the DNA is trapped between a surface on one side and an optically trapped microsphere on the other. Two more geometries consist of DNA trapped between two microspheres, held either by optical traps or micro-pipettes (Bustamante et al., 2021). The use of this method allowed greater insight into the properties of DNA structure and also into the force generation function of the motor proteins kinesin and myosin (Wang et al., 1997; Block et al., 1990; Finer et al., 1994). More recently, optical tweezers has been utilised to gain insight into a variety of biological processes, just as smFRET and TIRF have. These include but are not limited to protein folding (Bustamante et al., 2020), the cytoskeleton and extracellular matrix (Hu et al., 2017) and RNA transcription (Chen et al., 2019; Herbert et al., 2008).

Although measuring the force by itself is very useful, the incorporation of fluorescence microscopy to optical tweezers has enhanced its powers even further (recently termed 'fleezers') (Bustamante et al., 2021; Ma et al., 2019). As briefly mentioned above, there are several different types of fluorescence microscopy, including widefield, confocal and TIRF. All three have been combined with optical tweezers but TIRF can be difficult

to utilise as it is a surface based measurement and this causes issues when using dual bead systems (Heller et al., 2014). Furthermore, widefield microscopy leads to a high background noise and low signal-to-noise ratio and so is not often utilised for single-molecule imaging (Galletto et al., 2006). More commonly used is confocal scanning, which allows the tracking of single fluorescent proteins along trapped DNA while maintaining a high signal-to-noise ratio (Newton et al., 2019; Losito et al., 2021; Anand et al., 2021). The main drawback of confocal scanning is the tracking of biomolecules along DNA is diffraction limited (250 nm), unless combined with super-resolution microscopy such as STED which can increase the resolution to $\sim$ 50 nm. I will utilise the dual-trapped system with confocal scanning in this thesis by using a commercial system called the C-trap (LUMICKS). This allows easy combination of the two systems to monitor RNA transcription (Chapter 6).

2.3.1 Properties of Optical Tweezers

Optical tweezers allows more direct control over the biophysical properties of DNA (stretch and twist) than smFRET and TIRF allows, as these are surface based assays and once immobilised are difficult to manipulate. This method relies on the principle that photons (such as in laser beams) carry a momentum and can be used to 'trap' microspheres by transferring this momentum to them when close to the trapping laser (Ashkin, 1970; Bustamante et al., 2021). Upon trapping, two main forces are applied to the microsphere, scattering and gradient forces (F_{scat} and F_{grad}) (Figure 2.4A). F_{scat} arises from the photons hitting the sphere and being scattered in all directions. As the trapping beam is directional and has a high intensity, the resulting change in momentum of the particle is also in the direction of the beam (Neuman and Block, 2004). F_{grad} originates from the fact that these

same photons can also pass through the sphere or be refracted depending on its refractive index and the medium being used. If the refractive index of the sphere is greater than the surrounding media (as is the case with polystyrene beads often used) then the photons are refracted causing the sphere to be displaced slightly in the opposite direction (Bustamante et al., 2021) (Figure 2.4A). Due to the non-homogenous laser beam being used having a high intensity gradient at the focal centre, the trapping laser will attract the sphere that has been displaced back to the centre of the beam(Neuman and Block, 2004). In order for the microsphere to be stably trapped, F_{grad} must exceed F_{scat} to prevent the sphere being pushed out of the focused trapping laser. To achieve this, a high numerical aperture objective is used to obtain a steep gradient in the trapping laser (Neuman and Block, 2004).

Once trapped, the force (F) applied to the sphere can be described using Hooke's Law:

$$F = k * x \tag{2.7}$$

Where k is the trap stiffness and x is the displacement of the sphere from the focus. Optical tweezers allows the manipulation of both F and x where the sphere can be displaced and the force measured, or a force can be applied to the sphere and the displacement measured. Both of these must be accurately measured for precise optical tweezer experiments.

The displacement of the sphere (x) is usually measured using a quadrant photodiode (QPD) which can detect the deflection of the scattered light in both x and y axis (Figure 2.4B). The trap stiffness (k) is determined by capturing a sphere of known size and calibrating based on the Brownian motion of the sphere when no force is applied (Neuman and Block, 2004). In general, a high trap stiffness will cause the high frequency movements

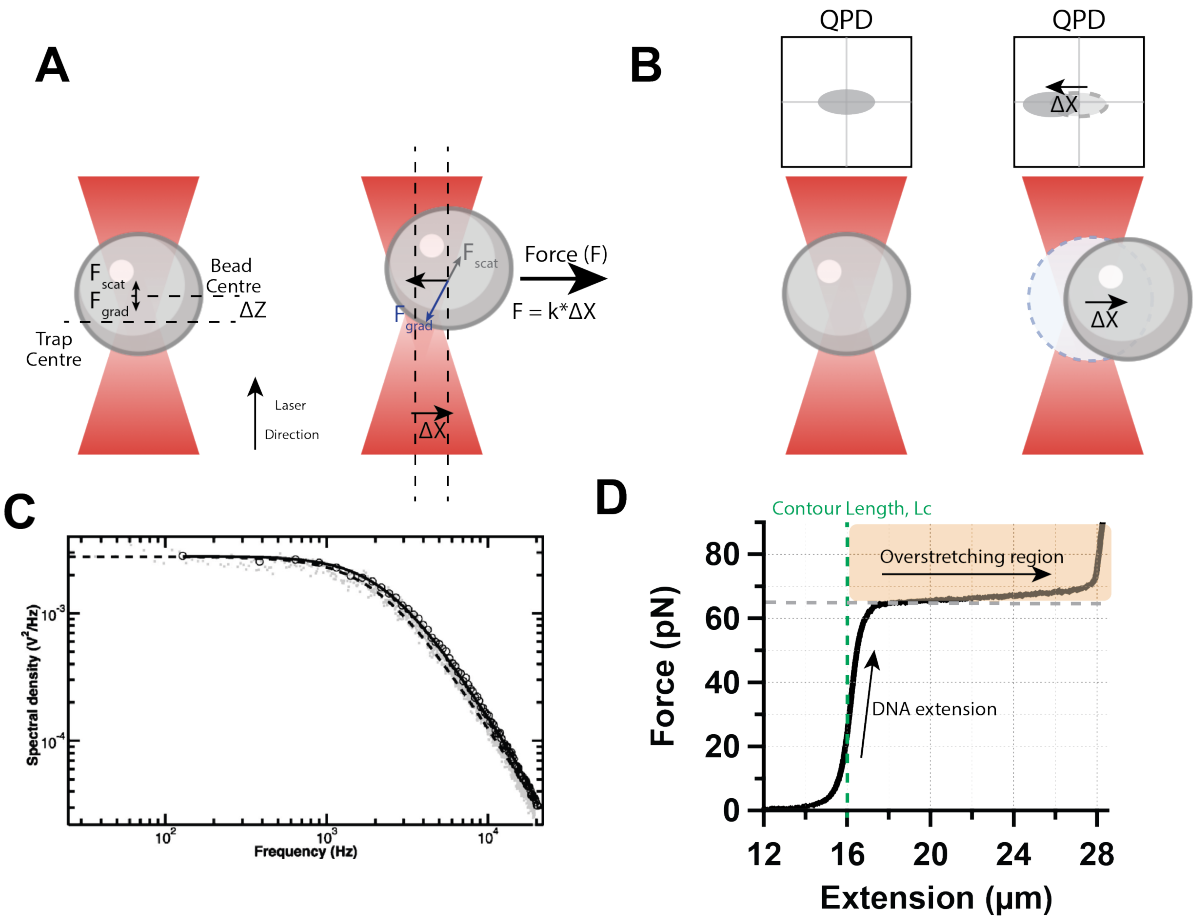


Figure 2.4: **Principles of Optical Tweezers** (A) Microsphere trapped in the trapping beam showing the F_{grad} and F_{scat} acting on it when no force (left) and a force (right) is applied. Adapted from Heller et al. (2014). (B) QPD detection of a displaced microsphere (Δx from the centre of trapping beam Adapted from Heller et al. (2014). (C) Determination of the corner frequency from a force calibration using a known microsphere. Adapted from Neuman and Block (2004). (D) Force-distance curve of $\lambda - DNA$ trapped between two trapped microspheres upon stretching. DNA is pulled from 0 - 85 pN showing the contour length of DNA and the overstretching region at higher forces.

to be more pronounced. The Brownian motion at these varying frequencies are measured, producing a power spectrum from which a corner frequency (f_0) is determined (Figure 2.4C). The trap stiffness is then determined using this calibrated f_0 via the following equation:

$$k = 12\pi^2 * f_0 * n * r \tag{2.8}$$

Where n is the viscosity of the medium used and r is the radius of the trapped bead. In this thesis two main sizes of polystyrene beads are used (1.78 µm and 4.45 µm).

2.3.2 Properties of optically trapped DNA molecules

Biomolecules such as DNA are too small to interact with the optical beam and so cannot be directly trapped, instead they are tethered to microspheres which can be trapped (as discussed above) and manipulated. DNA can generally be described as an elastic rod with several elastic properties affecting its structure (bending, twisting and stretching) (Bustamante et al., 2021). If dsDNA is held between two optically trapped beads it can be stretched in one direction and both the force and extension of the DNA can be measured, producing a force-distance (FD) curve (Figure 2.4D). When minimal force is applied to the DNA (0 - 5 pN) the DNA is in a naturally disordered state, as the extension is below the contour length (L_c) of the DNA. Upon further stretching, the DNA approaches the L_c and the force rapidly increases (Figure 2.4D) until an over-stretching region is reached past the L_c, where the DNA becomes more straight and rigid (Smith et al., 1996). In this region small changes in the end-to-end distance of the DNA results in large changes in the force. This is true until the 'over stretching' region is met, where the natural properties that hold dsDNA together begin to diminish including loss of base-pairing, where it will eventually form two single strands of DNA. In this region, large changes in distance lead to minimal change in the observed force (Smith et al., 1996) (Figure 2.4D). Due to its intrinsically elastic properties, dsDNA can often be modeled using the Worm-like chain (WLC) equation to determine the length (L_{DNA}), the contour length (L_c) or the persistence length (L_P) of dsDNA.

$$L_{DNA} = L_c(1 - \frac{1}{2}\sqrt{\frac{k_B T}{F * L_P} + \frac{F}{S}}) \tag{2.9}$$

Where, k_B is the Boltzman constant (4.11 pN/nm), T is the temperature in Kelvin and

Table 2.1: Biophysical parameters of DNA

DNA	L_c (nm bp^{-1})	L_P (nm)	S (pN)
dsDNA	0.34	50	1400
ssDNA	0.6	10	800

Note: Measured in 150 mM NaCl at pH 7, adapted from (Wang et al., 1997)

S is the stretching modulus (pN). The general properties of ds- and ss-DNA can be found in Table 2.1

ssDNA however does not fit this model as it is even more flexible than dsDNA due to each base being free to move and not limited by base-pairing. Instead, it is modelled by the freely-jointed chain (FJC)(Wang et al., 1997; Bustamante et al., 2021):

$$L_{DNA} = (coth(\frac{2 * F * L_P}{k_B T}) - (\frac{kBT}{2 * F * L_P})) * (1 + \frac{F}{S}) \tag{2.10}$$

Although most DNA used to study transcription is double-stranded and so uses the WLC model, the hybrid DNA I generate (Chapter 6) contains both double-stranded and single-stranded region and so both models are useful to probe the structure of hybrid DNA.

2.4 Single-Molecule studies of RNAPII elongation

As discussed in Section 1, RNA transcription is a highly complex and dynamic process involving multiple proteins, including RNAPII. Although ensemble assays have been heavily used to study these processes, and do provide insight into the start and end points of transcription, they often struggle to capture the complex dynamics of this process. Single-molecule imaging has allowed greater insight into this dynamic process where early works by the Weiss and Kapanidis labs probed the initiation mechanism of bacterial RNAPs (Kapanidis et al., 2006). More recent work using both smFRET and optical tweezers

has provided more insight into eukaryotic polymerases, but mainly focused on the initiation stage or transcription on long DNA templates requiring modification of RNAPII (Revyakin et al., 2012; Dangkulwanich et al., 2013; Baek et al., 2021). However, the dynamics of transcription over the nucleotide scale and backtracking is not fully understood and so requires novel smFRET techniques to monitor this.

Eukaryotic DNA is also highly complex and often contains multiple roadblocks to transcription including nucleosomes, bound proteins (DNA polymerases) and DNA damage, all of which can hinder RNAPII. The effect DNA damage has on transcription is still lacking mechanistic detail with most studies using structural approaches (Hantsche and Cramer, 2017; Martinez-Rucobo and Cramer, 2013; Bernecky et al., 2016; Wang et al., 2018). Although these are beneficial and provide good understanding, as mentioned earlier they are only a snapshot of possible conformations and does not provide the necessary detail as to the transitions between possible states. In this thesis, I utilise smFRET assays to probe these dynamics of RNAPII transcription and focus on the effect DNA damage has on this process.

Bibliography

R. Anand, E. Buechelmaier, O. Belan, M. Newton, A. Vancevska, A. Kaczmarczyk, T. Takaki, D. S. Rueda, S. N. Powell, and S. J. Boulton. HELQ is a dual-function DSB repair enzyme modulated by RPA and RAD51. *Nature 2021 601:7892*, 601 (7892):268–273, 12 2021. ISSN 1476-4687. doi: 10.1038/s41586-021-04261-0. URL https://www.nature.com/articles/s41586-021-04261-0.

J. Andrecka, R. Lewis, F. Brückner, E. Lehmann, P. Cramer, and J. Michaelis. Single-molecule tracking of mRNA exiting from RNA polymerase II. *Proceedings of the National Academy of Sciences of the United States of America*, 105(1):135–140, 1 2008. ISSN 00278424. doi: 10.1073/PNAS.0703815105/SUPPL{_}FILE/03815FIG13.PDF.

A. Ashkin. Acceleration and Trapping of Particles by Radiation Pressure. *Physical Review Letters*, 24(4):156, 1 1970. ISSN 00319007. doi: 10.1103/PhysRevLett.24.156. URL https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.24.156.

A. Ashkin, K. Schütze, J. M. Dziedzic, U. Euteneuer, and M. Schliwa. Force generation of organelle transport measured in vivo by an infrared laser trap. *Nature 1990 348:6299*, 348(6299):346–348, 1990. ISSN 1476-4687. doi: 10.1038/348346a0. URL https://www.nature.com/articles/348346a0.

D. Axelrod, T. P. Burghardt, and N. L. Thompson. TOTAL INTERNAL REFLECTION FLUORESCENCE. *Annual Review Biophysics and Bioengineering*, 13:247–268, 1984.

I. Baek, L. J. Friedman, J. Gelles, and S. Buratowski. Single-molecule studies reveal branched pathways for activator-dependent assembly of RNA polymerase II preinitiation complexes. *Molecular Cell*, 81(17):3576–3588, 2021. ISSN 10974164. doi: 10.1016/j.molcel.2021.07.025. URL https://doi.org/10.1016/j.molcel.2021.07.025.

C. Bernecky, F. Herzog, W. Baumeister, J. M. Plitzko, and P. Cramer. Structure of transcribing mammalian RNA polymerase II. *Nature*, 529(7587):551–554, 2016. ISSN 14764687. doi: 10.1038/nature16482. URL http://dx.doi.org/10.1038/nature16482.

S. M. Block, L. S. Goldstein, and B. J. Schnapp. Bead movement by single kinesin molecules studied with optical tweezers. *Nature 1990 348:6299*, 348(6299):348–352,

1990. ISSN 1476-4687. doi: 10.1038/348348a0. URL https://www.nature.com/articles/348348a0.

C. Bustamante, C. Bustamante, L. Alexander, K. MacIuba, and C. M. Kaiser. Single-Molecule Studies of Protein Folding with Optical Tweezers. *https://doi.org/10.1146/annurev-biochem-013118-111442*, 89:443–470, 6 2020. ISSN 15454509. doi: 10.1146/ANNUREV-BIOCHEM-013118-111442. URL https://www.annualreviews.org/doi/abs/10.1146/annurev-biochem-013118-111442.

C. J. Bustamante, Y. R. Chemla, S. Liu, and M. D. Wang. Optical tweezers in single-molecule biophysics. *Nature Reviews Methods Primers 2021 1:1*, 1(1):1–29, 3 2021. ISSN 2662-8449. doi: 10.1038/s43586-021-00021-6. URL https://www.nature.com/articles/s43586-021-00021-6.

C. Chen, B. Stevens, J. Kaur, D. Cabral, H. Liu, Y. Wang, H. Zhang, G. Rosenblum, Z. Smilansky, Y. E. Goldman, and B. S. Cooperman. Article Single-Molecule Fluorescence Measurements of Ribosomal Translocation Dynamics. *Molecular Cell*, 42: 367–377, 2011. doi: 10.1016/j.molcel.2011.03.024.

Z. Chen, R. Gabizon, A. I. Brown, A. Lee, A. Song, C. Díaz-Celis, C. D. Kaplan, E. F. Koslover, T. Yao, and C. Bustamante. High-resolution and high-accuracy topographic and transcriptional maps of the nucleosome barrier. *eLife*, 8:1–38, 2019. ISSN 2050-084X. doi: 10.7554/elife.48281.

T. D. Christian, L. J. Romano, and D. Rueda. Single-molecule measurements of synthesis by DNA polymerase with base-pair resolution. *Proceedings of the National Academy of Sciences*, 106(50):21109–21114, 2009. ISSN 0027-8424. doi: 10.1073/pnas.0908640106.

M. Dangkulwanich, T. Ishibashi, S. Liu, M. L. Kireeva, L. Lubkowska, M. Kashlev, and C. J. Bustamante. Complete dissection of transcription elongation reveals slow translocation of RNA polymerase II in a linear ratchet mechanism. *eLife*, 2013(2):1–22, 2013. ISSN 2050084X. doi: 10.7554/eLife.00971.

J. T. Finer, R. M. Simmons, and J. A. Spudich. Single myosin molecule mechanics: piconewton forces and nanometre steps. *Nature 1994 368:6467*, 368(6467):113–119, 1994. ISSN 1476-4687. doi: 10.1038/368113a0. URL https://www.nature.com/articles/368113a0.

T. Funatsu, Y. Harada, M. Tokunaga, K. Saito, and T. Yanagida. Imaging of single fluorescent molecules and individual ATP turnovers by single myosin molecules in aqueous solution. *Nature*, 374(6522):555–559, 1995.

H. L. Gahlon, L. J. Romano, and D. Rueda. Influence of DNA Lesions on Polymerase-Mediated DNA Replication at Single-Molecule Resolution. *Chemical Research in Toxicology*, 30(11):1972–1983, 11 2017. ISSN 15205010. doi: 10.1021/ACS.CHEMRESTOX.7B00224/ASSET/IMAGES/MEDIUM/TX-2017-00224M{_}0010.GIF. URL `https://pubs.acs.org/doi/abs/10.1021/acs.chemrestox.7b00224`.

R. Galletto, I. Amitani, R. J. Baskin, and S. C. Kowalczykowski. Direct observation of individual RecA filaments assembling on single DNA molecules. *Nature 2006 443:7113*, 443(7113):875–878, 9 2006. ISSN 1476-4687. doi: 10.1038/nature05197. URL `https://www.nature.com/articles/nature05197`.

T. Ha, T. H. Enderle, D. F. Ogletreet, D. S. Chemla, P. R. Selvin, and S. Weiss. Probing the interaction between two single molecules: Fluorescence resonance energy transfer between a single donor and a single acceptor. *Proceedings of the National Academy of Sciences*, 93:6264–6268, 1996.

T. Ha, A. Y. Ting, J. Liang, W. B. Caldwell, A. A. Deniz, D. S. Chemla, P. G. Schultz, and S. Weiss. Single-molecule fluorescence spectroscopy of enzyme conformational dynamics and cleavage mechanism. *Proceedings of the National Academy of Sciences of the United States of America*, 96(3):893–898, 2 1999a. ISSN 00278424. doi: 10.1073/PNAS.96.3.893/ASSET/F17D03FD-7BCB-4F45-B393-E955229DB0D3/ASSETS/GRAPHIC/PQ0294336004.JPEG. URL `www.pnas.org`.

T. Ha, X. Zhuang, H. D. Kim, J. W. Orr, J. R. Williamson, and S. Chu. Ligand-induced conformational changes observed in single RNA molecules. *Biophysics*, 96:9077–9082, 1999b. URL `www.pnas.org`.

M. Hantsche and P. Cramer. Conserved RNA polymerase II initiation complex structure. *Current Opinion in Structural Biology*, 47:17–22, 2017. ISSN 1879033X. doi: 10.1016/j.sbi.2017.03.013. URL `http://dx.doi.org/10.1016/j.sbi.2017.03.013`.

I. Heller, T. P. Hoekstra, G. A. King, E. J. Peterman, and G. J. Wuite. Optical tweezers analysis of DNA-protein complexes. *Chemical Reviews*, 114(6):3087–3119, 3 2014. ISSN

15206890. doi: 10.1021/CR4003006/ASSET/IMAGES/LARGE/CR-2013-003006{_}0034.JPEG. URL https://pubs.acs.org/doi/full/10.1021/cr4003006.

K. M. Herbert, W. J. Greenleaf, and S. M. Block. Single-Molecule Studies of RNA Polymerase: Motoring Along. *https://doi.org/10.1146/annurev.biochem.77.073106.100741*, 77:149–176, 6 2008. ISSN 00664154. doi: 10.1146/ANNUREV.BIOCHEM.77.073106.100741. URL https://www.annualreviews.org/doi/abs/10.1146/annurev.biochem.77.073106.100741.

T. Hirschfeld. Optical microscopic observation of single small molecules. *Applied Optics*, 15(12):2965, 1976.

J. Hohlbein, L. Aigrain, T. D. Craggs, O. Bermek, O. Potapova, P. Shoolizadeh, N. D. Grindley, C. M. Joyce, and A. N. Kapanidis. Conformational landscapes of DNA polymerase I and mutator derivatives establish fidelity checkpoints for nucleotide insertion. *Nature Communications 2013 4:1*, 4(1):1–11, 7 2013. ISSN 2041-1723. doi: 10.1038/ncomms3131. URL https://www.nature.com/articles/ncomms3131.

X. Hu, F. M. Margadant, M. Yao, and M. P. Sheetz. Molecular stretching modulates mechanosensing pathways. *Protein Science*, 26(7):1337–1351, 7 2017. ISSN 1469-896X. doi: 10.1002/PRO.3188. URL https://onlinelibrary.wiley.com/doi/full/10.1002/pro.3188https://onlinelibrary.wiley.com/doi/abs/10.1002/pro.3188https://onlinelibrary.wiley.com/doi/10.1002/pro.3188.

A. N. Kapanidis, E. Margeat, S. O. Ho, E. Kortkhonjia, S. Weiss, and R. H. Ebright. Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science*, 314(5802):1144–1147, 11 2006. ISSN 00368075. doi: 10.1126/SCIENCE.1131399/SUPPL{_}FILE/KAPANIDIS.SOM.PDF. URL https://www.science.org/doi/abs/10.1126/science.1131399.

J. R. Lakowicz. *Principles of fluorescence spectroscopy*. Springer, 2006. ISBN 0387312781. doi: 10.1007/978-0-387-46312-4.

M. Losito, Q. M. Smith, M. D. Newton, M. E. Cuomo, and D. S. Rueda. Cas12a target search and cleavage on force-stretched DNA. *Physical Chemistry Chemical Physics*, 23(47):26640–26644, 12 2021. ISSN 1463-9084. doi: 10.1039/D1CP03408A. URL

https://pubs.rsc.org/en/content/articlehtml/2021/cp/d1cp03408ahttps://pubs.rsc.org/en/content/articlelanding/2021/cp/d1cp03408a.

G. Ma, C. Hu, S. Li, X. Gao, H. Li, and X. Hu. Simultaneous, hybrid single-molecule method by optical tweezers and fluorescence. *Nanotechnology and Precision Engineering*, 2(4):145–156, 12 2019. ISSN 2589-5540. doi: 10.1016/J.NPE.2019.11.004.

A. M. Malinen, J. Bakermans, E. Aalto-Setälä, M. Blessing, D. L. Bauer, O. Parilova, G. A. Belogurov, D. Dulin, and A. N. Kapanidis. Real-Time Single-Molecule Studies of RNA Polymerase–Promoter Open Complex Formation Reveal Substantial Heterogeneity Along the Promoter-Opening Pathway. *Journal of Molecular Biology*, 434(2): 167383, 1 2022. ISSN 0022-2836. doi: 10.1016/J.JMB.2021.167383.

F. W. Martinez-Rucobo and P. Cramer. Structural basis of transcription elongation. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1829(1):9–19, 2013. ISSN 18749399. doi: 10.1016/j.bbagrm.2012.09.002. URL http://dx.doi.org/10.1016/j.bbagrm.2012.09.002.

A. Mazumder, A. Wang, H. Uhm, R. H. Ebright, and A. N. Kapanidis. RNA polymerase clamp conformational dynamics: long-lived states and modulation by crowding, cations, and nonspecific DNA binding. *Nucleic Acids Research*, 49(5):2790–2802, 3 2021. ISSN 0305-1048. doi: 10.1093/NAR/GKAB074. URL https://academic.oup.com/nar/article/49/5/2790/6137300.

J. B. Munro, R. B. Altman, N. O'Connor, and S. C. Blanchard. Identification of Two Distinct Hybrid State Intermediates on the Ribosome. *Molecular Cell*, 25(4):505–517, 2 2007. ISSN 10972765. doi: 10.1016/J.MOLCEL.2007.01.022/ATTACHMENT/7FEDA20B-05AD-4F3F-9622-9C6A6B3A73E2/MMC1.PDF. URL http://www.cell.com/article/S1097276507000457/fulltexthttp://www.cell.com/article/S1097276507000457/abstracthttps://www.cell.com/molecular-cell/abstract/S1097-2765(07)00045-7.

K. C. Neuman and S. M. Block. Optical trapping. 2004. doi: 10.1063/1.1785844. URL http://rsi.aip.org/rsi/copyright.jsp.

M. D. Newton, B. J. Taylor, R. P. Driessen, L. Roos, N. Cvetesic, S. Allyjaun, B. Lenhard, M. E. Cuomo, and D. S. Rueda. DNA stretching induces Cas9 off-target activity.

Nature Structural & Molecular Biology 2019 26:3, 26(3):185–192, 2 2019. ISSN 1545-9985. doi: 10.1038/s41594-019-0188-z. URL https://www.nature.com/articles/s41594-019-0188-z.

A. Revyakin, Z. Zhang, R. A. Coleman, T. R. Strick, and N. Hernandez. single-molecule resolution Transcription initiation by human RNA polymerase II visualized at. *Genes Dev. Related Content Genes Dev*, 26:1643–1647, 2012. doi: 10.1101/gad.194936.112. URL http://genesdev.cshlp.org/subscriptions.

N. C. Robb, T. Cordes, L. C. Hwang, K. Gryte, D. Duchi, T. D. Craggs, Y. Santoso, S. Weiss, R. H. Ebright, and A. N. Kapanidis. The Transcription Bubble of the RNA Polymerase–Promoter Open Complex Exhibits Conformational Heterogeneity and Millisecond-Scale Dynamics: Implications for Transcription Start-Site Selection. *Journal of Molecular Biology*, 425(5):875–885, 3 2013. ISSN 0022-2836. doi: 10.1016/J.JMB.2012.12.015.

D. Rueda and N. G. Walter. Single molecule fluorescence control for nanotechnology. *Journal of Nanoscience and Nanotechnology*, 5(12):1990–2000, 12 2005. ISSN 15334880. doi: 10.1166/jnn.2005.505.

S. B. Smith, Y. Cui, and C. Bustamante. Overstretching B-DNA: The Elastic Response of Individual Double-Stranded and Single-Stranded DNA Molecules. *Science*, 271(5250): 795–799, 2 1996. ISSN 00368075. doi: 10.1126/SCIENCE.271.5250.795. URL https://www.science.org/doi/abs/10.1126/science.271.5250.795.

M. Steiner, K. S. Karunatilaka, R. K. O. Sigel, and D. Rueda. Single-molecule studies of group II intron ribozymes. *PNAS*, 105(37):13853–13858, 2008. URL www.pnas.org/cgi/content/full/.

A. I. Tomescu, N. C. Robb, N. Hengrung, E. Fodor, and A. N. Kapanidis. Single-molecule FRET reveals a corkscrew RNA structure for the polymerase-bound influenza virus promoter. *Proceedings of the National Academy of Sciences of the United States of America*, 111(32), 8 2014. ISSN 10916490. doi: 10.1073/PNAS.1406056111/SUPPL{_}FILE/PNAS.201406056SI.PDF. URL www.pnas.org/cgi/doi/10.1073/pnas.1406056111.

B. Valeur and M. N. Berberan-Santos. *Molecular Fluorescence*. Wiley-VCH Verlag & Co, Weinheim, second edition, 2013.

M. D. Wang, H. Yin, R. Landick, J. Gelles, and S. M. Block. Stretching DNA with optical tweezers. *Biophysical Journal*, 72(3):1335–1346, 1997. ISSN 00063495. doi: 10.1016/S0006-3495(97)78780-0.

W. Wang, C. Walmacq, J. Chong, M. Kashlev, and D. Wang. Structural basis of transcriptional stalling and bypass of abasic DNA lesion by RNA polymerase II. *Proceedings of the National Academy of Sciences of the United States of America*, 115(11):E2538–E2545, 2018. ISSN 1091-6490. doi: 10.1073/pnas.1722050115. URL `http://www.ncbi.nlm.nih.gov/pubmed/29487211http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5856558`.

O. Willhoft, M. Ghoneim, C. L. Lin, E. Y. Chua, M. Wilkinson, Y. Chaban, R. Ayala, E. A. McCormack, L. Ocloo, D. S. Rueda, and D. B. Wigley. Structure and dynamics of the yeast SWR1-nucleosome complex. *Science*, 362(6411), 10 2018. ISSN 10959203. doi: 10.1126/SCIENCE.AAT7716/SUPPL{_}FILE/AAT7716-WILLHOFT-SM.PDF. URL `https://www.science.org/doi/abs/10.1126/science.aat7716`.

R. Zhao and D. Rueda. RNA folding dynamics by single-molecule fluorescence resonance energy transfer. *Methods*, 49(2):112–117, 10 2009. ISSN 1046-2023. doi: 10.1016/J.YMETH.2009.04.017.

X. Zhuang, L. E. Bartley, H. P. Babcock, R. Russell, T. Ha, D. Herschlag, and S. Chu. A single-molecule study of RNA catalysis and folding. *Science*, 288(5473):2048–2051, 6 2000. ISSN 00368075. doi: 10.1126/SCIENCE.288.5473.2048/SUPPL{_}FILE/1049406S1{_}THUMB.GIF. URL `https://www.science.org/doi/abs/10.1126/science.288.5473.2048`.

X. Zhuang, H. Kim, M. J. Pereira, H. P. Babcock, N. G. Walter, and S. Chu. Correlating structural dynamics and function in single ribozyme molecules. *Science*, 296(5572):1473–1476, 2002. ISSN 00368075. doi: 10.1126/science.1069013.

Chapter 3

Materials and Methods

3.1 Preparation of DNA and RNA Oligonucleotides

3.1.1 Deprotection of DNA oligonucleotides

Most DNA and RNA oligonucleotides were purchased from Integrated DNA Technologies (IDT). The DNA oligonucleotides containing the 8-Oxo Guanine lesion and the cyclopyrimidine dimer lesion (TT-Dimer) were synthesised by the Keck Oligo synthesis resource at Yale.

Oligonucleotides are synthesised with protecting groups (usually a benzoyl group) as the bases adenine, cytosine and guanine contain an exocyclic amino group and can cause cleavage and degradation of the synthesised oligonucleotide. Usually, oligonucleotides arrive already deprotected, but the oligonucleotide containing the CPD group first had to be deprotected in house as these can interfere with the normal functioning of the oligonucleotide (Glen Research). The oligonucleotide support was placed in the middle of two syringes, one to flush in the desired solution, and one to collect the eluent. Firstly, the oligonucleotide support was washed by flushing a mix of thiophenol/triethylamine/tetrahydrofuran (SIGMA) in a 1:2:2 (v/v/v) ratio via a syringe for 45 minutes at 1 mL/min. Then, the support was further washed by flowing tetrahydrofuran followed by methanol (99%, Fischer Scientific) at 1 mL/min, and this step repeated 5 times. A final wash step was carried out by flushing acetonitrile (HPLC grade, Fischer Scientific) 3 times over the column, also at 1 mL/min. In order to elute the DNA oligonucleotide off the support, fresh ammonium hydroxide (28-30%, SIGMA) was flowed over the column at a slower rate of 0.2 mL/min over a 1-hour period. The eluent was left to incubate overnight in the dark at room temperature. Finally, the ammonium hydroxide/DNA solution was dried using a vacuum concentrator (Eppendorf Concentrator Plus,

Eppendorf), resuspended in 200 µL nuclease free water and concentration determined via measuring UV absorbance at 260nm (BioDrop µLite, BioDrop).

3.1.2 Purification of DNA and RNA oligonucleotides

Before being used, all DNA and RNA oligonucleotides were first purified via denaturing polyacrylamide gel electrophoresis (PAGE) (8 M Urea, 18% polyacrylamide:bis-acrylamide, 1X TBE). All gels were run at a constant power (20 W) for 2 – 3 hours. The DNA or RNA was visualised via UV-shadowing, a technique where a TLC plate is placed behind the gel and is exposed to UV light, exposing the oligonucleotide band at the desired length. Bands were excised and purified either by the crush and soak method in DNA/RNA elution buffer (40 mM Tris-HCl pH 7.8, 480 mM NaOAC pH5, 0.5 mM EDTA) or by using the ZR-small RNA PAGE recovery kit (Zymo Research).

3.1.3 Fluorescent labelling of DNA and RNA oligonucleotides

Oligonucleotides that needed to be labelled were ordered either with a 5' C6-NH2 group (5AmC6 (IDT), for 5' end labelling) or an internal amino modified thymine base for internal labelling (dT-C6-NH2, iAmMC6dT (IDT)) at the desired residue (Figure 3.1C). After PAGE purification, DNA/RNA oligonucleotides (25 µL @ 100-150 µM) were mixed with 1X labelling buffer (750 mM $NaHCO_3$) and mono-reactive Cy3 or Cy5 dye (Cytiva, PA23001) resuspended in 10 µL DMSO (99%, SIGMA) to a final volume of 50 µL (Figure 3.1A&B). The labelling reaction was left for a minimum of 2 hours at 4°C. To remove any unreacted dye, the reaction was subjected to ethanol precipitation and resuspended in 100 µL TEAA buffer (50 mM Acetonitrile, 2.8 mM triethylamine , pH 7.0).

Finally, labelled and non-labelled oligonucleotides were separated via reverse-phase HPLC

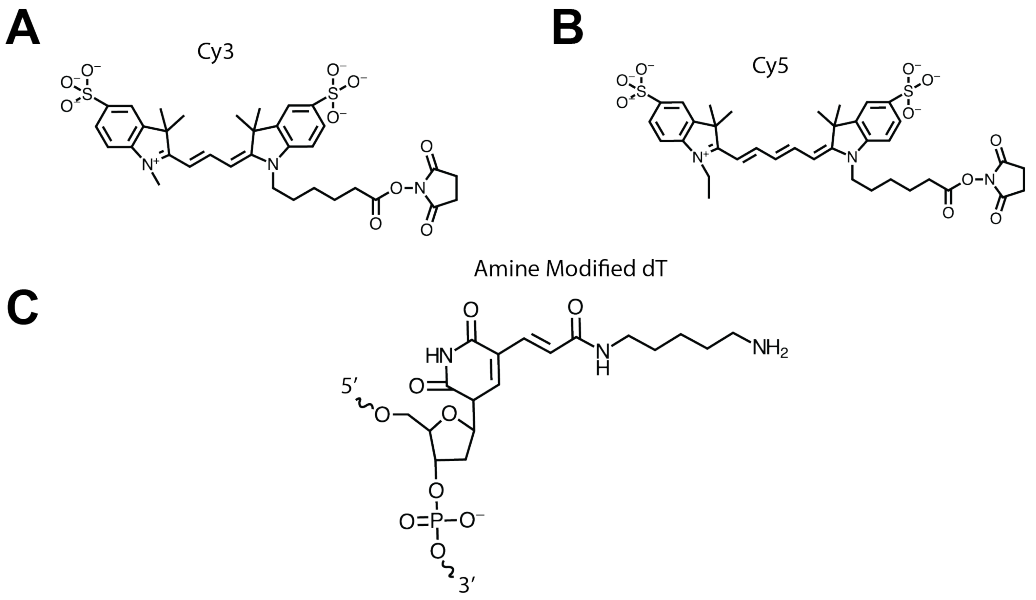


Figure 3.1: **Chemical structures of fluorescent dyes.** (A-B) Structure of mono-reactive succinimidyl ester form of Cy3 (A) and Cy5 (B) used for labelling . (C) Structure of modified amine base used for internal labelling of nucleic acids (IDT).

(Discovery BIO Wide Pore C8 5 µm column, 568323-U) and eluted in TEAA buffer. Labelled product was dried under vacuum, resuspended in 30 µL nuclease-free water and the concentration measured via both UV absorbance and at the corresponding wavelength of the relevant dye as before.

3.2 Purification of eukaryotic RNA Polymerase II

All RNA Polymerase used in this thesis was purified with the help of Zhong Han in the Svejstrup laboratory. Unless otherwise stated all steps were carried out at 4°C.

3.2.1 Purification of yeast RNAPII

Endogenous yeast RNAPII was purified by our collaborators and in particular Zhong Han. Briefly, a yeast strain which encoded a FLAG tag on the RPB3 subunit of RNAPII was used to specifically pull down endogenous RNAPII. Approx 100L of this yeast strain was

grown in YPD. The culture was then grown to an OD600 of between 5-6 before being centrifuged. The pellet was resuspended in lysis buffer (50 mM Hepes pH7.6, 400 mM ammonium sulfate, 10 mM $MgSO_4$, 1 mM EDTA, 10% glycerol and protease inhibitors) and placed in a freezer mill, where the powder is then stored at -80°C until ready to be purified. Typically, 12 tubes of powder are thawed in a water bath at 37°C for 1 hour and resuspended in 15 mL of lysis buffer. This was centrifuged for 1 hour at 40,000 RPM at 4°C and the supernatant collected. The lysate was then applied to an anti-FLAG resin at 1 mL/min and washed in wash buffer (lysis buffer without glycerol or protease inhibitors) and eluted off the column in elution buffer (50 mM Hepes pH 7.6, 110 mM KOAc, 5 mM MgOAc, 1 mM EDTA, FLAG peptide (0.1 mg/ml)). RNAPII containing fractions were then applied to a MonoQ column (anion exchange chromatography) to clean the sample further. Fractions were eluted in a gradient of elution buffer 2 (50 mM Hepes pH 7.6, 2 M KOAc, 5 mM MgOAc, 1 mM EDTA, 10 mM DTT) where it typically eluted at 55% of elution buffer 2. Peak fractions were collated together and concentrated and dialysed using a 100kDa Amicon centrifugal unit in storage buffer (50 mM Hepes pH 7.6, 110 mM KOAc, 5 mM MgOAc, 1 mM EDTA, 10 mM DTT, 5% glycerol). Samples were flash frozen in liquid nitrogen and stored at -80°C.

3.2.2 Purification of mammalian RNAPII

Mammalian RNAPII, was purified from from *Bos taurus* as previously described (Bernecky et al., 2016). Calf thymus (500 g) was homogenised in 1L of Buffer A (50 mM Tris-HCl pH7.8, 10µM $ZnCl_2$, 10% glycerol and protease inhibitors (Sigma)) using a 2L blender (Waring). The homogenised thymus was mixed for 30 minutes, centrifuged at 6000 RPM for 20 minutes and the resulting supernatant passed through a nylon cloth

(Bioprep). Next, a 10% PEI solution was added to the thymus mixture to a final concentration of 0.05% causing precipitation of DNA and DNA bound proteins out of solution. This was then centrifuged again at 12,000xg for 30 minutes, and the RNAPII containing pellet was then resuspended in Buffer B (Buffer A + 150 mM ammonium sulfate). A final centrifugation step at 15,000 RPM for 40 minutes is carried out and the supernatant loaded onto a Fast Flow Q Sepharose column. This anion exchange column contains a positively charged stationary phase and a negatively charged mobile phase. Protein was eluted off the column in Buffer C (Buffer A + 500 mM ammonium sulfate). This eluent contains many negatively charged proteins including RNA Polymerase II. To separate the RNAPII from the other proteins, the eluent was applied to an antibody column at a flow rate of 0.2 ml/min which targets the hypo-phosphorylated CTD of RNAPII (antibody code: 8WG16). RNAPII was eluted in 10 mL fractions in elution buffer (Buffer C + 40% 1,2 propanediol) and dialysed against storage buffer (50 mM Tris-HCl pH 7.8, 10 µM $ZnCl_2$, 5 mM DTT, 150 mM ammonium sulfate, 10% glycerol). Finally, this was concentrated to a final concentration of 1-2 mg/mL (measured via absorbance at 280 nm) using a 100kDa Amicon centrifugal unit. Samples were flash frozen in liquid nitrogen and stored at -80°C.

3.3 *In vitro* transcription assay

Transcription elongation complexes (TECs) were assembled by annealing the RNA primer and template strand (TS) DNA (20 µL TS at 200 nM, 20 µL RNA at 200 nM) in 1X transcription buffer (20 mM Tris-HCl pH 7.8, 100 mM NaCl, 3 mM $MgCl_2$, 10 µM $ZnCl_2$, 2 mM DTT) and heated to 95°C for 2 minutes followed by a slow cool to room temperature (Oh et al., 2019). Next, RNA Polymerase II (2.5 µL at 2.2 µg/mL) was added and

incubated for 20 minutes at room temperature to bind to the RNA:DNA hybrid. Finally, the corresponding non template strand (NTS) was mixed (20 µL at 400 nM), incubated at 37°C for 10 minutes to allow annealing to the TEC, and topped up with 1X transcription buffer to a final volume of 100 µL.

Transcription was started upon addition of NTPs at desired concentration and terminated via addition of STOP buffer (20 mM Tris-HCl pH 7.8, 95% formamide, 50 mM EDTA, 0.2% SDS) at various time points. All reactions were carried out at room temperature. Extension of the RNA was visualised via the 5' Cy5 dye after running on a denaturing 18% PAGE and imaged using a fluorescent gel imager (Amersham 680 RGB, Cytiva). Quantification of bands and further analysis were carried out using FIJI and Igor8 (Wavemetrics).

3.4 TIRF Microscopy

3.4.1 Slide passivation

To create single-molecule imaging chambers (Figure 3.2) holes were drilled (Dremmel) into quartz slides using a diamond tipped drill bit (Kingsley North) (Figure 3.2). Slide preparation and passivation was carried out as per the protocol published from the Rueda lab previously (Lamichhane et al., 2010). Slides were scrubbed with alconox paste for a total of 4 minutes each, followed by rinsing with ddH_2O, ethanol and then ddH_2O again. All slides were then boiled in basic piranha solution (350 mL ddH_2O, 70 mL ammonium hydroxide, 70 mL hydrogen peroxide) for 45 minutes. Slides were removed and dried briefly with a blue flame to remove any leftover organic residue. Cleaned slides and coverslips (Fischer-Scientific) were placed in staining jars containing 1 M KOH and

sonicated for 20 minutes. This was followed by rinsing with ddH_2O and further sonication in methanol. Two layers of coatings were then applied to both the slides and coverslips. Firstly, an aminopropylsilane layer was added by incubating slides and coverslips in a mixture of methanol (100%, 200 mL), aminopropylsilane (2 mL, Vectabond) and glacial acetic acid (10 mL) and incubated for 20 minutes, with a 1 minute round of sonication half way through. Slides and coverslips were washed with methanol and dried via N_2 gas. The second coating was a mixture of methoxy-PEG-SVA (mPEG, MW5000) and biotin-PEG-SVA (bPEG, MW5000), this allowed immobilisation of constructs via biotin-neutravidin linkage. This mixture was created by combining mPEG (90 mg) and bPEG (10 mg) with sodium bicarbonate buffer (640 µL, 84 mM). The PEG solution was added to each slide (120 µL), with a coverslip placed on top to create a PEG layer in between the two. Slides were then left overnight in a dark and humid environment. After the PEGylation step, slides and coverslips were rinsed with ddH_2O and dried via N_2. Slides were stored with respective coverslips in falcon tubes filled with N_2, vacuum sealed and placed in a dark box at -20°C until ready to use. Chambers were constructed by addition of two layers of double-sided tape, to create the desired number and orientation of chambers, the coverslip placed on top, and edges sealed via epoxy resin (Figure 3.2) (Lamichhane et al., 2010).

3.4.2 smFRET experiments

All experiments were carried out on a home-built prism based total internal reflection fluorescence microscope (TIRFM) with a 60X objective . 532 nm (Vortran Stradus) and 647 nm (Vortran Stradus) lasers were used for excitation of Cy3 and Cy5 dyes respectively. For smFRET experiments where only the donor fluorophore was excited, a 542 nm long pass filter (Chroma, 542LP ET) was placed in the emission pathway. Finally, the emission

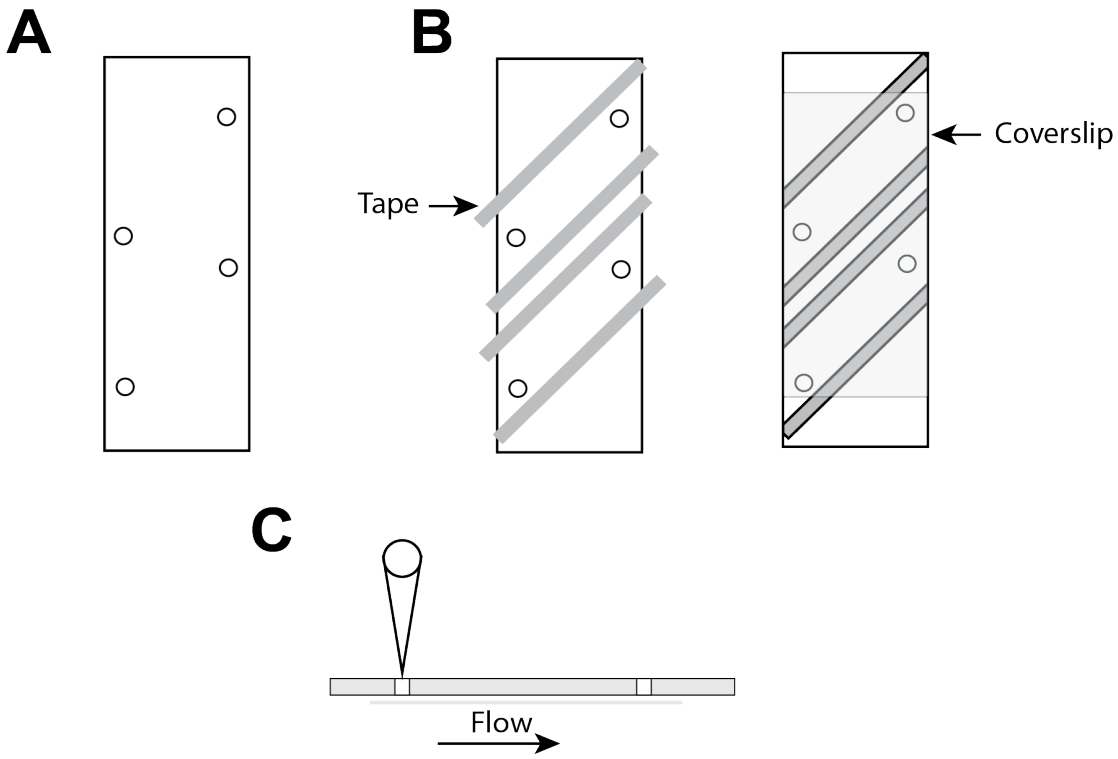


Figure 3.2: **Assembly of smFRET slides.** (A) Drilled hole layout for 2-chamber design slide. (B) Positioning of tape and coverslip to form the desired chambers. (C) Injection of the sample via a pipette tip into slide chamber showing direction of flow applied.

beam was split via a two-way beam splitter which contains a 635 nm long pass filter (Chroma, T635LPXR) to split the Cy3 and Cy5 emission. Fluorescence emission was detected using an Andor iXonUltra EMCCD camera. All experiments were measured at 60 ms exposure time, 3.5 mW 532 nm excitation and 150-200 gain, unless stated otherwise.

Imaging chambers were first passivated with 0.2 mg/mL BSA in T50 buffer (50 mM Tris-HCl pH 7.5, 50 mM NaCl). Next, neutravidin (0.2 mg/mL) was flowed into the chamber and incubated for 10 minutes. Excess neutravidin was washed out with 1X transcription buffer, before immobilisation of desired fluorescent TEC (final 10-20 pM) for 10 minutes. Non-bound TECs were flushed out using imaging buffer (1X transcription buffer, 0.2 mg/mL BSA, 125 mM PCA, 10 µM PCD, Trolox) and the initial FRET measurement taken. To measure transcription, NTPs (10-100 µM) were added to the chamber in fresh

imaging buffer, and the FRET was then measured again for a given time period. Unless stated otherwise, single-molecules were observed for a total of 5 minutes per field-of-view (fov). After 5 minutes, a new fov was found and again the observed FRET was captured for 5 minutes. This was repeated for a total of 50 minutes. This was done to capture the multiple round of backtracking and translocating dynamics that RNAPII observes. In order to capture real time transcription events on undamaged DNA templates, such as in Ch4, imaging was only observed for the first 5 minutes as after this point transcription would be complete. This was then repeated on a fresh sample 3-4 times to capture a high enough sample size.

3.4.3 smFRET Analysis

Extraction of intensity values for each single-molecule required mapping of donor and acceptor channels. This was achieved by using a fluorescent bead slide (2% w/v 0.2 µM FluoSpheres carboxylate-modified microspheres, red fluorescent 580/605). A quartz slide was first cleaned using alconox and water as above. Two pieces of tape were added to create a 1cm wide chamber by placement of a clean coverslip on top. The bead solution was added to the chamber, incubated for 45 minutes before sealing the edges with epoxy. Each day a new bead image was acquired (30 ms exposure time, 0 gain, 100-200 frames) which contained $>$ 100 molecules.

Subsequent extraction of peak intensity values and calculation of FRET efficiencies were calculated via a home-built IDL and MATLAB scripts. A donor to acceptor leakage must also be considered and was estimated to be 10%, which was applied to all calculated FRET values based on the FRET equation (Equation 2.7). Extracted traces were subjected to further analysis using ebFRET (Van De Meent et al., 2013) to identify FRET

states and transitions. Histograms of FRET values were made in Igor8 (wavemetrics) with a minimum of 100 trajectories for each condition tested. FRET trajectories of transcribing molecules were post-synchronized from initial 0.95 FRET state from 10 s before the transcription starting point using a MATLAB script (kindly provided by J. Puglisi, Stanford University). Post-synchronization histograms (PSH) were built by binning the smFRET using 0.05 FRET bins, 10 s time bins and a 0.7 FRET threshold to exclude false decreases due to noisy signals (Blanchard et al., 2004; Senavirathne et al., 2015).

3.5 Optical Tweezer Spectroscopy

Two main constructs were used for the studies of RNAPII transcription that used the optical-tweezers setup. One contained a single-stranded gap at a desired location along λDNA, this gap allowed the RNA primer and RNAPII to bind. The second contained a TEC that was formed as above, then ligated in between two longer handles to form a 18 kb length TEC. (Figures 3.3 and 3.4)

3.5.1 Biotinylation of DNA

Biotinylated λ-DNA was generated by filling in the naturally occurring cos-ends of λ-DNA (5' GGGCGGCGACCT) (Figure 3.3A). For a single biotinylation reaction; λ-DNA (21 µL at 0.03 mg/mL) was mixed in a tube containing NEB Buffer 2 (5µL of 10X), dGTP (0.5 µL at 10mM), dTTP (0.5 µL at 10mM), biotin-dCTP (10 µL at 400 µM) and biotin-dATP (10 µL at 400 µM) (these will place 5-7 biotin molecules on each side) and made up to 50 µL. Klenow Fragment (2.5U) was then added to initiate the fill-in reaction. The reaction was placed in a PCR machine at 37°C for 30 minutes, then 70°C for 15 minutes and finally held at 4°C until ready. Unincorporated biotin-dNTPs, excess dNTPs and Klenow

fragment were removed from the biotinylated-DNA using a QIAquick PCR purification kit and eluted in 50µL of nuclease free water. The final concentration of the biotinylated λ-DNA was determined via absorbance at 260 nm as above, with the usual concentration was ∼60 ng/µL. This was stored at 4°C until ready to use. All pipetting steps are done with extreme care to avoid nicking of the long DNA.

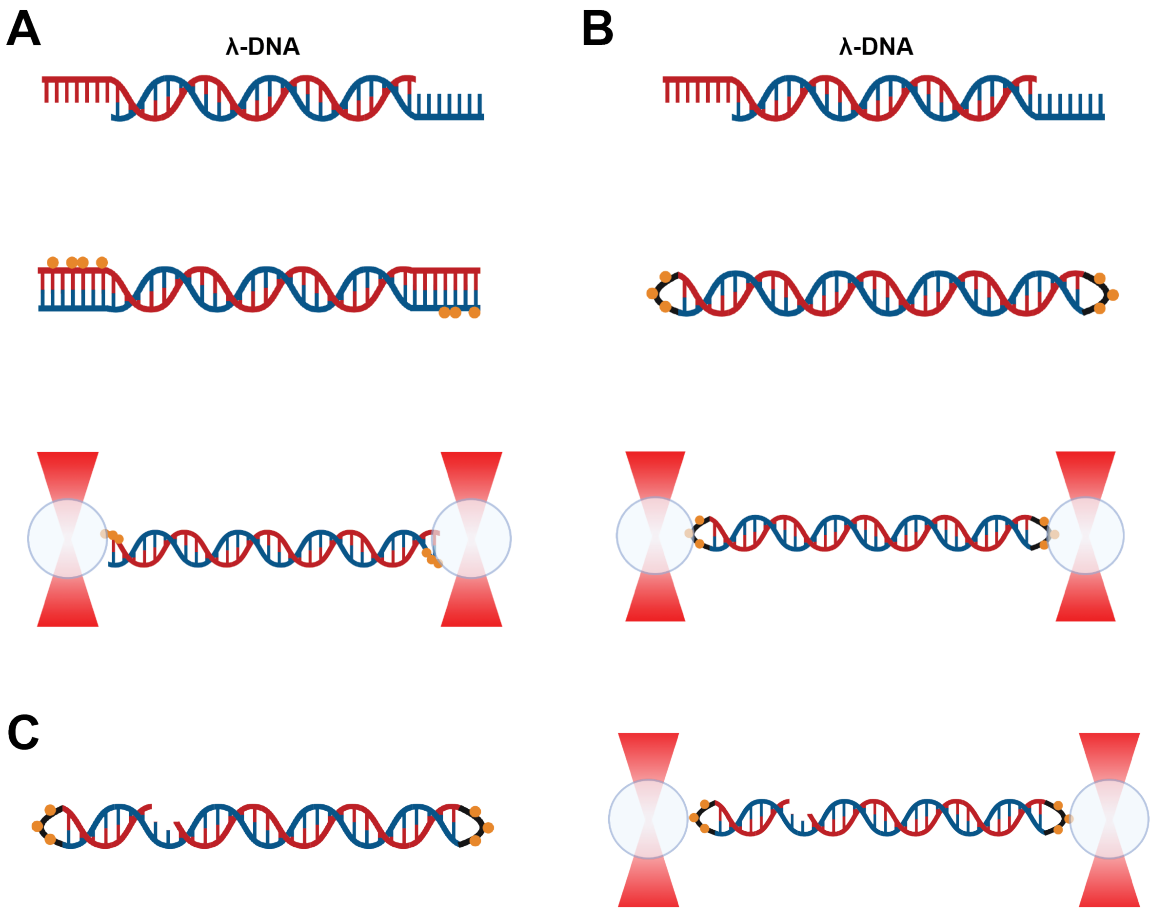


Figure 3.3: **DNA Constructs for Optical Tweezers.** (A) Formation of unconstrained biotinylated λ-DNA held between optical traps. (B) Formation of tortionally constrained biotinylated λ-DNA held between optical traps using the end-cap system. (C) Formation of gapped biotinylated -DNA made from tortionally constrained λ-DNA and nicked via Cas9 nickase held between optical traps. Biotin molecules shown in orange circles

3.5.2 Generation of hybrid ss-/ds-DNA constructs

Hybrid λ-DNA was generated by nicking tortionally constrained λ-DNA twice in specific locations using a Cas9 D10A nickase (IDT) (King et al., 2019; Belan et al., 2021).

Tortionally constrained DNA was generated by addition of biotinylated end-caps onto the overhangs of λ-DNA. End-caps were purified via PAGE and HPLC purified as described previously (Figure 3.3B and Table 3.1).

Table 3.1: DNA Sequences for tortionally constrained DNA

Name	Sequence (5' to 3')	Modification
End Cap 1	AGGTC GCCGC CCGGA GTTGA ACG**TT**	**T** = Biotin dT
	T**TTT**A CGTTC AACTC C	5' Phosphate
End Cap 2	GGGCG GCGAC CTCAA GTTGG ACAA**T**	**T** = Biotin dT
	T**TTTT** TGTCC AACTT G	5' Phosphate

First, λ-DNA (Life Technologies Limited, 41.6 µL at 300 ng/µL) was phosphorylated in T4 DNA ligase buffer (5.5 µL 10X) with T4 PNK (0.5 µL, 10U/µL) to a final volume of 55 µL. The end caps were also phosphorylated (2 µL at 100 µM) in T4 ligase buffer with T4 PNK to a final volume of 20 µL. All reactions were incubated for 1 hour at 37°C.

Next, phosphorylated end-cap 1 (1 µL at 10 µM) was added to the phosphorylated λ-DNA (56 µL at 14 nM) in T4 DNA ligase buffer (50 µL) topped up with ddH2O (400 µL). To aid in annealing, the mixture was heated to 80°C for 5 minutes before being placed directly on ice. T4 DNA ligase (NEB, 0.5 µL to final 400U) was then added and incubated for 2 hours at 25°C (or 12 hours at 16°C). Phosphorylated end-cap 2 was subsequently added (10 µL at 10 µM), heated to 80°C as before, more T4 DNA ligase added (4 µL) and then incubated for 2 hours at 25°C. Tortionally constrained DNA can be stored for up to 2 weeks at 4°C or be directly used in the subsequent nicking step.

Two Cas9 nickase RNP complexes must be generated, one for each site to be nicked forming the ssDNA gap. Cas9 is mutated (D10A) so one of the two nuclease domains is inactivated, allowing only one nick to be generated when coupled to target DNA (Chiang et al., 2016). crRNA (targeting specific sequence) is annealed to tracrRNA in T50 buffer as described before. Cas9 D10A nickase was mixed (5 µL at 5 µM) to each of the guide

RNAs in a 1:1 ratio. To nick the tortionally constrained DNA, hybrid-DNA (175 µL) was mixed with both Cas9 nickase RNPs (2 µL) in Cas9 digestion buffer (20 µL at 10X, 200 mM HEPES pH7.5, 1M NaCl 50 mM $MgCl_2$, 1 mM EDTA)(Figure 3.3C). The reaction was allowed to continue for 1 hour at 37°C before being stopped upon addition of 10 mM EDTA and stored at 4°C until used.

3.5.3 Ligation method to generate long TEC constructs

DNA handles with a 4 nucleotide overhang were obtained by restriction digest of pBR322 plasmid. pBR322 was cut with AatII and BmtI-HF to form handle 1 (H1) and with AvaI and StyI to generate handle 2 (H2). pBR322 (2 µL at 1 µg/µL) was mixed with the two required enzymes (2 µL at 10 U/µL each) and 1X Cutsmart buffer (5 µL at 10X) to a final volume of 50 µL. DNA was incubated at 37°C for 30 minutes and purified using the QIAquick PCR purification kit (Qiagen, 28104) and checked via 0.8% agarose gel electrophoresis. The intermediate DNA oligonucleotide (10 µL at 10 nM), RNA primer (10 µL at 10 nM), DNA handles (5µL, total 100 ng each), 10X T4 DNA ligase buffer (5 µL) and T4 DNA ligase (1 µL) were mixed and incubated at 16°C overnight. Ligated samples were checked via 0.8% agarose gel electrophoresis.

3.5.4 Autosticky PCR method to generate long TEC constructs

DNA handles with a 5' overhangs were generated via 'autosticky PCR', where the PCR primer contains the overhang of choice, followed by an abasic site and then the annealing region (Table 3.1). One primer also contained 3x biotin groups. A master mix containing λ-DNA (10 µL at 10 ng/µL), primer 1 and primer 2 (10 µL at 25 µM each), 5X HF buffer (100 µL), dNTP mix (10 µL at 10 mM of each dATP, dGTP, dCTP and dTTP) and

Phusion DNA Polymerase (5 µL at 2 u/µL) were mixed, topped up to 500µL with ddH_2O and split into 50 µL reactions. These were placed in a thermocycler using the cycling conditions in Table 3.2. Handles were purified using QIAquick PCR purification kit and checked via 0.8% agarose gel electrophoresis. TECs were formed as per Section 3.2.4. The TEC and two handles were ligated together and incubated at 16°C for 12 hours, before checking for ligation efficiency on a 0.8% agarose gel (Figure 3.4).

Table 3.2: Thermocycler Conditions for autosticky PCR

Step	Temperature °C	Time	No. of cycles
Initial Denaturation	98	30	1
Denaturation	98	10s	
Annealing	64.3	30s	30
Extension	72	30s/Kb	
Final Extension	72	10 min	1
Hold	4	Infinite	

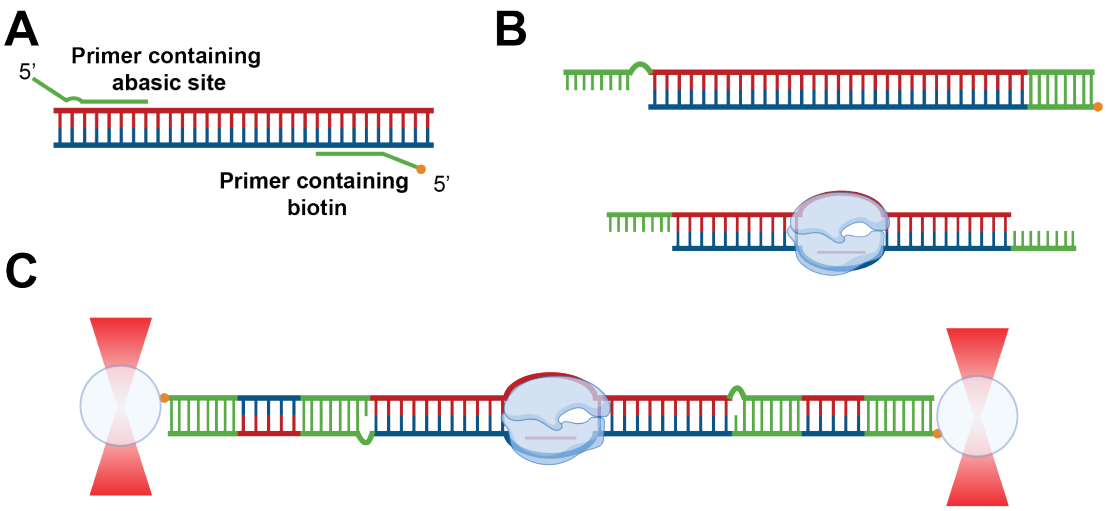


Figure 3.4: **Schematic of DNA constructs generated via autosticky PCR.** (A) Fragments of λ-DNA are replicated using a the autosticky PCR method. (B) λ-DNA fragments containing overhangs (green) are mixed and ligated to an intermediate TEC containing complimentary overhangs (green). (C) Full complex of handles and ligated intermediate shown held between optical traps. Biotin molecules shown in orange circles.

3.5.5 Optical tweezers preparation

Optical tweezer experiments were performed using a C-trap (LUMICKS) containing a microfluidic system that allows for laminar flow of up to 5 different channels at once. Before use, the fluidic system is first cleaned using bleach (1mL of 5% (NaClO)) followed with 1 mL each of ddH_2O, sodium thiosulfate (25 mM, to neutralise the bleach) and ddH_2O again. This was carried out for 15 minutes at each step. Once cleaned, the channels were passivated in Pluronic F128 (0.5% w/v in PBS) and also BSA (0.1% w/v in PBS).

3.5.6 Optical tweezers assay

Experiments varied depending on the type of DNA/protein needed to test but the channels always contained similar contents. Channel 1 always contained the streptavidin coated beads (4.8/1.78 µm, 0.005% w/v, SPHEROTECH) and channel 2 the specific DNA construct that was to be used (diluted to approx. 50 ng/mL). Channel 3 contained 1X TB supplemented with imaging mix (0.2mg/mL BSA, 125 mM PCA, 10 µM PCD, Trolox). Channel 4 and 5 then varied depending on the experiment to be carried out. This included RNAPII (5-50 nM) and Cy5 labelled RNA primers (10 - 1000 nM). For channel layout see Figure 3.5.

3.5.7 Optical tweezers analysis

Data from the C-trap was exported as '.h5' files and are further analysed using a series of home built Python Jupyter notebooks (https://github.com/singlemoleculegroup). This was done in the lumicks.pylake environment which contains the necessary packages to extract and analyse FD curves, scans and kymographs that are generated from the C-

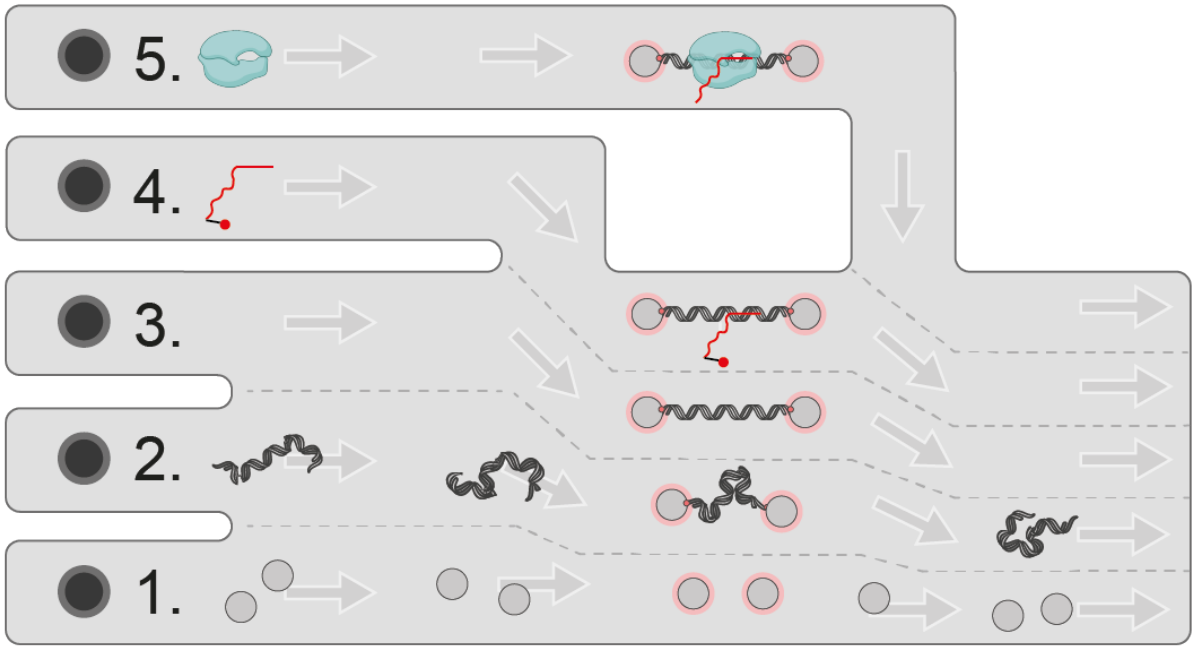


Figure 3.5: **Schematic of C-trap microfluidic system.** 5 channel microfluidic chip used within the C-trap to allow for laminar flow between channels. (1) Bead Channel, (2) DNA Channel, (3) Buffer Channel, (4 & 5) Protein/RNA and experimental Channels. Adapted from (Losito et al., 2021)

trap. Further analysis on scans and kymographs were carried out in FIJI to determine fluorescence intensity. Using known pixel size of the image we could map back any fluorescent spots to a specific gnomic location along the DNA using Gaussian fits in IGOR pro (wavemetrics).

Bibliography

O. Belan, G. Moore, A. Kaczmarczyk, M. D. Newton, R. Anand, S. J. Boulton, and D. S. Rueda. Generation of versatile ss-dsDNA hybrid substrates for single-molecule analysis. *STAR Protocols*, 2(2):100588, 6 2021. ISSN 2666-1667. doi: 10.1016/J.XPRO.2021.100588.

C. Bernecky, F. Herzog, W. Baumeister, J. M. Plitzko, and P. Cramer. Structure of transcribing mammalian RNA polymerase II. *Nature*, 529(7587):551–554, 2016. ISSN 14764687. doi: 10.1038/nature16482. URL `http://dx.doi.org/10.1038/nature16482`.

S. C. Blanchard, R. L. Gonzalez, H. D. Kim, S. Chu, and J. D. Puglisi. tRNA selection and kinetic proofreading in translation. *Nature Structural and Molecular Biology*, 11 (10):1008–1014, 10 2004. ISSN 15459993. doi: 10.1038/nsmb831. URL `http://www.nature.com/natstructmolbiol`.

T. W. W. Chiang, C. Le Sage, D. Larrieu, M. Demir, and S. P. Jackson. CRISPR-Cas9D10A nickase-based genotypic and phenotypic screening to enhance genome editing. *Scientific Reports 2016 6:1*, 6(1):1–17, 4 2016. ISSN 2045-2322. doi: 10.1038/srep24356. URL `https://www.nature.com/articles/srep24356`.

Glen Research. Deprotection Guide. Technical report. URL `https://www.glenresearch.com/media/productattach/import/`.

G. A. King, F. Burla, E. J. Peterman, and G. J. Wuite. Supercoiling DNA optically. *Proceedings of the National Academy of Sciences of the United States of America*, 116(52):26534–26539, 12 2019. ISSN 10916490. doi: 10.1073/PNAS.1908826116/-/DCSUPPLEMENTAL. URL `/pmc/articles/PMC6936703//pmc/articles/PMC6936703/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC6936703/`.

R. Lamichhane, A. Solem, W. Black, and D. Rueda. Single-molecule FRET of protein–nucleic acid and protein–protein complexes: Surface passivation and immobilization. *Methods*, 52(2):192–200, 2010. doi: 10.1016/j.ymeth.2010.06.010.

M. Losito, Q. M. Smith, M. D. Newton, M. E. Cuomo, and D. S. Rueda. Cas12a target search and cleavage on force-stretched DNA. *Physical Chemistry Chemical Physics*, 23(47):26640–26644, 12 2021. ISSN 1463-9084. doi: 10.1039/D1CP03408A. URL `https://pubs.rsc.org/en/content/articlehtml/2021/cp/d1cp03408ahttps://pubs.rsc.org/en/content/articlelanding/2021/cp/d1cp03408a`.

J. Oh, J. Xu, J. Chong, and D. Wang. Structural and biochemical analysis of DNA lesion-induced RNA polymerase II arrest. *Methods*, (October 2018):0–1, 2019. ISSN 10462023. doi: 10.1016/j.ymeth.2019.02.019. URL `https://linkinghub.elsevier.com/retrieve/pii/S1046202318303001`.

G. Senavirathne, J. G. Bertram, M. Jaszczur, K. R. Chaurasiya, P. Pham, C. H. Mak, M. F. Goodman, and D. Rueda. Activation-induced deoxycytidine deaminase (AID) co-transcriptional scanning at single-molecule resolution. *Nature Communications*, 6, 2015. ISSN 20411723. doi: 10.1038/ncomms10209.

J.-W. Van De Meent, J. E. Bronson, F. Wood, R. L. Gonzalez, and C. H. Wiggins. Hierarchically-coupled hidden Markov models for learning kinetic rates from single-molecule data. *Proceedings of the International Conference on Machine Learning*, 28 (2):361–369, 2013.

Chapter 4

Visualisation of RNAPII transcription elongation dynamics via single-molecule FRET

4.1 Abstract

Transcription elongation is a highly complex process that has been probed for decades using a variety of techniques. Although ensemble and structural assays have provided invaluable insight into this process, they often lack information about the heterogeneity and dynamics of transcription. Single-molecule techniques can be employed to resolve and detect dynamics along the transcription pathway.

Current single-molecule assays either require labelling of proteins or mainly focus on the initiation stage of transcription. In order to monitor the elongation phase specifically at the single-molecule level without labelling RNAPII, a novel FRET based assay was developed.

Using this assay, I was able to form stable and efficient transcription elongation complexes that allowed me to monitor transcription elongation via RNAPII. I was able to observe that RNAPII transcribes in a predominantly unidirectional manner, with occasional pauses as is expected for unhindered polymerases. These pausing events are even more prevalent and longer when encountering AT-rich sequences, which have been shown to induce premature termination of transcription. I further show that this premature termination is not due to the template strand sequence and hence any RNA hairpin formation but instead suggests the non-template strand plays a more important role in this process.

4.2 Results

4.2.1 Visualisation of FRET dependent on RNA primer length

In order to visualise transcription elongation specifically, a TEC must be formed, which allows the bypass of the complex initiation stage (Nogales et al., 2017). TECs consist of 4 main components, a template strand (TS), an RNA primer that will then act as a landing pad for RNAPII to bind to and finally a complimentary non-template strand (NTS). These have been utilised effectively in ensemble and structural assays (Oh et al., 2019; Bernecky et al., 2017). In order to study these at the single-molecule level, fluorescent dyes had to be attached to monitor transcription in real time using smFRET. The acceptor dye was placed on the 5' end of the RNA, so when the polymerase transcribes the TS, the RNA will increase in length but the 5' dye will remain present (Figure 4.4 and Table 4.1). In this combination, the 5' Cy5 dye on the RNA will begin within the exit tunnel of RNAPII and as the polymerase transcribes the template strand, the 5' RNA will thread through the tunnel until it reaches a length greater then 19nt. At this point the 5' RNA will be outside of the exit tunnel. As mentioned in Section 1.6, one of the main aims of designing this assay without modifying the protein was to use higher eukaryotic samples and the following optimisation was done using mammalian RNAPII.

Different positions for the location of the donor fluorophore were tested on both the TS and NTS, placed upstream of the RNA binding site by 2, 4 or 8 nt in order to achieve a high FRET efficiency in the pre-transcription state (before NTPs are added) (Table 4.1). This allows for maximum visualisation of transcription over a longer distance (up to 16 nt). In all cases good immobilisation of the TEC to the slide surface was observed where a high density of donor fluorophores were detected (approx. 300 molecules) (Figure 4.1A).

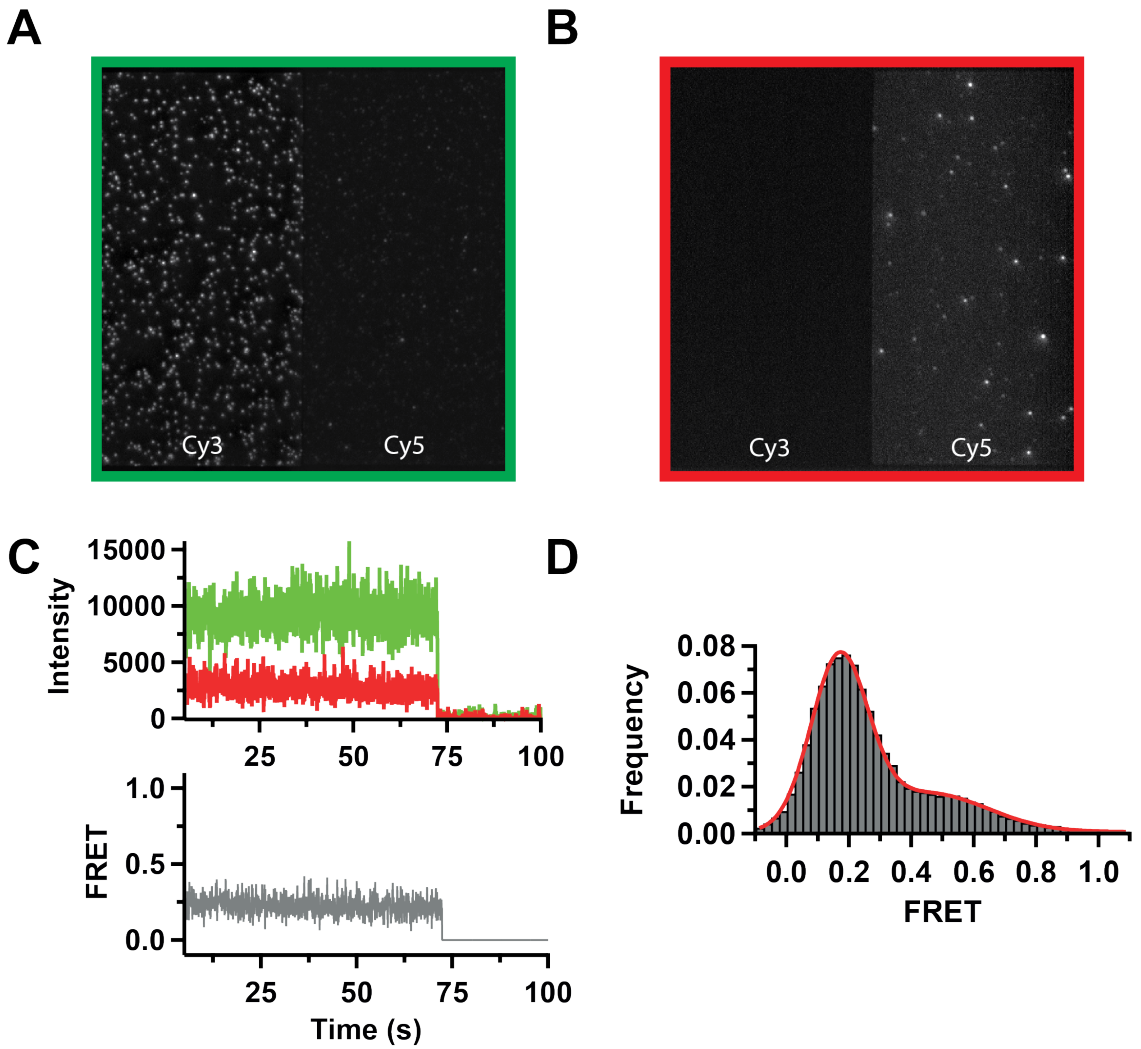


Figure 4.1: **Poor FRET efficiencies detected within TEC with long RNA Primer .** (A) TIRF image showing Cy3 labelled TS immobilised to the surface but no FRET molecules in Cy5 channel. 4mW Cy3 Excitation. (B) TIRF image showing Cy5 labelled RNA immobilised to the surface via annealing with complimentary TS (A) in Cy5 channel. 4mW Cy5 Excitation (A - B) Cy3 channel on left and Cy5 channel right. (C) (top panel) Example trace of donor (Cy3) and acceptor (Cy5) intensity over time in absence of NTPs showing static intensity over time in both acceptor (red) and donor (green). (bottom panel) Example FRET trace of RNAPII in absence of NTPs showing static low FRET, calculated using equation 2.7 from data in (C).(D) Histogram of RNAPII in absence of NTPs. N = 75, peak fitted to double Gaussian fit

However, minimal FRET was detected in all of the constructs using the RNA primer that was 21 nt in length. I confirmed that this lack of FRET was not due to an absence of the Cy5-labelled RNA primer, which was seen bound to the surface by direct Cy5 excitation (647 nm), although the number of RNA molecules present was significantly less than the number of corresponding TS molecules (10%, $\sim$ 30) (Figure 4.1B). No FRET was observed when the donor fluorophore was placed 8 or 4 nt upstream of the RNA on both

the TS and NTS, with FRET only being observed when placed 2 nt upstream of the RNA primer. In this conformation >80% of molecules exhibited an undesired low FRET state of 0.2 suggesting the dyes were far apart (Figure 4.1C&D).

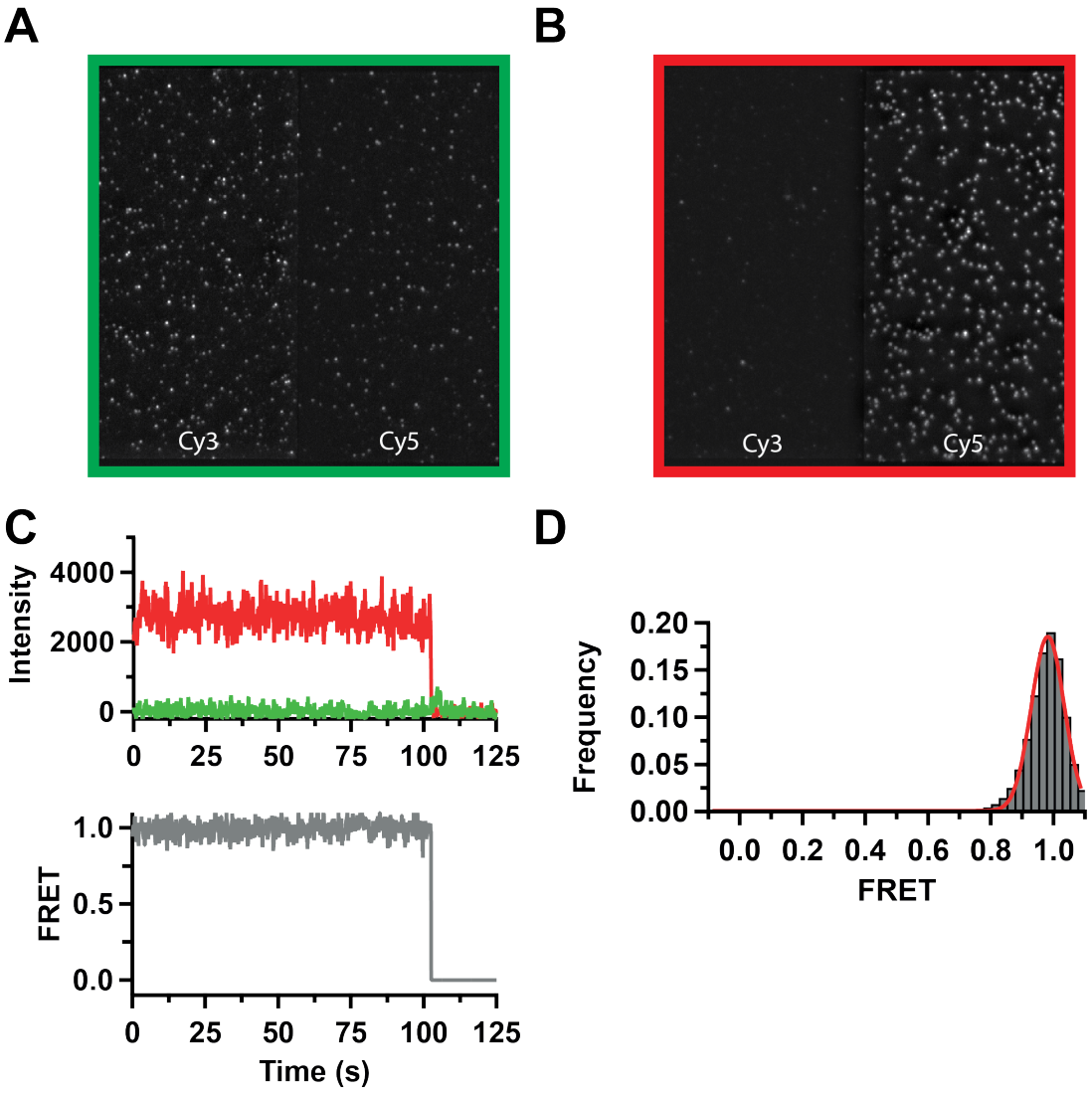


Figure 4.2: **Decreasing RNA length increases FRET efficiency.** (A) TIRF image showing Cy3 labelled TS immobilised to the surface and a high number of FRET molecules in Cy5 channel. 4mW Cy3 Excitation. (B) TIRF image showing Cy5 labelled RNA immobilised to the surface via annealing with complimentary TS (A) in Cy5 channel. 4mW Cy5 Excitation (A - B) Cy3 channel on left and Cy5 channel right. (C) (top panel) Example trace of donor (Cy3) and acceptor (Cy5) intensity over time in absence of NTPs showing static intensity over time in both acceptor (red) and donor (green). (bottom panel) Example FRET trace of RNAPII in absence of NTPs showing static high FRET, calculated using equation 2.7 from data in (C).(D) Histogram of RNAPII in absence of NTPs. N = 180, peak fitted to Gaussian fit

As a high amount of the donor fluorophore was present on the surface and the RNA signal was also present but I observed no FRET, it was hypothesised that the low FRET value

was possibly due to the long RNA length. If the RNA is too long in the pre-transcription state then the acceptor fluorophore will be too far away from the donor fluorophore for high FRET transfer. To test this, the RNA length was decreased from 21 to 12 nucleotides and the FRET measured with the donor fluorophore placed 2 nt upstream of the RNA primer, as this produced the most promising results before (Table 4.1). In this new conformation, not only did I observe a high immobilisation efficiency of the TS (approx. 300 molecules), but also a high number of molecules in the Cy5 channel upon direct excitation was observed suggesting that over 90% of the TS molecules immobilised also contained a corresponding RNA primer (Figure 4.2A). Furthermore, under donor excitation a high number of Cy5 signal was visualised, suggesting the positions of these dyes induced FRET (Figure 4.2B). Further analysis of these molecules showed a high FRET state of 0.95 and minimal dynamics were observed in this conformation (Figure 4.2C&D). Over 90% of bound TS molecules showed a corresponding NTS signature with a similar number if molecules being detected, suggesting that the FRET I observe arises due to full TECs and not any intermediate complexes (Figure 4.11).

4.2.2 Visualisation of RNAPII dynamics using smFRET

Based on the optimisation work done in Sections 4.2.1 a TEC construct was developed that incorporated a 12 nt RNA primer (9 nt annealing region with a 3 nt overhang, labelled with 5' Cy5), a TS containing a Cy3 fluorophore labelled adjacent to the RNA primer binding position and a fully complimentary NTS (Figure 4.3). As mentioned above, in this combination the 5' Cy5 fluorophore will be inside the RNAPII exit tunnel. It should be noted that for the rest of the thesis, yeast RNAPII was used instead of mammalian RNAPII. Although the assay was developed initially for mammalian samples, due

to COVID and inability to source suitable calf thymus I was unable to purify clean mammalian RNAPII and so yeast RNAPII was used instead. Purification of yeast RNAPII is much more established and was gifted to us by our collaborators (Svejstrup group).

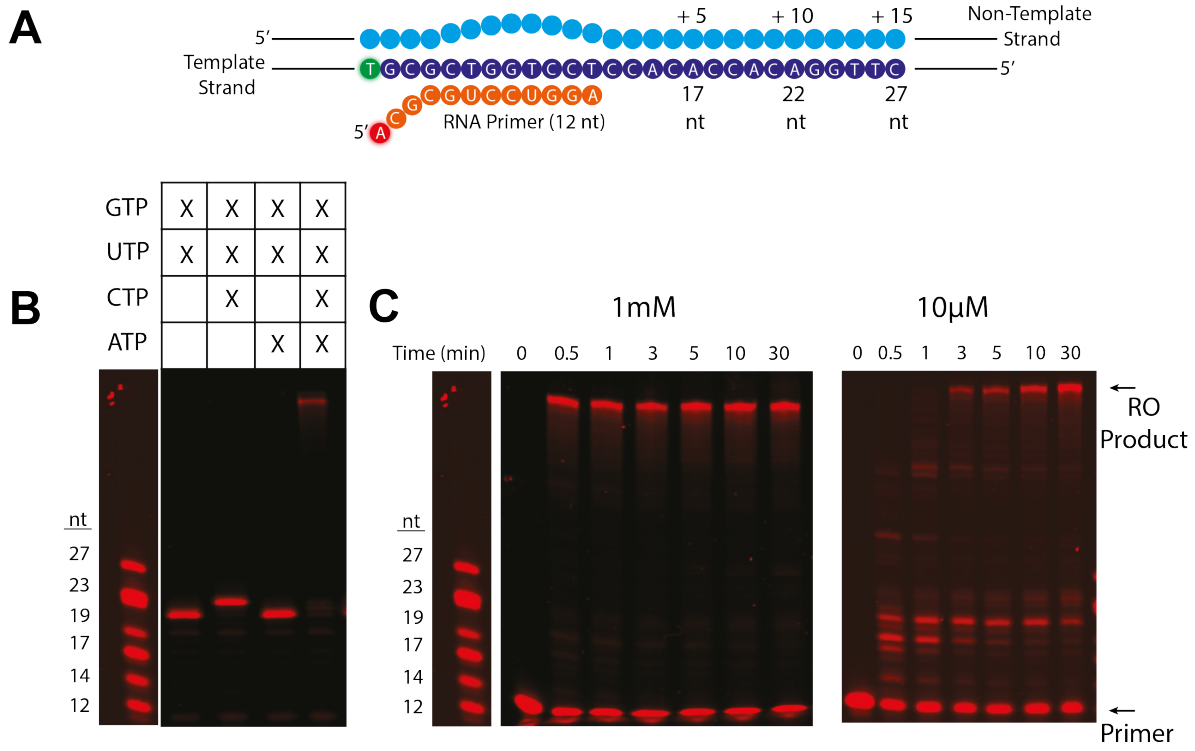


Figure 4.3: **RNAPII transcribes in an NTP dependent manner.** (A) Sequence of template strand (TS) and RNA primer where RNAPII will bind to and transcribe. Green base = Cy3 dye; Red base = Cy5 dye. NTS is fully complimentary but when RNAPII is present, a transcription bubble is formed permitting transcription. (B) Denaturing PAGE showing RNA extension with subsets of NTPs at 1 mM [NTP], measured via Cy5 fluorescence. (C) Denaturing PAGE showing RNA extension over time from primer (lower bands, 12 nucleotides) to full length transcript (top bands, 69 nucleotides) at 1 mM (left) and 10µM (right) [NTP], measured via Cy5 fluorescence.

I started by visualising the dynamics of RNAPII while transcribing an undamaged DNA template (Figure 4.3A). I first tested the bulk activity of the labelled TEC using the undamaged DNA template (Figure 4.3B-C). If fully functional, the presence of all 4 RNA nucleotide bases (ATP, GTP, CTP and UTP) will induce the polymerase to transcribe and synthesise the nascent RNA based upon the TS sequence, causing an increase in RNA length. Furthermore, because of the sequence of the template I could control the

exact number of nucleotides transcribed. Addition of only UTP and GTP permitted transcription of exactly 10 nucleotides due to the presence of the first stop region (GG), addition of CTP allows transcription of a further 2 nucleotides to a second stop region (AA), while addition of ATP allowed full transcription to occur (Figure 4.3B and Table 4.2). At saturating [NTP] (1 mM), RNAPII showed complete transcription within 30 s, which increased to 3 min when [NTP] was decreased to 10 µM showing a clear NTP dependence (Figure 4.3C).

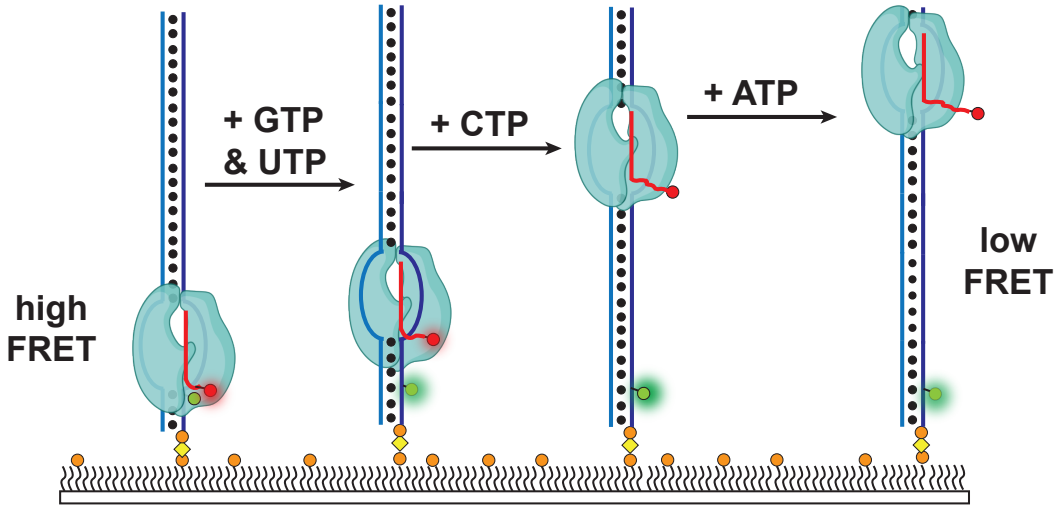


Figure 4.4: **Schematic of smFRET TIRF based assay.** smFRET set up of immobilised TEC to the biotinylated-PEG surface via neutravidin (yellow diamond). Cy3 (green circle) was labelled on the TS while Cy5 (red circle) on the 5' growing RNA chain (blue). RNAPII (blue) bound to transcription bubble which can transcribe upon addition of NTP's leading to a decrease in observed FRET as polymerase transcribes further from start site.

Once validated via the *in vitro* bulk assay, the transcription process was further probed using the smFRET assay. TEC's were assembled, surface immobilised and fluorescently labelled with an immobile donor fluorophore (Cy3) on the template strand and with a FRET acceptor (Cy5) on the RNA primer 5' end (Figure 4.4). In this initial configuration (pre-transcription state), the FRET pairs are expected to be in close proximity, yielding static high FRET trajectories (Figure 4.4). In the presence of NTPs, RNAPII will insert the complimentary NTP into the growing RNA chain defined by the template strand

sequence. In turn, the RNAPII will displace the growing 5' RNA chain away from the slide surface and further from the immobile FRET donor. Thus, increasing the distance between the FRET pairs resulting in an observable decrease in FRET (Figure 4.4). Furthermore, as established in the ensemble assay, this process is very fast and so to capture real time dynamics and potential intermediates, a limiting [NTP] was used to slow down the reaction (10 µM) rather than the 1 mM often used in ensemble assays.

In the absence of NTPs, 95% of FRET trajectories exhibit high static FRET (0.95) indicating that the RNAPII remains stably bound at the primer:template complex, unable to move (Figure 4.5).

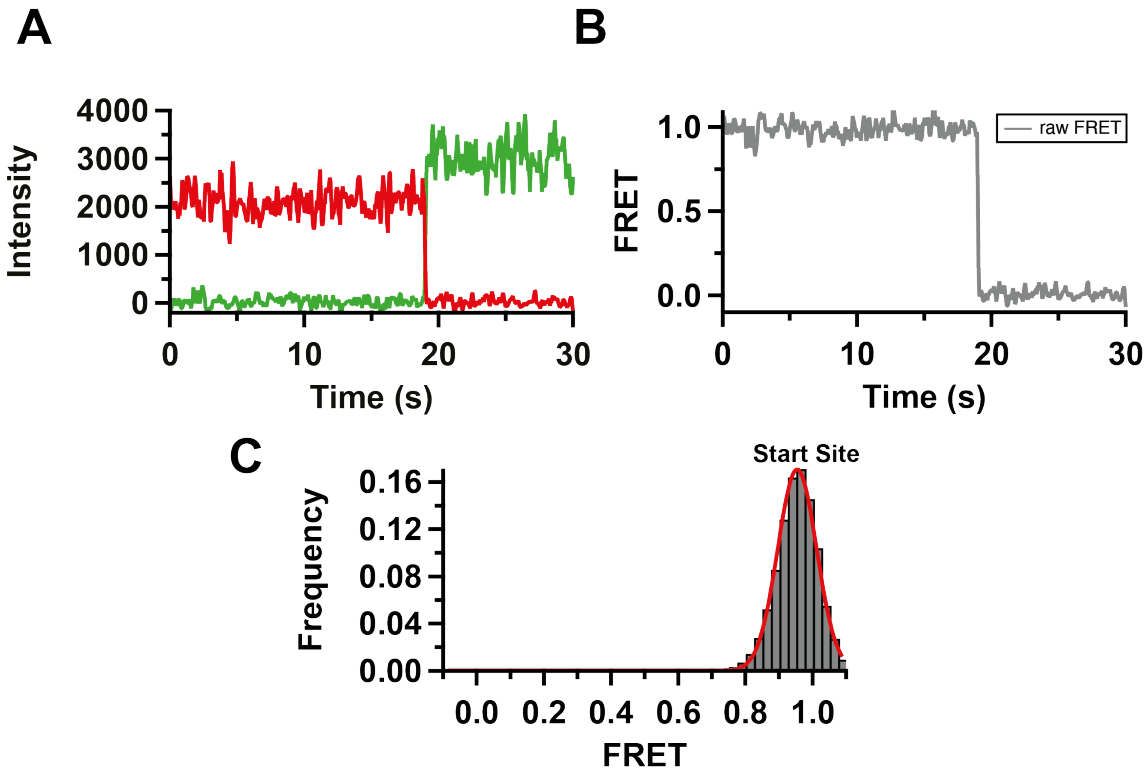


Figure 4.5: **RNAPII remains in pre-transcription state without NTPs.** (A) Example trace of donor (Cy3) and acceptor (Cy5) intensity over time in absence of NTPs showing static intensity over time in both acceptor (red) and donor (green). (B) Example FRET trace of RNAPII in absence of NTPs showing static FRET, calculated using equation 2.7 from data in (A).(C) Histogram of RNAPII in absence of NTPs. N = 131, peak fitted to Gaussian fit.

Addition of all 4 NTPs, allowing complete transcription, showed a significant change in

the observed FRET traces and ratio. Now, trajectories revealed a step-wise decrease in the observed FRET ratios (Figure 4.6A). The observed FRET decreases unidirectionally from the pre-transcription state (0.95) to a complete loss of FRET within ∼1 min (Figure 4.6A). A histogram of the raw FRET states revealed several transcription intermediates between 0.6 and 0 FRET (Figure 4.6B). To further probe the observed FRET intermediates, I used Hidden-Markov Modelling (HMM)(Van De Meent et al., 2013).

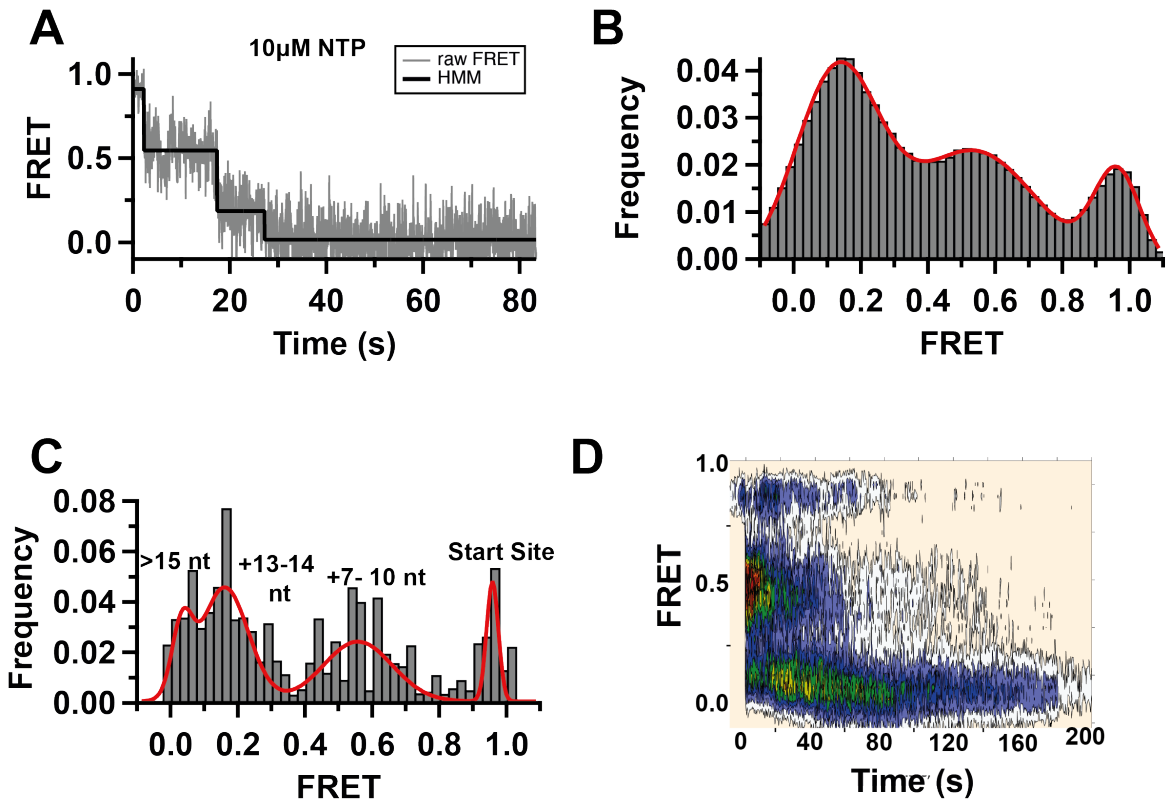


Figure 4.6: **RNAPII transcription leads to loss of FRET.** (A) Example FRET trace of RNAPII in presence of 10µM NTPs showing predominately unidirectional transcription, grey trace shows raw data, black trace shows idealised FRET states detected via HMM analysis. NTPs were added and FRET data was captured in real-time for 5 minutes. (B) Histogram of raw FRET of RNAPII in presence of all 4 NTPs. , peaks fitted to a triple Gaussian. (C) Histogram of idealised FRET of RNAPII in presence of all 4 NTPs. Peaks fitted to a quadruple Gaussian. (D) Post synchronisation histogram showing multiple RNAPII molecules transcribing over time from high FRET to low FRET with minimal pauses during transcription. (A - D) N = 128

A histogram of the resulting idealised FRET states identified at least three intermediates at 0.60, 0.15 and 0.03 (Figure 4.6A & C). To better determine the location of the observed intermediates, I performed an empirical FRET calibration with known RNA

lengths (Figure 4.12). This was done by annealing RNAs of different lengths at known positions along the template strand. In turn, RNAPII was bound at a known position downstream of the original binding site, for example + 2nt from the start site, and the observed FRET could be measured in this position. This was then repeated several times over a range of + 2 - +15 nt, where a poly-U tail was added to the 5' end to mimic the increase in RNA length that would be expected if RNAPII had 'transcribed' to that position (Figure 4.12). The observed FRET can then be plotted against the known position and a empirical calibration is therefore determined. Based on this calibration, I estimate that the first intermediate (0.6) to correspond to a position of +7 - 10 nt transcribed, the second intermediate (0.15) corresponding to +13 -14 nt transcribed, and the final intermediate (0.03) corresponding to addition of >15 nt (Figure 4.6C). Incorporation of further NTPs yield FRET values below our experimental resolution of ∼0.01 FRET. Post-synchronisation the transcribing FRET trajectories at the initial transcription event further confirmed the presence of the same intermediates (Blanchard et al., 2004). The resulting post-synchronization histogram (PSH) showed multiple FRET densities ranging from 0.95 to 0, confirming RNAPII transcribed the template in a unidirectional manner (Figure 4.6D). Consistent with the HMM analysis, the same FRET intermediates were observed in both the HMM and PSH analysis, with pauses ranging from 20 - 100 s at 0.6 and 0.15 (Figure 4.6D). Eventually, the FRET drops to zero indicating run-off transcription or acceptor photobleaching. Overall, this data shows that the smFRET assay has allowed me to track a single elongating RNAPII molecule on a given template DNA up to 15 nt, resolving several transcription intermediates along the DNA pathway.

4.2.3 AT-rich DNA induces premature disengagement from DNA

The smFRET assay developed above was further utilised to probe the effect of the DNA sequence on RNAPII, in collaboration with the Svejstrup group. The sequence used above (Figure 4.3) utilised a mix of all 4 bases with no AT-rich regions, which have been known to interrupt biological processes (Sigurdsson et al., 2010). The sequence of the TS and corresponding NTS were changed to contain an AT-rich region immediately downstream of the start site (Figure 4.7A Table 4.3). This sequence was designed as such based on *in vivo* data from the Svejstrup lab, and is based on the Cyc1 gene sequence. In order to probe the dissociation of the polymerase and nascent RNA, the biotin on the complex was utilised in the ensemble assay. This allowed observation of two different fractions, the RNA bound to the bead (B) or RNA in solution (S) that has dissociated with RNAPII. In normal transcription, the polymerase should transcribe the full template unhindered and not dissociate until the end of the template is reached and it runs off. Addition of UTP only should permit transcription of specifically 5 nucleotides due to a presence of G in the TS. When the UTP is washed away and replaced with CTP, this would allow progression past this point by 1 nucleotide but is then stalled by the next AT-tract region. Washing away the CTP and replacing with UTP should then allow transcription of a further 8 nucleotides (to a second G stop position) (Table 4.3). Normal transcription behaviour is observed upon addition of UTP and CTP, where the first 6 nucleotides are transcribed, with an RNA band of corresponding length found in the B fraction, but nothing in the S fraction (Figure 4.7B). This shows that RNAPII pauses at the relative stop positions waiting for the presence of the next NTP, but does not fall off. However, addition of more UTP showed perturbation of transcription when attempting to transcribe the next AT-rich region. The majority of DNA is paused at a site 1-2 nt downstream of the 1st G

stop position and only a small amount makes it to the expected length of 26 nucleotides. Furthermore, a band of the same length now appears in the S fraction. This can only appear if RNAPII falls off the construct into solution along with the RNA (Figure 4.7B). The presence of this band in the S fraction only occurs when transcribing the second AT-rich region, and not the first suggesting this region induces premature termination of RNAPII transcription.

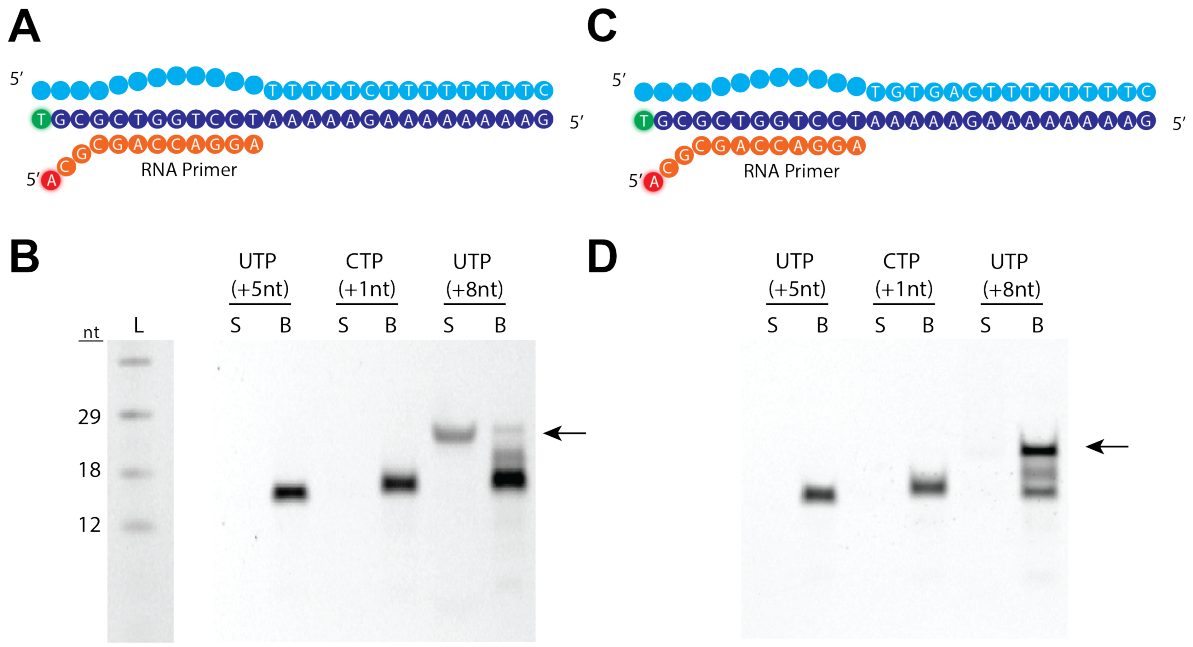


Figure 4.7: **RNAPII disengages on AT-rich containing sequences.** (A) Sequence of template strand (TS, dark blue) and RNA primer (orange) with AT-rich TS and fully complimentary NTS (light blue). Green base = Cy3 dye; Red base = Cy5 dye. (B) Denaturing PAGE showing RNA extension from primer (12 nucleotides) to different length products upon addition of subset of NTPs where construct shows pausing at AT-rich sequence (arrow), 1 mM NTPs, measured via Cy5 fluorescence. S = supernatent, B = Beads. (C) Sequence of template strand (TS, dark blue) and RNA primer (orange) with AT-rich TS but not fully complimentary NTS (light blue). Green base = Cy3 dye; Red base = Cy5 dye. (D) Denaturing PAGE showing RNA extension from primer (12 nucleotides) to different length products upon addition of subset of NTPs where construct shows no pausing at AT-rich sequence (arrow), 1 mM NTPs, measured via Cy5 fluorescence. S = supernatent, B = Beads. Assay carried out by Zhong Han.

4.2.4 AT-rich DNA induces RNAPII pausing before disengagement from DNA

To further validate the sequence-dependent intrinsic transcription termination observed in the bulk experiments, I utilised the smFRET assay above to visualise the dynamics of a single RNAPII complex on a DNA molecule containing the intrinsic termination sequence (Figure 4.4). Transcription elongation complexes (TECs) were assembled with a FRET donor (Cy3) on the template strand (TS) and a FRET acceptor (Cy5) on the 5'-end of the RNA primer as before. Like above, in the absence of nucleotides, the vast majority of FRET trajectories appear to be static at 0.95 FRET (95%, N = 105), indicating that the RNAPII remains stably bound to the primer:template complex, unable to transcribe as expected (Figure 4.8A).

In the presence of all nucleotides (UTP, CTP, GTP and ATP) the observed FRET traces changed dramatically, where the FRET decreased over time moving from the high FRET conformation (0.95) to lower FRET values ranging from 0.2 - 0.8. HMM analysis was carried out which showed that the polymerase followed a similar trend to that seen with the sequence with no AT-rich regions, the polymerase transitions through several intermediate states in a mostly unidirectional and step-wise manner with occasional backtracks, as expected for transcribing TECs (Nudler, 2012) (Figure 4.8C). RNAPII transitioned through at least 3 different intermediates from the high pre-transcription state (0.95) to an eventual loss of FRET as seen via the TDP (FRET = 0.84, 0.37 and 0.19) (Figure 4.9A). Although this followed a similar trend to the previous sequence used, where no AT-rich regions are present, a few differences were also found. Firstly, the number of molecules completing full transcription decreased when the AT-rich region was placed in the transcription pathway. Bulk analysis showed that over 80% of molecules did not

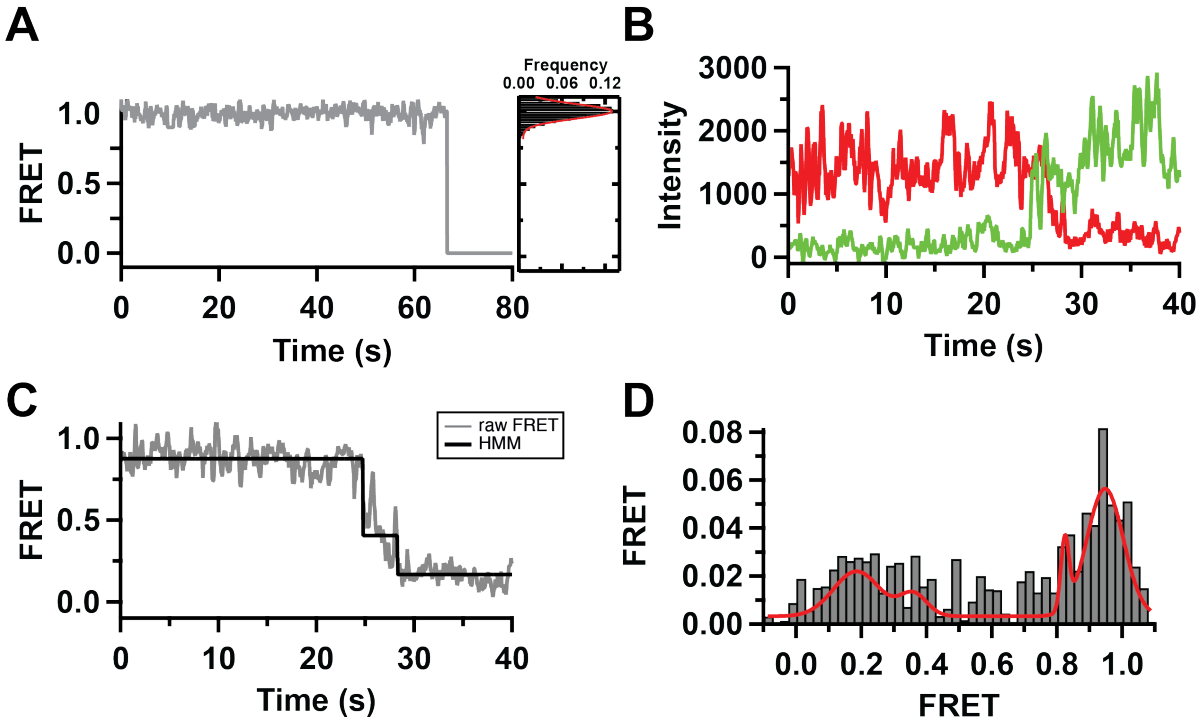


Figure 4.8: **AT-rich DNA perturbs transcription.** (A) Example FRET trace of RNAPII in absence of NTPs showing static FRET and histogram of RNAPII in absence of NTPs. N = 1XX, peak fitted to Gaussian fit. NTPs were added and FRET data was captured in real-time for 5 minutes. (B) Example trace of donor (Cy3) and acceptor (Cy5) intensity over time in absence of NTPs showing static intensity over time in both acceptor (red) and donor (green). (C) Example FRET trace of RNAPII in presence of 10µM NTPs showing decrease in FRET over time and pausing at FRET 0.2. grey trace shows raw data and black trace shows idealised FRET states detected via HMM analysis (D) Histogram of RNAPII in absence of NTPs. N = 108, peaks fitted to multi Gaussian fit.

progress into the second AT-rich region but stayed paused just before it (Figure 4.7). This was observed in the smFRET assay where a higher proportion of the observed FRET was found in the high FRET region (0.7 - 1.0) when transcribing on the AT-rich sequence compared to the sequence used above. These molecules also transitioned within this region for a significantly longer period of time, over 80s compared to under 50s before (Figure 4.9B). Secondly, a significant fraction of molecules (25%) exhibited a long pause at lower FRET values ranging from 0.1 - 0.4 when encountering the second AT-rich region. This long pause of more than 100s is not detected in the sequence without the AT-rich regions, where the lower FRET values show a gradual decrease over time as the polymerase transcribes through the template (Figure 4.9B). Using the empirical RNA ladder (Figure

4.12), I can determine that the 0.2 FRET pauses occur approximately the fourteenth nucleotide from the initial binding site, consistent with the dissociation site observed in the bulk biochemical experiments (Figure 4.7). Together, the smFRET data and the bulk biochemical data show that the RNAPII pauses seven nucleotides into the second AT-rich region, stalls and falls off the DNA unable to rectify this blockage leading to premature termination.

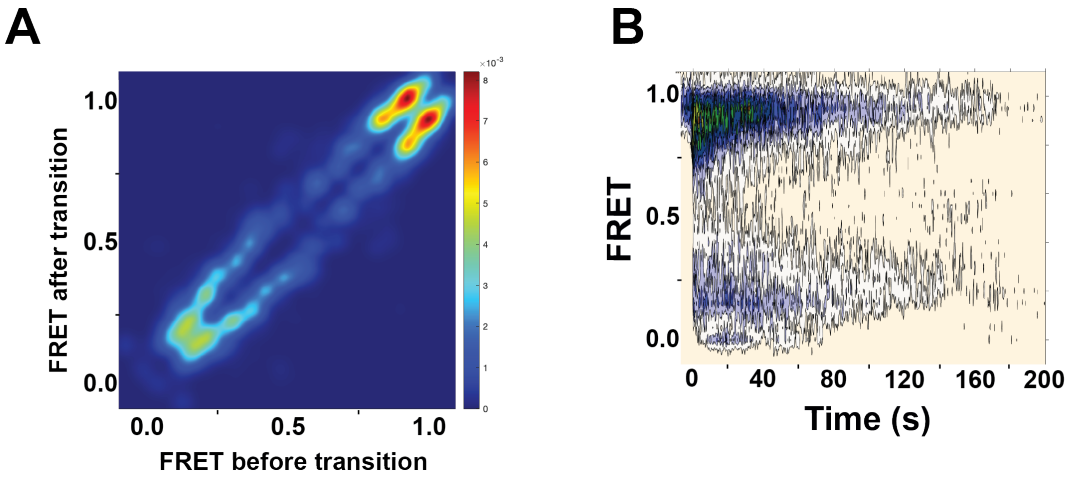


Figure 4.9: **AT-rich DNA induces RNAPII pausing.** (A) Transition density plot showing RNAPII transitioning through several intermediates along AT-rich containing DNA. (B) Post-synchronisation histogram of RNAPII encountering AT-rich region showing mainly pausing at second AT-rich site with multiple encounters at high FRET also. (A-B) N = 108

4.2.5 NTS substitution allows permissive transcription with no premature termination

Possible explanations for the observed premature termination could be due to the presence of the AT-rich sequence in the template strand or the formation of an RNA hairpin in the nascent RNA chain. To test for these possibilities, I modified the NTS sequence which maintains the base composition of the second AT-rich region and potential secondary structures on the nascent RNA (Artsimovitch and Landick, 2000). Now, the region that

was complimentary to the first AT-rich region was changed to be non-complimentary (Figure 4.7C Table 4.3). As this was within the first AT-rich region, where no falling off was detected before, I expected to see no difference in the transcription activity of RNAPII. Surprisingly, now RNAPII transcribes the full template without any issues, and no RNA can be detected in the S fraction at any stage, including the second AT-rich region where it was see before (Figure 4.7D).

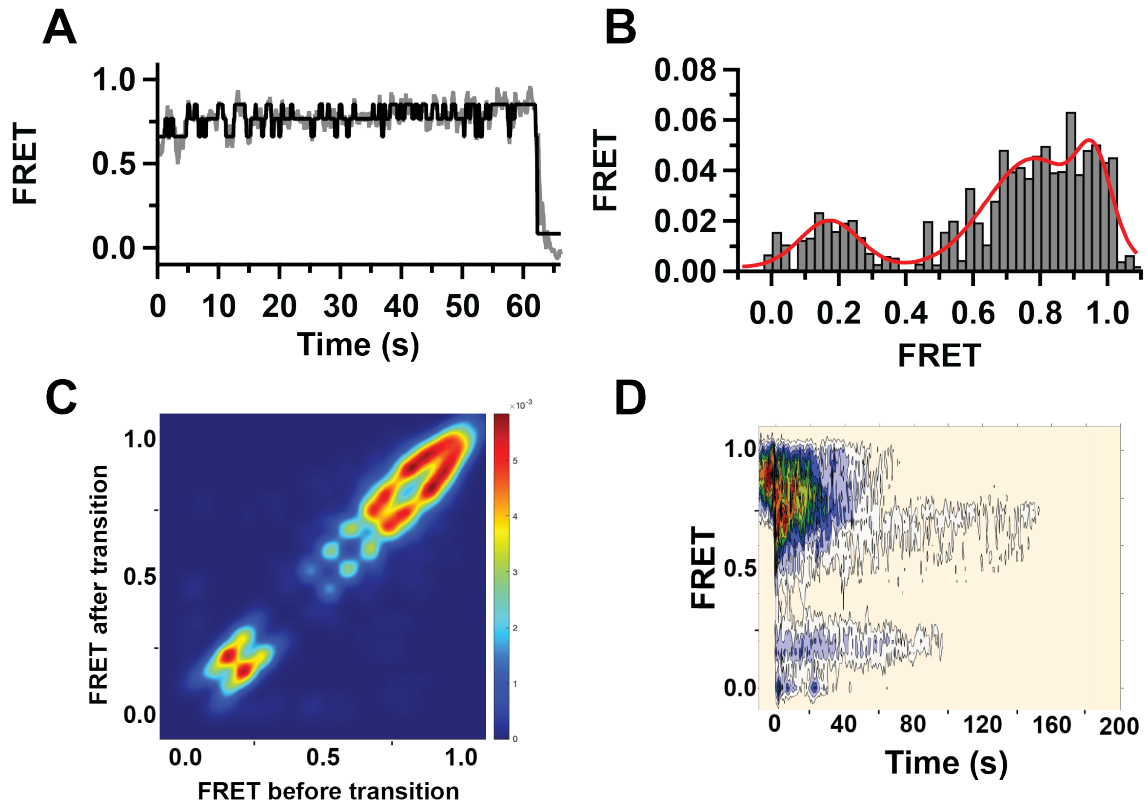


Figure 4.10: **NTS substitution does not induce RNAPII pausing in AT-rich region.** (A) Example FRET trace of RNAPII in presence of 10µM NTPs showing complete transcription of template, grey trace shows raw data, black trace shows idealised FRET states detected via HMM analysis. (B) Histogram of idealised FRET of RNAPII in presence of all 4 NTPs, peaks fitted to a Gaussian fit. (C) Transition density plot showing RNAPII transitioning through several intermediates along NTS mutation construct. (D) Post-synchronisation histogram of RNAPII showing mainly transitions in the first AT-rich site but minimal transitions in the low FRET range where second AT-rich region is. (A-D) N = 109.

smFRET experiments on this modified NTS and in the presence of all NTPs also revealed a different behaviour of the active TEC. The main differences are that the fraction of molecules that show the pause at the low FRET values of 0.1 - 0.4 decreases by at least two-fold (from 25% to 9%) and a slight increase in the fraction of trajectories that

complete transcription was also observed. Most importantly, these molecules showed complete transcription much faster (< 30s) than with the fully complimentary sequence and exhibit no pausing at the intermediate 0.3 FRET state (Figure 4.10A). However, a higher fraction of traces do now exhibit transitions within the mid and high FRET range (0.6 - 1.0), with the FRET histogram showing similar peaks at FRET = 0.96, 0.78 and 0.18 that were detected in the AT-rich sequence, but now contained another peak at 0.71 which was not present before (Figure 4.10B). Despite this increased occupancy, the polymerase transcribes through this region quickly with a similar rate to that seen in the sequence containing no AT-rich region at all (< 40 s) and faster than when the AT-rich sequence is fully complimentary (> 80s) (Figure 4.10D). These results indicate that the RNAPII stalling and premature termination observed is not caused by the AT-content of the TS or potential secondary structures in the nascent RNA. Instead a more prominent role for the NTS strand is suggested in this process.

4.3 Discussion

Single-molecule imaging of transcription has allowed new and greater insight into multiple stages of the transcription pathway, including initiation and elongation kinetics (Kapanidis et al., 2006; Dangkulwanich et al., 2013). However, many of these assays require modification to RNAPII, which is not always possible on higher eukaryotic samples such as mammalian samples. As a result, I have here developed a new single-molecule FRET based assay that allows for direct visualisation of RNAPII transcription elongation without the need to modify the protein, instead utilising fluorescently labelled DNA and RNA.

Initial attempts at generating a TEC provided mixed results, where I observed good TEC immobilisation but a poor FRET ratio was detected (Figure 4.1). In most cases no FRET was detected, only when the donor fluorophore was placed 2 nucleotides upstream of the RNA primer binding site did I observe a low FRET of 0.2. The most likely cause of the low FRET is due to the length of RNA used in this initial setup. Initially, I designed the RNA primer to have a 9 nucleotide binding region and a 12nt overhang, 21 nucleotides in total length. This was based on previous research that showed the RNAPII exit pore accommodates between 7 - 11 nucleotides of RNA, and I wanted the acceptor fluorophore to be outside of this channel so the protein would not interfere with the FRET (Chen et al., 2009). However, when I decreased the size of the primer to 12 nucleotides in total (9 nucleotides primer, 3 nucleotide overhang) I now observed not only a higher ratio of TEC:RNA signal ($> 80\%$) but also a significant number of molecules exhibited a high static FRET ratio as desired (Figure 4.2). This is likely for two main reasons. Firstly, when the primer is 12 nt in length, the 5' Cy5 will be positioned inside the exit tunnel of RNAPII, whereas when it is 21 nt the 5' Cy5 is protruding from the exit pore by approximately 3 nt, in this position the RNA is more flexible and less constrained than

when the primer is 12nt and this may result in a lower and more variable FRET ratio. However, as the RNA is still in close proximity to the donor fluorophore and we are within the persistence length of RNA we would still expect to observe some FRET. The second and more prominent reason for the lack of observed FRET is that when the primer is so long (21 nt), the ability of RNAPII to bind to the 9 nt binding region end and 'thread' the other 12 nt through the exit pore is quite challenging. If it cannot orientate the RNA properly, this will result in a non optimal FRET ratio, as was observed in our experiments (Figure 4.2).

Using these optimised labelling positions, the ability of this TEC to transcribe was next tested using mainly ensemble assays. Upon addition of NTPs, RNAPII was able to transcribe and synthesise nascent RNA encoding the template strand sequence in an NTP dependent manner, with minimal pausing detected. In the assay when 1mM [NTP] was used, transcription was complete within 1 minute, compared to over 10 minutes when this was lowered to 10 µM (Figure 4.3). In order to capture the real time dynamics of RNAPII transcription elongation, I used the lower [NTP] (10 µM) as it was likely that transcription would already be complete by the time imaging was started if using higher concentrations, meaning I would miss important information. This is fairly common practice in single-molecule imaging to slow a reaction down to capture important intermediates, but does not interfere with the function of RNAPII (Figure 5.11) (Dangkulwanich et al., 2013).

When immobilised onto the single-molecule slide, the TEC behaved as expected where a high static FRET was observed in the absence of NTPs (Figure 4.5). Although a high FRET state was expected based on labelling positions, the static nature of this complex was initially surprising. As the two dyes are in such close proximity (1-2 nm) we would expect a high level of interaction between the two dyes in multiple orientations.

However, as mentioned above the 5' Cy5 is located within the exit tunnel of RNAPII in this position and surrounded by amino acids. This in turn will confine the free rotation of the dye yielding the static traces that we observe. Because of this relative positions, it is unlikely that any dynamics we observe are due to dye rotations, but instead dye to RNAPII translocation along the DNA.

Addition of NTPs then lead to a decrease in the observed FRET over time in a predominately unidirectional manner. It is known that RNAPII does backtrack and pause occasionally on undamaged DNA and these events were also detected in our assay. RNAPII transitioned through several intermediates along the pathway identified via HMM analysis showing peaks at at least 4 positions determined by the empirical RNA ladder (start site, 9, 12 and 13 nucleotides). Although some short lasting pausing and backtracking events were observed, these almost always resulted in complete transcription of the template within a minute (Figure 4.6).

These dynamics were changed quite dramatically when encountering an AT-rich sequence along the template strand. The AT-rich TS induced significant pausing and backtracking of RNAPII as seen by the smFRET assay (Figure 4.8). Instead of RNAPII transcribing through the complete template, RNAPII now often backtracked to the pre-transcribed high FRET state (0.95) and paused significantly ($> 60s$) at a position within the AT-rich sequence (Figure 4.9). This pause was then proceeded by RNAPII falling off the template into solution (Figure 4.7).

As mentioned above, in both constructs tested short pauses during transcription were observed, something that is commonly observed in other published works and it is hypothesised that it is due to a sequence dependency rather than intrinsic pauses by RNAPII. While transcribing on a clean undamaged template RNAPII exhibited three main short

lived pause states, at FRET 0.6, 0.2 and 0.1 (Figure 4.6). However, when encountering the AT-tract DNA RNAPII paused in different positions, FRET 0.8, 0.4 and 0.2 (Figure 4.8). There are likely more pauses present, such as FRET 0.6, but due to the statistics being too low we could not determine if these were real pauses or not and so are not included in further analysis. Based on this initial evidence, we can initially suggest that the differences in the pause positions could be due to the sequence which is different between the two constructs. However, as the AT-tract is a pause inducing sequence we cannot make this claim fully from the evidence at hand., but only speculate that these pauses are sequence induced. To further probe these pauses another construct that mimics the undamaged template could be generated, containing a mix of A:T and G:C base pairs but that differs from the current sequences used. If the pauses are sequence dependent, we would expect to see different pause positions as observed currently, although likely within a similar time-scale.

Poly(A) regions within eukaryotic genomes are widely known to trigger the termination of RNAPII by cleavage of nascent RNA (Tran et al., 2001; White et al., 2013). It is thought that the polymerase pauses within this region, slowing down the polymerase to allow cleavage factors to bind. Furthermore, it has been shown using ensemble assays that AT-rich sequences can perturb and prematurely terminate transcription (Saeki and Svejstrup, 2009; Sigurdsson et al., 2010). Consistent with this hypothesis, our results show that the presence of AT-rich DNA does in fact reduce the rate of transcription and induces long lived pause states of RNAPII before becoming disengaged from the DNA. These AT-induced pauses could in turn allow for a termination mechanism whereby the polymerase is slowed down sufficiently for termination factors to bind and carry out their function. AT-rich DNA is known to cause weakening of the duplex and this reduced

duplex affinity, often found in promoter DNA, could be the cause of RNAPII pausing as it disrupts the already engaged RNAPII (Rajewska et al., 2012). Although it may account for the pausing of RNAPII, the disengagement off the DNA is not fully understood in this setting. In bacteria, the presence of these AT-rich sequences can often lead to the formation of RNA hairpins which physically block RNAPII from transcribing further than this position (Artsimovitch and Landick, 2000). This hypothesis was tested in our system but strikingly, it was the sequence of the NTS and not the TS that induces the pausing and premature termination from RNAPII (Figure 4.10). Changing the NTS sequence upstream of the pause site not only reduced the amount of RNAPII pausing but also prevented disengagement from the complex (Figure 4.10 & 4.7). This change maintains the RNA sequence and any possible secondary structure it may form and so any effect I see therefore rules out a role of hairpin formation on RNAPII disengagement, and instead highlights a more significant role for the NTS upstream of the dissociation region.

Although the NTS itself does not encode information to generate the nascent RNA, it does play an important function in transcription and contacts RNAPII in several locations. It has been shown in bacteria that mutating NTS upstream of a AT-rich region can affect termination by reducing the ability of the NTS to fully re-anneal to the TS (Ryder and Roberts, 2003). It is thought this re-annealing can drive disengagement from the DNA. Interestingly, a similar role of AT-tracts has been shown to be important also in termination of RNA Polymerase III (another eukaryotic RNAP) and so could be a more general feature of eukaryotic RNAPs than previously thought (Arimbasseri et al., 2014; Arimbasseri and Maraia, 2015).

Although these initial results indicate that AT-rich DNA induces long pauses this has only been tested on a single construct and therefore any conclusions we draw from this must

considered with a caveat. In order to fully address this hypothesis another AT-rich DNA construct should be tested with varying T-rich tracts and different lengths of homo-Ts to determine if this pause state is observed for all A-T tract regions or just this specific case. It should be noted, recent work with our collaborators has been carried out since writing of the thesis to address these concerns with both biochemical and genome wide sequencing results. These results show that long pauses and intrinsic termination does seem to be a general feature of A-T rich DNA and is not just observed in the single construct that we use. A further thing to consider is that the RNAPII used in most experiments was purified from yeast via the RPB3 FLAG tag. As mentioned in Chapter 1, the CTD of RNAPII is prone to phosphorylation on varying residues which can alter the regulation and function of RNAPII. Proteins purified from yeast can undergo post-translational modifications and as we do not use an affinity purification system for a specific phosphorylation state it is possible we will have a mix of RNAPII phosphorylation states.

The phosphorylation state of RNAPII may affect how it performs once bound to the DNA in our assay, with some phosphorylation markers causing active complexes to form while other leading to non-active complexes. It could also lead to different outcomes when encountering a specific damage or roadblock to transcription. This will be discussed further in Chapter 5 discussion in relation to the heterogeneous damage response. In order to address these possible issues, there are several experiments or alterations that could be made. Firstly, a second purification step for specific phosphorylation states could be added to the purification system. This would lead to a homogenous mix of RNAPII molecules. I hypothesise that we would observe similar results in relation to the dynamic nature of RNAPII, although some changes in active vs. non active complexes may be seen. A second option would be to perform post purification checks on the phosphorylation state

of RNAPII (such as a western blot) to determine if a heterogenous sample is present, and if so what the relative mix of states is.

The development of a new smFRET assay has allowed direct visualisation of the elongation pathway and dynamics without needing to label or modify RNAPII, allowing use of endogenous and higher eukaryotic samples. Furthermore, this assay has provided greater insight into the effect AT-rich DNA has on RNAPII and that the upstream NTS can help control premature termination.

4.4 Supplementary Figures

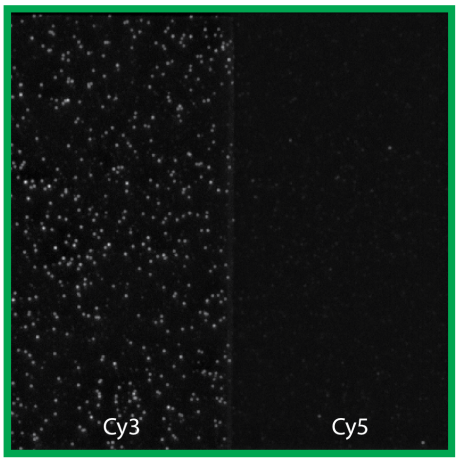


Figure 4.11: **NTS labelled TEC showing full complex bound to the surface.** TIRF image showing Cy3 labelled NTS immobilised to the surface via annealing to TS (biotin). The high number of bound molecules shows that the TS I visualise in other assays will also contain their relative NTS. 4mW Cy3 Excitation; Cy3 channel on left and Cy5 channel right.

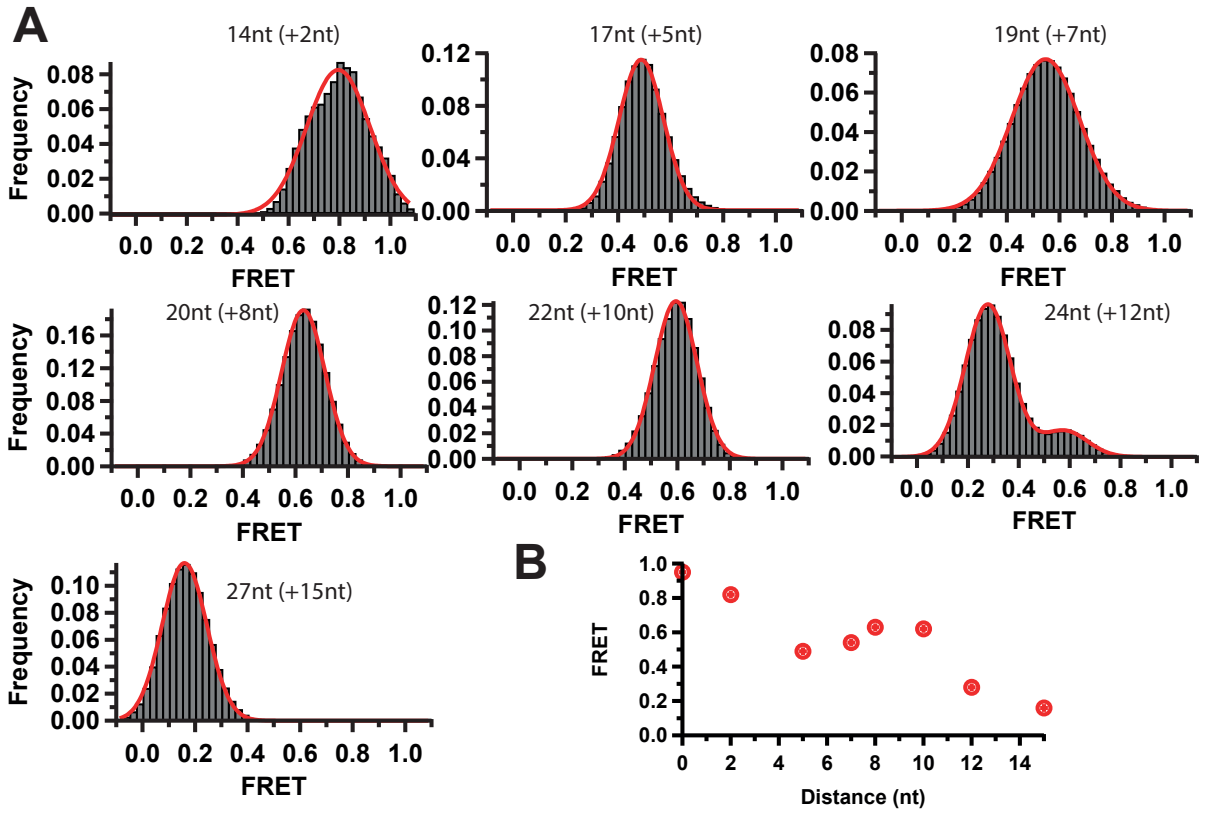


Figure 4.12: **RNA calibration.** (A) Histograms of FRET for each RNA primer used of varying lengths from 14 - 27nt (+xnt shows how far the RNA primer is from the original primer site). Peaks fitted to Gaussian fits. (B) FRET peaks from Gaussian fits plotted against number of nucleotides from start site.

4.5 DNA and RNA Sequences

Table 4.1: DNA and RNA sequences used in initial optimisation of smFRET assay

Name	Sequence (5' to 3')	Modification
TSP2	TACGC TGAGC TACGT AGCAA GTCGT GTCGC CTGCG TCCGT GTGGT GTCGC GATGG CTGTG **X**CTGA CTGAT	3' Biotin **X** = internal amine dT
TSP4	TACGC TGAGC TACGT AGCAA GTCGT GTCGC CTGCG TCCGT GTGGT GTCGC GATGG CTGTG TC**X**GA CTGAT	3' Biotin **X** = internal amine dT
TSP8	TACGC TGAGC TACGT AGCAA GTCGT GTCGC CTGCG TCCGT GTGGT GTCGC GATGG CTGTG TCTGA C**X**GAT	3' Biotin **X** = internal amine dT
NTSP2	ATCTG TCTG**X** CACAG CCATC GCGAC ACCAC ACGGA CGCAG GCGAC ACGAC TTGCT ACGTA GCTCA GCGTA **X**CCTA	**X** = internal amine dT
NTSP4	AATCTG TC**X**GT CACAG CCATC GCGAC ACCAC ACGGA CGCAG GCGAC ACGAC TTGCT ACGTA GCTCA GCGTA	**X** = internal amine dT
NTSP8	ATC**X**G TCTGT CACAG CCATC GCGAC ACCAC ACGGA CGCAG GCGAC ACGAC TTGCT ACGTA GCTCA GCGTA	**X** = internal amine dT
TS optimised	TACGC TGAGC TACGT AGCAA GTCGT GTCGC CTGCG TCCGT GTGGT GTCGC GATGG CTGTG **X**CTGA CGTTT	3' Biotin **X** = internal amine dT
NTS optimised	AAACG TCAGA CACAG CCATC GCGAC ACCAC ACGGA CGCAG GCGAC ACGAC TTGCT ACGTA GCTCA GCGTA	
Primer 21nt	rUrUrArCrA rGrCrCrArU rCrGrCrGrA rCrArCrCrA rC	5'Amine Group
Primer 12nt	rArCrArCrA rGrCrCrArU rCrG	5'Amine Group

Table 4.2: DNA and RNA Sequences used for smFRET assay

Name	Sequence (5' to 3')	Modification
TS	CATGC GTACA CGCGT TTGTA CACCA	3' Biotin
Undamaged	CATGG CATTC CACTT ATCAC TTGGA	**X** = internal
	CACCA CACCT CCTGG TCGCG **X**CCTA	amine dT
	TAGTT TGTTT CCTGA GT	
NTS	ACTCA GGAAA CAAAC TATAG GACGC	
	GACCA GGAGG TGTGG TGTCC AAGTG	
	ATAAG TGGAA TGCCA TGTGG TGTAC	
	AAACG CGTGT ACGCA TG	
RNA Primer	rArCrGrCrG rArCrCrArG rGrA	5'Amine Group

Table 4.3: DNA and RNA Sequences used for AT-rich DNA assay

Name	Sequence (5' to 3')	Modification
TS AT-rich	CATGC GTACA CGCGT TTGTA CAGAA	3' Biotin
	AAAAA AGAAA AATCC TGGTC GCG **X**C	**X** = internal
	CTACC TATAG TTTGT TTCCT GAGT	amine dT
NTS	ACTCA GGAAA CAAAC TATAG GACGC	
AT-rich	GACCAG GATTT TTCTT TTTTT TCTGT	
	ACAAA CGCGT GTACG CATG	
NTS Anti-	ACTCA GGAAA CAAAC TATAG GACGC	
termination	GACCAG GATGT GACTT TTTTT TCTGT	
	ACAAA CGCGT GTACG CATG	
RNA Primer	rArCrGrCrG rArCrCrArG rGrA	5'Amine Group

Bibliography

A. G. Arimbasseri and R. J. Maraia. Mechanism of transcription termination by RNA polymerase III utilizes a nontemplate-strand sequence-specific signal element. *Molecular cell*, 58(6):1124, 6 2015. ISSN 10974164. doi: 10.1016/J.MOLCEL.2015.04.002. URL /pmc/articles/PMC4475470//pmc/articles/PMC4475470/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4475470/.

A. G. Arimbasseri, K. Rijal, and R. J. Maraia. Comparative overview of RNA polymerase II and III transcription cycles, with focus on RNA polymerase III termination and reinitiation. *Transcription*, 5(1):1–13, 12 2014. ISSN 21541272. doi: 10.4161/TRNS.27369. URL /pmc/articles/PMC4214234//pmc/articles/PMC4214234/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4214234/.

I. Artsimovitch and R. Landick. Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proceedings of the National Academy of Sciences of the United States of America*, 97(13):7090–7095, 2000. ISSN 00278424. doi: 10.1073/pnas.97.13.7090.

C. Bernecky, J. M. Plitzko, and P. Cramer. Structure of a transcribing RNA polymerase II-DSIF complex reveals a multidentate DNA-RNA clamp. *Nature Structural and Molecular Biology*, 24(10):809–815, 2017. ISSN 15459985. doi: 10.1038/nsmb.3465.

S. C. Blanchard, R. L. Gonzalez, H. D. Kim, S. Chu, and J. D. Puglisi. tRNA selection and kinetic proofreading in translation. *Nature Structural and Molecular Biology*, 11(10):1008–1014, 10 2004. ISSN 15459993. doi: 10.1038/nsmb831. URL http://www.nature.com/natstructmolbiol.

C. Y. Chen, C. C. Chang, C. F. Yen, M. T. Chiu, and W. H. Chang. Mapping RNA exit channel on transcribing RNA polymerase II by FRET analysis. *Proceedings of the National Academy of Sciences of the United States of America*, 106(1):127–132, 1 2009. ISSN 00278424. doi: 10.1073/PNAS.0811689106/SUPPL{_}FILE/0811689106SI.PDF. URL www.pymol.org.

M. Dangkulwanich, T. Ishibashi, S. Liu, M. L. Kireeva, L. Lubkowska, M. Kashlev, and C. J. Bustamante. Complete dissection of transcription elongation reveals slow translo-

cation of RNA polymerase II in a linear ratchet mechanism. *eLife*, 2013(2):1–22, 2013. ISSN 2050084X. doi: 10.7554/eLife.00971.

A. N. Kapanidis, E. Margeat, S. O. Ho, E. Kortkhonjia, S. Weiss, and R. H. Ebright. Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science*, 314(5802):1144–1147, 11 2006. ISSN 00368075. doi: 10.1126/SCIENCE.1131399/SUPPL{_}FILE/KAPANIDIS.SOM.PDF. URL https://www.science.org/doi/abs/10.1126/science.1131399.

E. Nogales, R. K. Louder, and Y. He. Structural Insights into the Eukaryotic Transcription Initiation Machinery. *Annual Review of Biophysics*, 46(1):59–83, 2017. ISSN 1936-122X. doi: 10.1146/annurev-biophys-070816-033751. URL http://www.annualreviews.org/doi/10.1146/annurev-biophys-070816-033751.

E. Nudler. RNA polymerase backtracking in gene regulation and genome instability. *Cell*, 149(7):1438–1445, 2012. ISSN 00928674. doi: 10.1016/j.cell.2012.06.003. URL http://dx.doi.org/10.1016/j.cell.2012.06.003.

J. Oh, J. Xu, J. Chong, and D. Wang. Structural and biochemical analysis of DNA lesion-induced RNA polymerase II arrest. *Methods*, (October 2018):0–1, 2019. ISSN 10462023. doi: 10.1016/j.ymeth.2019.02.019. URL https://linkinghub.elsevier.com/retrieve/pii/S1046202318303001.

M. Rajewska, K. Wegrzyn, and I. Konieczny. AT-rich region and repeated sequences – the essential elements of replication origins of bacterial replicons. *FEMS Microbiology Reviews*, 36(2):408–434, 3 2012. ISSN 0168-6445. doi: 10.1111/J.1574-6976.2011.00300.X. URL https://academic.oup.com/femsre/article/36/2/408/565738.

A. M. Ryder and J. W. Roberts. Role of the Non-template Strand of the Elongation Bubble in Intrinsic Transcription Termination. *Journal of Molecular Biology*, 334(2):205–213, 11 2003. ISSN 0022-2836. doi: 10.1016/J.JMB.2003.09.039.

H. Saeki and J. Q. Svejstrup. Stability, Flexibility, and Dynamic Interactions of Colliding RNA Polymerase II Elongation Complexes. *Molecular Cell*, 35(2):191–205, 2009. ISSN 10972765. doi: 10.1016/j.molcel.2009.06.009. URL http://dx.doi.org/10.1016/j.molcel.2009.06.009.

S. Sigurdsson, A. B. Dirac-Svejstrup, and J. Q. Svejstrup. Evidence that Transcript Cleavage Is Essential for RNA Polymerase II Transcription and Cell Viability. *Molecular Cell*, 38(2):202–210, 2010. ISSN 10972765. doi: 10.1016/j.molcel.2010.02.026. URL http://dx.doi.org/10.1016/j.molcel.2010.02.026.

D. P. Tran, S. J. Kim, N. J. Park, T. M. Jew, and H. G. Martinson. Mechanism of Poly(A) Signal Transduction to RNA Polymerase II In Vitro. *Molecular and Cellular Biology*, 21(21):7495, 11 2001. ISSN 0270-7306. doi: 10.1128/MCB.21.21.7495-7508.2001. URL /pmc/articles/PMC99921//pmc/articles/PMC99921/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC99921/.

J.-W. Van De Meent, J. E. Bronson, F. Wood, R. L. Gonzalez, and C. H. Wiggins. Hierarchically-coupled hidden Markov models for learning kinetic rates from single-molecule data. *Proceedings of the International Conference on Machine Learning*, 28 (2):361–369, 2013.

E. White, K. Kamieniarz-Gdula, M. J. Dye, and N. J. Proudfoot. AT-rich sequence elements promote nascent transcript cleavage leading to RNA polymerase II termination. *Nucleic Acids Research*, 41(3):1797, 2 2013. ISSN 03051048. doi: 10.1093/NAR/GKS1335. URL /pmc/articles/PMC3561976//pmc/articles/PMC3561976/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC3561976/.

Chapter 5

DNA damage induces heterogeneous stalling and backtracking of RNA Polymerase II

5.1 Abstract

DNA damage has been shown to cause severe transcriptional stress to RNA Polymerase II including pausing, backtracking and stalling of the polymerase. If not rectified, the damage can lead to transcriptional arrest and genome instability. Our current understanding of transcription and the effect DNA damage has, have predominately come from structural and cellular studies. Although these studies have provided insight into this system, they can often miss key mechanistic steps.

The following chapter describes a project in collaboration with Dong Wang (University of California San Diego, UCSD) to investigate the effect of different DNA damages on RNA Polymerase II mediated transcription elongation. Using biochemical and structural techniques they were able to show that specific lesions cause stalling of the polymerase while others allow bypass. They have also shown the effect of TC-NER factors such as Rad26 on this stalling. However, the key mechanistic and dynamic steps involved in this process are yet to be elucidated.

In order to probe these dynamics, I used the single-molecule FRET assay developed in Chapter 4 and inserted three different damages (abasic lesions, 8-Oxo Guanine and TT-Dimer) into the template. I show that while all of these damages inhibit transcription, transcriptional stalling is a heterogeneous process with some damages causing static stalling and others dynamic stalling and backtracking. Furthermore, repair factors such as Rad26 can alter these dynamics to induce the repair pathway.

5.2 Results

As discussed in Section 1.4, DNA damage can pose a severe roadblock to transcription often inducing stalling, backtracking or even transcriptional arrest (Lans et al., 2019). How these damages interact with RNAPII and how repair factors such as Rad26 interact with stalled RNAPII has been studied intensely over the past decade, using predominantly biochemical and structural studies. Using the smFRET assay developed in Chapter 4, I aim to study damages including abasic lesions, oxidative damage and UV-induced lesions at the single-molecule level to observe how they interact with elongating RNAPII. These damages were chosen for two main reasons. Firstly, all of these have been shown to cause severe disruptions to transcription elongation and there is a wealth of both biochemical and structural data to directly compare to (Xu et al., 2017; Wang et al., 2018; Oh et al., 2019). Secondly, although they all induce stalling of RNAPII, they do so to varying degrees and hence should be good targets for single-molecule analysis to observe any heterogeneity within the RNAPII stalling mechanism.

5.2.1 RNAPII dynamically stalls in absence of all nucleotides

To investigate the effect of DNA damage on transcription elongation dynamics, a smFRET TEC construct based upon the one used in Section 4.2.3 was utilised. In order to study the effect of roadblocks, different pausing or damaged lesions were inserted at a specific position along the template strand (Figure 4.4 and Table 5.1). It should be noted that this work is part of a soon to be published paper in collaboration with Jesper Svejstrup (University of Copenhagen) and Dong Wang (UCSD) that at time of writing was in the submission phase. The work laid out below can also be found in the corresponding paper where there will be overlap.

Table 5.1: DNA and RNA Sequences used

Name	Sequence (5' to 3')	Modification
TS Abasic	CATGC GTACA CGCGT TTGTA CACCA	3' Biotin
	CATGG CATTC CACTT ATCAC TTG**Y**A	**Y** = Abasic Lesion
	CACCA CACCT CCTGG TCGCG **X**CCTA	**X** = internal
	TAGTT TGTTT CCTGA GT	amine dT
TS 8-Oxo	CATGC GTACA CGCGT TTGTA CACCA	3' Biotin
Guanine	CATGG CATTC CACTT ATCAC TTG**Y**A	**Y** = 8OG
	CACCA CACCT CCTGG TCGCG **X**CCTA	**X** = internal
	TAGTT TGTTT CCTGA GT	amine dT
TS	CATGC GTACA CGCGT TTGTA CACCA	3' Biotin
TT-Dimer	CATGG CATTC CACTT ATCAC **Y**GGA	**Y** = TT-Dimer
	CACCA CACCT CCTGG TCGCG **X**CCTA	**X** = internal
	TAGTT TGTTT CCTGA GT	amine dT
NTS	ACTCA GGAAA CAAAC TATAG GACGC	
	GACCA GGAGG TGTGG TGTCC AAGTG	
	ATAAG TGGAA TGCCA TGTGG TGTAC	
	AAACG CGTGT ACGCA TG	
RNA Primer	rArCrGrCrG rArCrCrArG rGrA	5'Amine Group

It has already been shown that addition of all 4 NTPs to a construct with no damage permits full transcription and a unidirectional decrease in FRET can be measured over time (Figure 4.6). It was noted during this analysis that although the FRET was predominately unidirectional, there was a significant amount of dynamic behaviour in both the forward and backwards direction. Due to these dynamics on the undamaged template, I decided to probe this further by artificially stalling the polymerase at specific positions along the template by limiting certain types of NTP. Addition of UTP and GTP allows transcription of only 10 nucleotides due to the subsequent base requiring CTP (Figure 4.3A and Table 5.1). As expected, addition of UTP and GTP caused a decrease in the observed FRET ratio, moving from 0.95 to 0.55 while transitioning through an intermediate at 0.65 (Figure 5.1A), corresponding to a position 1-3 nt upstream of the damage site validated via an empirical RNA calibration (Figure 5.1A and Supp. Figure 4.12). Surprisingly, rather than stalling and remaining bound at the site of damage, the observed FRET ratios were highly dynamic where the polymerase showed multiple rounds of backtracking

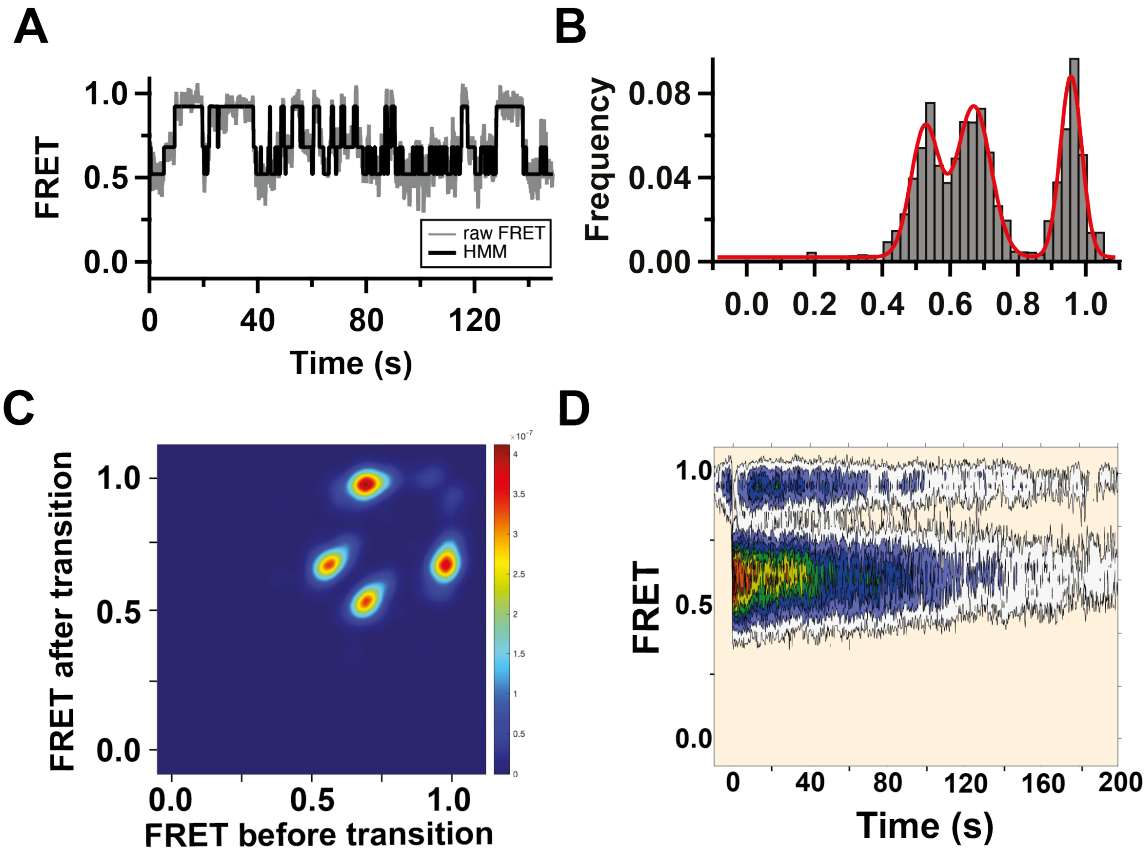


Figure 5.1: **RNAPII dynamically stalls when subsequent NTP is absent.** (A) (Left panels) Example FRET traces of RNAPII in presence of 10µM NTPs showing dynamic behaviour in presence of GTP & UTP (top panel) and GTP, UTP & CTP (bottom panel), grey trace shows raw data, black trace shows idealised FRET states detected via HMM analysis (Right panels) Histogram of idealised FRET of RNAPII in presence of respective NTP subset. N = 200 and 242 respectively, peaks fitted to a Gaussian fit. NTPs were added and FRET was captured over a 50-minute time period with several 5-minute windows.(B) Transition density plot showing RNAPII transitioning through two main intermediates along template DNA in absence of CTP and ATP (left panel) and ATP only (right panel).

and forward translocation between the high FRET state (0.95) and the site of damage (0.55). Using HMM analysis two different types of transitions were detected, short range transitions to and from the site of damage (0.55) and a slightly higher FRET state (0.65), corresponding to a few nucleotides movement, and longer range transitions from this state to the pre-transcribed high FRET state (0.95) (Figure 5.1B).

Dwell times of both the long range and short range transitions were extracted and in all cases showed an increase in comparison to the polymerase on an undamaged DNA template ($<$ 0.5-1 s), suggesting that the polymerase is inhibited significantly when lacking

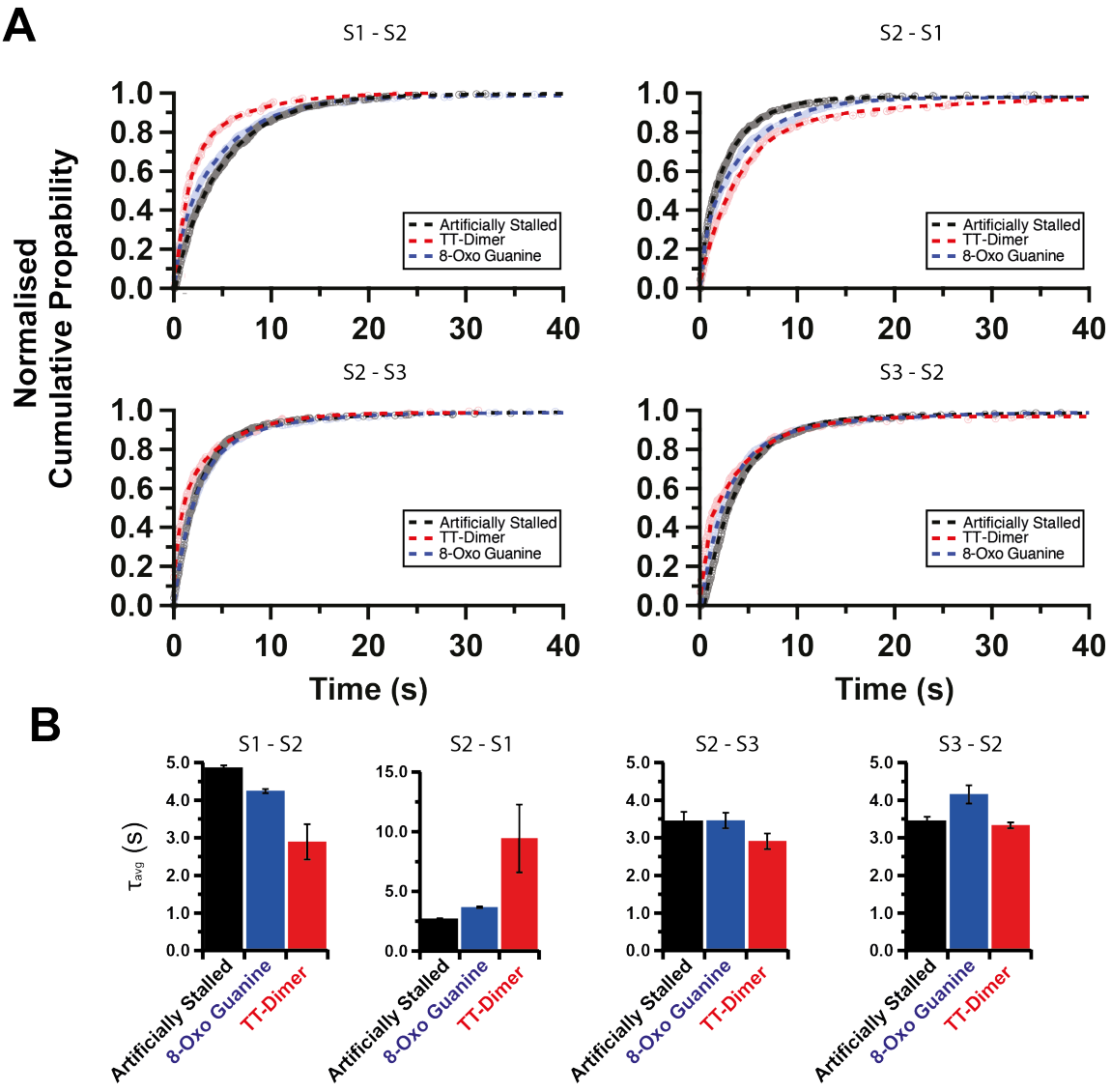


Figure 5.2: **Dwell times of RNAPII transitions for the 4 main observed transitions.** (A) Cumulative dwell time analysis of each transition observed for the 4 transitions. S1 - S2 (0.95 to 0.6) S2 - S3 (0.6 to 0.5) and their respective reverse transitions. Data was fitted to a double exponential. Markers indicate raw data, dotted line indicates double exponential fit (B) Average mean lifetimes where $\tau avg(s)$ was calculated from the weighted lifetimes of the fast and slow transitions determined from the double exponential fit from (A). Black = artificially stalled, blue = 8-Oxo Guanine damage, red = TT-Dimer damage. Error bars calculated from exponential fit errors.

the subsequent nucleotide. RNAPII remained stalled at the artificially stalled site with a mean lifetime of 3.5 s ± 0.1 s before transitioning back two nucleotides to the intermediate FRET state (0.65). Once in this position, RNAPII remained bound in this state for a similar but slightly less mean lifetime (2.7 s ± 0.1 s) before backtracking to the high FRET state (0.95). Once fully backtracked, RNAPII remained stuck in that state for

Table 5.2: Dwell time analysis of RNAPII Transitions

FRET State	Rad26	Damage	τ_{fast} (s)	τ_{slow} (s)	Ratio τ_{fast}	Ratio τ_{slow}	τ_{avg} (s)
S1 - S2	- Rad26	AS	1.1 ± 0.1	5.3 ± 0.1	0.10	0.90	4.9 ± 0.1
		8OG	0.7 ± 0.1	5.2 ± 0.1	0.21	0.75	4.3 ± 0.1
		TT	1.3 ± 0.1	6.2 ± 1.2	0.69	0.34	2.9 ± 0.5
	+ Rad26	8OG	2.0 ± 0.1	10.7 ± 1.0	0.54	0.18	4.2 ± 0.3
		TT	1.2 ± 0.1	7.1 ± 0.2	0.43	0.38	4.0 ± 0.1
S2 - S3	- Rad26	AS	2.4 ± 0.1	11.4 ± 1.7	0.86	0.12	3.5 ± 0.2
		8OG	1.9 ± 0.1	6.6 ± 0.5	0.65	0.32	3.5 ± 0.2
		TT	0.7 ± 0.1	4.8 ± 0.3	0.40	0.47	2.9 ± 0.2
	+ Rad26	8OG	1.5 ± 0.1	6.3 ± 0.2	0.50	0.24	3.1 ± 0.1
		TT	0.9 ± 0.1	6.0 ± 0.3	0.59	0.18	2.1 ± 0.1
S2 - S1	- Rad26	AS	0.7 ± 0.1	3.4 ± 0.1	0.23	0.72	2.7 ± 0.1
		8OG	0.5 ± 0.1	4.7 ± 0.1	0.22	0.73	3.7 ± 0.1
		TT	3.9 ± 0.1	37.6 ± 16.8	0.82	0.16	9.4 ± 2.8
	+ Rad26	8OG	0.9 ± 0.1	3.5 ± 0.1	0.24	0.45	2.6 ± 0.1
		TT	0.8 ± 0.1	4.1 ± 0.1	0.29	0.44	2.8 ± 0.1
S3 - S2	- Rad26	AS	3.5 ± 0.1	-	1.0	-	3.5 ± 0.1
		8OG	2.5 ± 0.1	10.7 ± 1.0	0.81	0.21	4.2 ± 0.2
		TT	0.3 ± 0.1	4.5 ± 0.1	0.26	0.66	3.3 ± 0.1
	+ Rad26	8OG	1.3 ± 0.1	6.7 ± 0.1	0.52	0.33	3.4 ± 0.1
		TT	1.2 ± 0.1	6.5 ± 0.1	0.40	0.45	4.0 ± 0.1

Note: AS = Artificially Stalled; 8OG = 8-Oxo Guanine; TT = TT-Dimer

4.9 s ± 0.1 s before translocating forwards to the site of damage again, and this cycle repeated several times over several minutes. (Figure 5.2 black bars and Table 5.2)

5.2.2 Abasic lesions cause static stalling of RNAPII

Abasic sites (Figure 5.3A) can occur spontaneously within the cell due to radiation sources but are also formed as an intermediate during the base excision repair pathway (BER) (Thompson and Cortez, 2020). It is expected that around 15,000 AP sites occur every day within a human genome. This can be a detrimental issue for polymerases where missing or incorrect bases within the growing RNA strand can lead to mutated protein synthesis. An abasic site was placed within the TEC 11 nucleotides downstream of the start site to test its effect on RNAPII. *In vitro* assays like the ones carried out on the undamaged template

showed a significant reduction in transcript length, with stalling at the expected site of damage observed. No RNA corresponding to the full length of the template was detected suggesting RNAPII cannot bypass an abasic lesion within the template strand (Figure 5.3B). Two main lengths of transcript were detected, the first being 1 nt before the site of damage (+10 nucleotides), with the other at the site of damage itself (+11 nucleotides) showing insertion of 1 NTP into the abasic site. Fast stalling within 30 seconds can be seen at the +10 nucleotides position, corresponding to the position where the abasic lesion is within the active site but no NTP has been added to the growing RNA chain. Eventually, more than 90% of stalled polymerases showed slow insertion (several minutes) of one nucleotide opposite the abasic lesion (Figure 5.3B). The efficiency of insertion for this extra NTP was decreased to 50% when the [NTP] was decreased to 10µM and the rate of insertion also decreased taking up to 30 minutes (Figure 5.11). The type of NTP inserted was further probed and it was found that in accordance with previous literature ATP was preferentially inserted opposite the abasic site (Figure 5.3B) (Wang et al., 2018).

Although the abasic lesion caused stalling of RNAPII as expected in the ensemble assay like the artificially stalled complex, the observed FRET traces were significantly different. Upon addition of all NTPs, over 90% of molecules exhibited a highly static and reproducible FRET state of 0.59, with almost no molecules showing any backwards motion to the high FRET state as was the case with the artificially stalled polymerase (Figure 5.3C). This suggests that abasic lesions are highly disruptive to RNAPII, with almost complete stalling of RNAPII at the damaged site and minimal dynamic or backtracking behaviour detected, highlighting the inability of RNAPII to recover from this blockage.

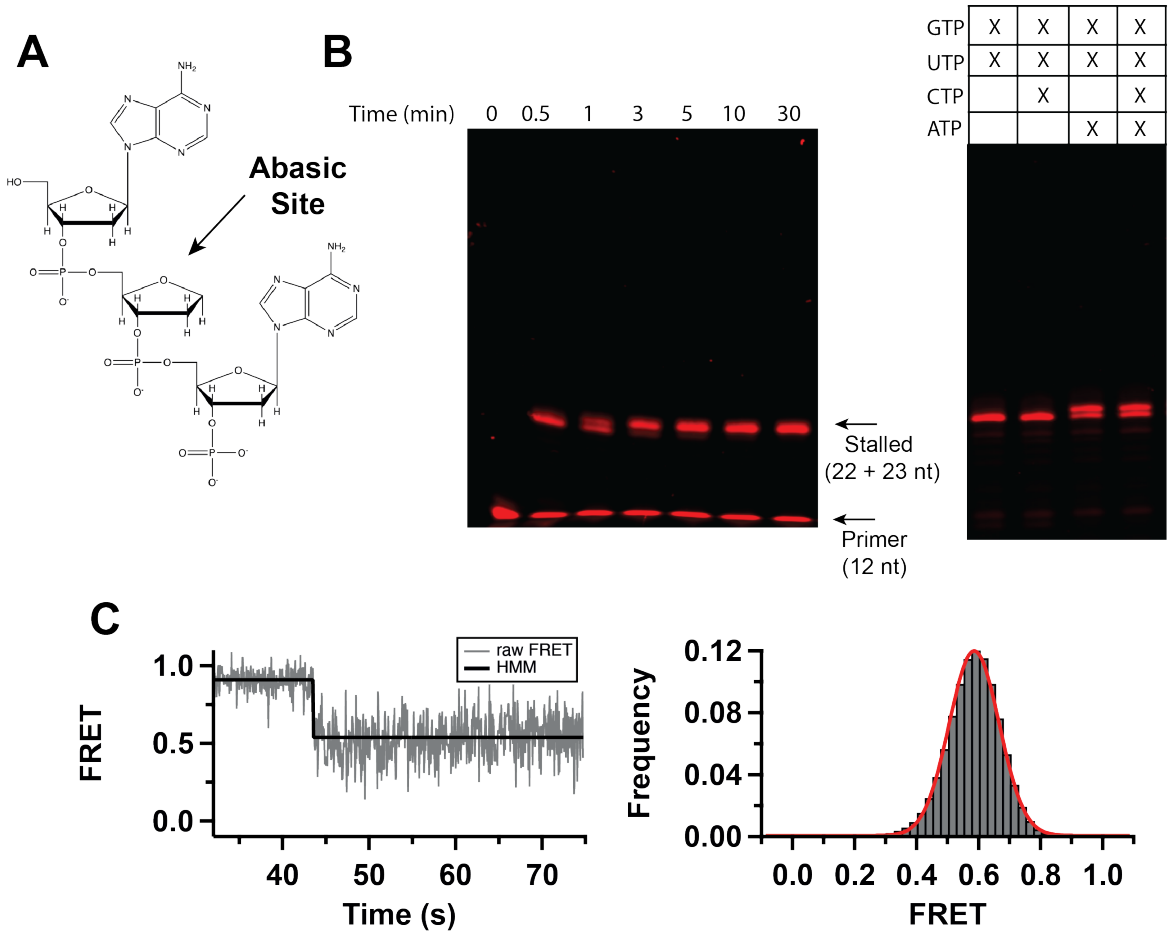


Figure 5.3: **Abasic lesions cause static stalling of RNAPII.** (A) Structure of abasic lesion within DNA backbone showing normal AMP on either side. (B) (left panel) denaturing PAGE showing RNA extension over time from primer (12 nt) to stalled transcript (22 nt) followed by slow insertion of ATP (23 nt) (right panel) denaturing PAGE visualising RNA extension with subset of NTPs showing specifically ATP insertion into the abasic site, 1 mM NTPs, measured via Cy5 fluorescence. (C) Example FRET trace of RNAPII in presence of 10µM NTPs showing static behaviour upon encountering an abasic lesion. Grey trace shows raw data, black trace shows idealised FRET states detected via HMM analysis. NTPs were added and FRET was captured over a 50-minute time period with several 5-minute windows. (D) Histogram of idealised FRET of RNAPII stalled at abasic lesion showing a single FRET peak of 0.6. N = 191, peaks fitted to a Gaussian fit.

5.2.3 Oxidative damage and bulky lesions induce dynamic stalling of RNAPII

RNAPII can also encounter multiple other damages during transcription leading to transcriptional stress, including oxidative damage such as 8-Oxo Guanine (Oxo-G) and cyclobutane pyrimidine dimers (TT-Dimer). It is known that these damages cause stress on polymerases leading to pauses along the gene and possible stalling of transcription (Lans

et al., 2019). By inserting either one of these sites into the FRET TEC, I am able to visualise how these different damages affect RNAPII like I did for the abasic lesion.

The presence of these damages along the template DNA also led to a drastic decrease in transcription efficiency. Run-Off (RO) transcripts of the expected full length decreased to 50% (this was even lower at 10µM where only 10% showed RO transcripts) when an 8-Oxo Guanine group is placed 11 nucleotides downstream of the start site, with RNAPII predominantly stalling around the site of damage (Figure 5.4A).

Visualisation of the RNA transcript lengths via the bulk transcription gels showed that most of the transcripts had a length of 12 nucleotides, 1 nucleotide further than the Oxo-G site. This suggests that RNAPII can initially bypass the Oxo-G lesion from the active site by inserting an NTP, but once it enters further into the RNAPII core, causes stalling after one further nucleotide.

The presence of the TT-dimer, placed 13 nucleotides downstream of the start site led to stalling more similar to the abasic lesion than the oxidative damage (Figure 5.4B). No RO transcripts were detected and two different transcript lengths were detected. Firstly, complete stalling at the site of damage was detected within 30 seconds followed by a slow incorporation of 1 nt opposite the TT-Dimer after 10 minutes(5.4B). The lack of bypass and more stalling states suggests a stronger stalling effect of the TT-dimer than an 8-Oxo Guanine group on RNAPII.

Both of these damages were tested using the smFRET based TEC assay in the same way as with the abasic lesion. The presence of an 8-Oxo Guanine group led to a decrease in the observed FRET as expected, where 90% of molecules showed a minimum FRET state of 0.51. Unlike abasic lesions however, the oxidative damage induced highly dynamic

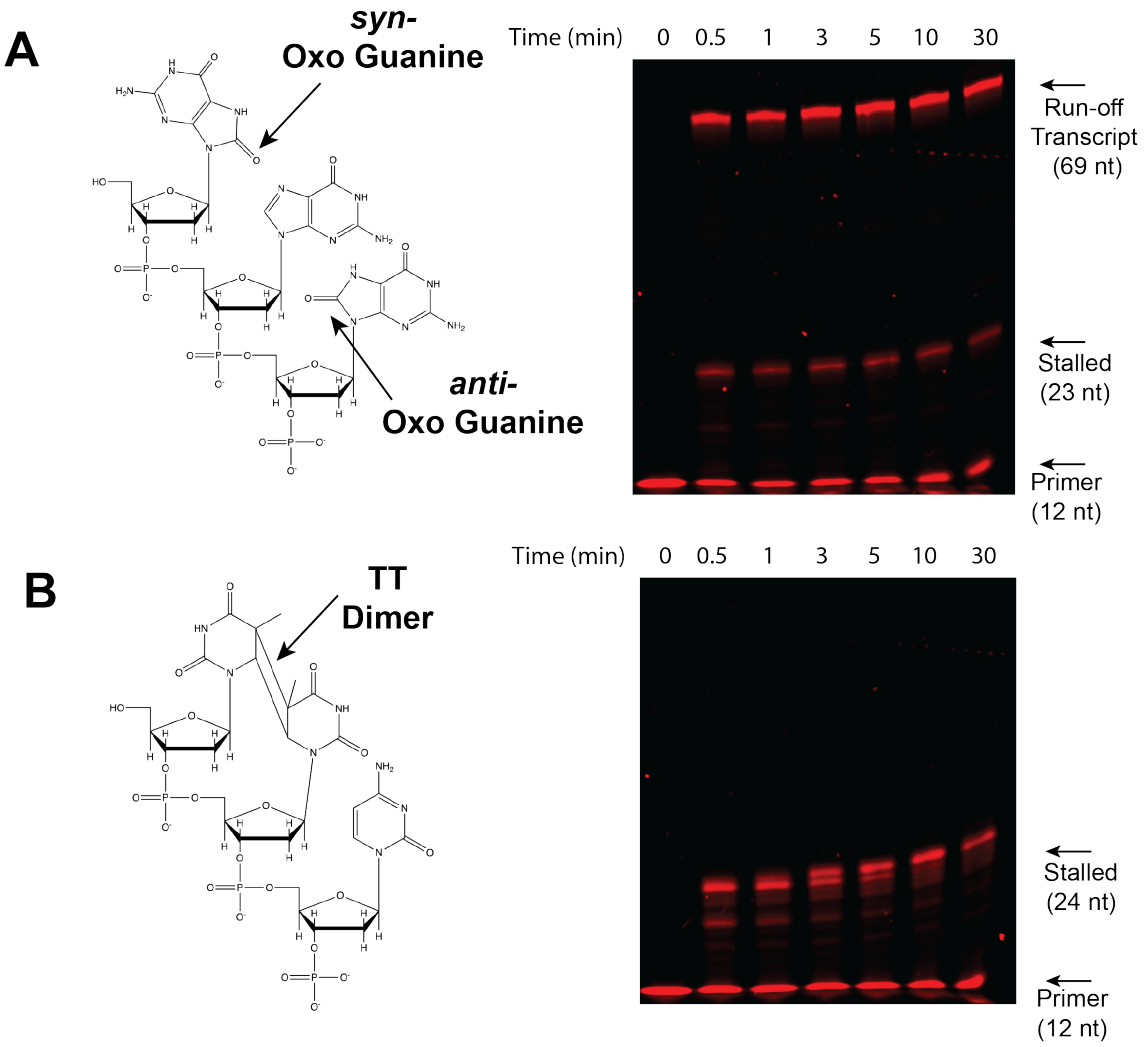


Figure 5.4: **Ensemble assay of RNAPII encountering TT-Dimer and 8-Oxo Guanine.** (A) Structure of 8-Oxo Guanine within DNA backbone showing both the syn- and anti- form (left panel) and Denaturing PAGE showing RNA extension over time from primer (12 nucleotides) to stalled transcript (23 nucleotides) and bypass of the damage to full length transcript (69 nucleotides) (right panel). (B) Structure of TT-Dimer within DNA backbone (bottom panel) and Denaturing PAGE showing RNA extension over time from primer (12 nucleotides) to stalled transcript (23 nucleotides) and addition of 1 NTP into the first T of TT-Dimer (24 nucleotides) (A-B) 1 mM NTPs, measured via Cy5 fluorescence.

stalling (80% of molecules), more closely related to that seen with the artificially stalled complex (Figure 5.5). Using the HMM analysis, I found that three main FRET states were observed. The first at 0.96 (corresponding to RNAPII in the pre-transcription state), and two further FRET states at 0.61 and 0.51. These FRET states and the transitions between them very closely resembled those seen with the artificially stalled complex, although a slightly higher proportion of molecules exhibited the FRET states corresponding to the

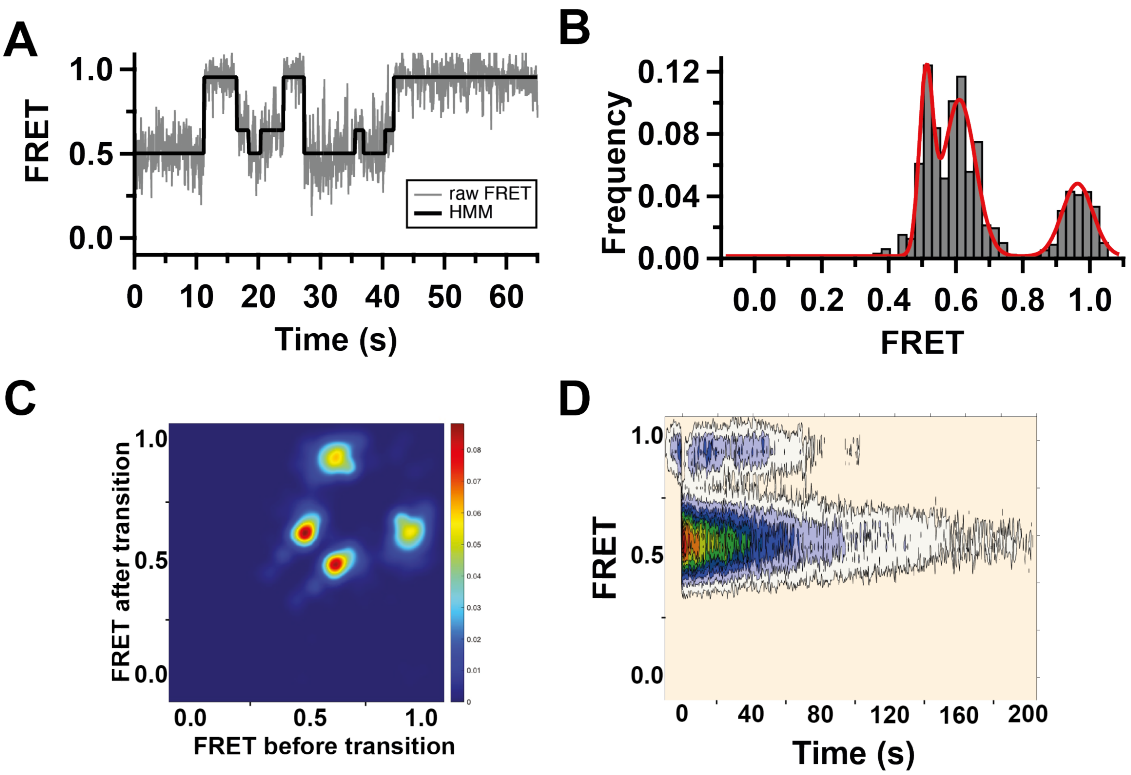


Figure 5.5: **8-Oxo Guanine causes dynamic stalling of RNAPII.** (A) Example FRET trace of RNAPII in presence of 10µM NTPs showing dynamic behaviour in presence of all 4 NTPs, grey trace shows raw data, black trace shows idealised FRET states detected via HMM analysis. NTPs were added and FRET was captured over a 50-minute time period with several 5-minute windows. (B) Histogram of idealised FRET of RNAPII in presence of all 4 NTPs showing peaks at 0.95, 0.6 and 0.5, peaks fitted to a Gaussian fit. N = 174 (C) Transition density plot showing RNAPII transitioning through two main intermediates along template DNA. (B) Post-synchronisation histogram of RNAPII encountering oxidative damage showing mainly stalling at damage site with multiple encounters at high FRET also.

site of damage (0.51 and 0.61) rather than the pre-transcribed state (0.95) (Figure 5.5B-D).

As well as observing similar dynamics, RNAPII observed comparable mean lifetimes at each state observed as when artificially stalled. The polymerase remained at the site of damage for 4.2 s ± 0.2 s before backtracking to the FRET intermediate of 0.65, where it remained for a further 3.7 s ± 0.1 s before backtracking further to the high FRET state (0.95) (Figure 5.2and Table 5.2). When artificially stalled, RNAPII remained bound around the site of damage for a total of 6.1 s while RNAPII remained bound for a total of 7.7 s (Table 5.2). Once in this position, RNAPII transitioned out of this state after approx. 3-4 s. (Figure 5.2and Table 5.2) also comparable to when RNAPII is artificially

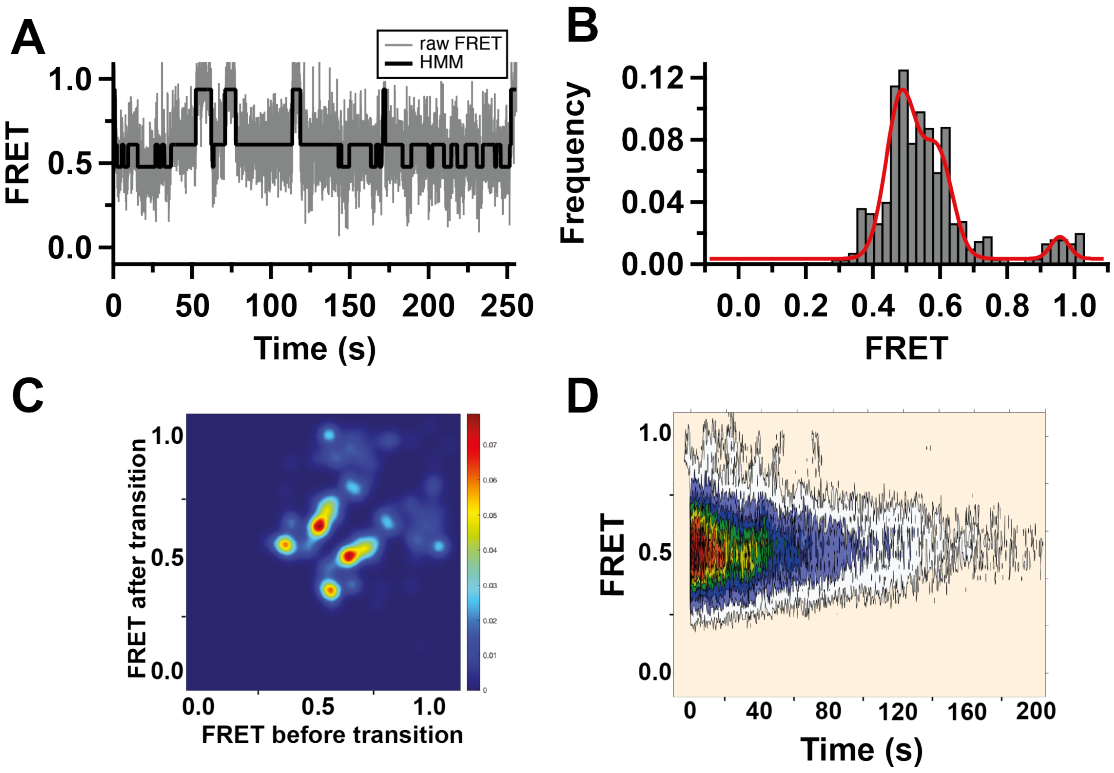


Figure 5.6: **TT-Dimer causes both dynamic and static stalling of RNAPII.** (A) Example FRET trace of RNAPII in presence of 10µM NTPs showing dynamic behaviour in presence of all 4 NTPs, grey trace shows raw data, black trace shows idealised FRET states detected via HMM analysis. NTPs were added and FRET was captured over a 50-minute time period with several 5-minute windows. (B) Histogram of idealised FRET of RNAPII in presence of all 4 NTPs showing peaks at 0.95, 0.6 and 0.5, peaks fitted to a Gaussian fit. N = 113 (C) Transition density plot showing RNAPII transitioning through two main intermediates along template DNA. (B) Post-synchronisation histogram of RNAPII encountering a TT-Dimer showing mainly stalling at damage site with minimal encounters at high FRET.

stalled suggesting RNAPII transitions to and from the damage sites with equal propensity. These high levels of dynamics highlights that although RNAPII is stalling at the site of damage, the molecule will undergo multiple rounds of backtracking to a 'pre-damaged' state before re-attempting to bypass the oxidative damage unlike what was seen for the abasic lesion.

This is in contrast to the FRET data observed when a TT-dimer is present in the template DNA region. Firstly, no traces were observed where the FRET eventually decreased to 0 like in the Oxo-G template, further validating the bulk data showing no bypass of a TT-dimer. Secondly, the percentage of molecules which showed dynamics dropped to

approx. 50%, with the other 50% showing highly static FRET levels at 0.5. Although RNAPII does explore similar states to that seen with the 8-Oxo Guanine and artificially stalled templates (0.96, 0.59 and 0.49), the frequency of the observed FRET was significantly different (Figure 5.6A & B). While the artificially stalled and oxidative damage complex spent a significant amount of time in the pre-transcribed high FRET state (20-30%), RNAPII spends minimal time in this state when encountering a TT-Dimer (5%) and instead occupies the FRET states corresponding to the site of damage (0.59 and 0.49). Finally, the mean lifetimes observed also varied significantly from the other damages tested, particularly for the short range transitions (Figure 5.6C). Although RNAPII remains bound at the site of damage for a similar time (3.3 s $\pm$ 0.1 s) before backtracking to the mid FRET state, once here RNAPII remains bound for approx 3x longer with a mean lifetime of 9.4 s $\pm$ 2.8 s (Figure 5.2 and Table 5.2). When RNAPII does eventually backtrack fully to the high FRET state (0.95), it only remains bound there for a short amount of time (2.9 s $\pm$ 0.5 s) before translocating back towards the lesion (Figure 5.2 and Table 5.2).

Overall, not only does a TT-Dimer not permit bypass of RNAPII, but also much fewer backtracking transitions are seen, as well as less dynamics observed. This data shows that while RNAPII backtracks frequently and rapidly to the pre-transcribed state when encountering an 8-Oxo Guanine group or is artificially stalled, it remains mostly at the site of damage when encountering a TT-Dimer. Based on both the bulk biochemical and smFRET data, this suggests that the presence of the TT-dimer causes a much stronger transcriptional stress effect on RNAPII than the 8-Oxo Guanine group.

The presence of these dynamics was surprising given that the abasic lesion caused such static stalling at the site of damage. To validate that these dynamics are due to the

damage and protein dependent, the excess NTPs were washed away from the surrounding solution. If the dynamics were NTP dependent, I would expect the dynamic behaviour to stop and become statically stalled in one state like what was observed with the abasic lesion. However as expected, washing away excess NTPs had minimal effect on RNAPII dynamics, where both the forward and backwards translocation steps were still detected with similar ratios as before (Figure 5.7 vs Figures 5.5 and 5.6).

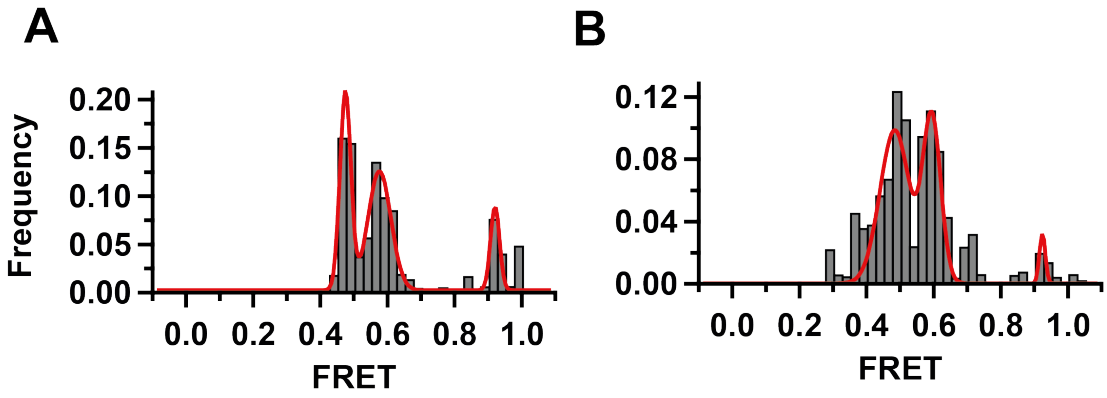


Figure 5.7: **RNAPII dynamics stalled at damages are not NTP dependent.** (A) Histogram of idealised FRET of RNAPII stalled at site of 8-Oxo Guanine with excess NTPs washed away showing similar peaks at 0.47, 0.58 and 0.92, peaks fitted to a Gaussian fit. N = 78 (B) Histogram of idealised FRET of RNAPII stalled at site of TT-Dimer with excess NTPs washed away showing similar peaks at 0.48, 0.59 and 0.92, peaks fitted to a Gaussian fit. N = 123

5.2.4 Rad26 stimulates bypass of DNA lesions by preventing backtracked state

There are several pathways within the cell to detect and fix damage along DNA, one such pathway is TC-NER. TC-NER is responsible for fixing bulky adducts and recruits a factor called Cockayne syndrome group B (CSB) in humans or Rad26 in yeast and is one of the first to arrive at the site of stalled RNAPII. Mutations within the CSB protein are affiliated with Cockayne syndrome and so is important to understand how it interacts with

stalled RNAPII. It has recently been shown that recruitment of Rad26 allows bypass of some lesions such as oxidative damage but not others such as abasic sites and bulky lesions by binding upstream of stalled RNAPII (Xu et al., 2017; Oh et al., 2020). However, the exact mechanism of how Rad26 interacts with RNAPII at these damages is still unknown (Xu et al., 2017). In collaboration with the Wang group I aimed to elucidate such insights.

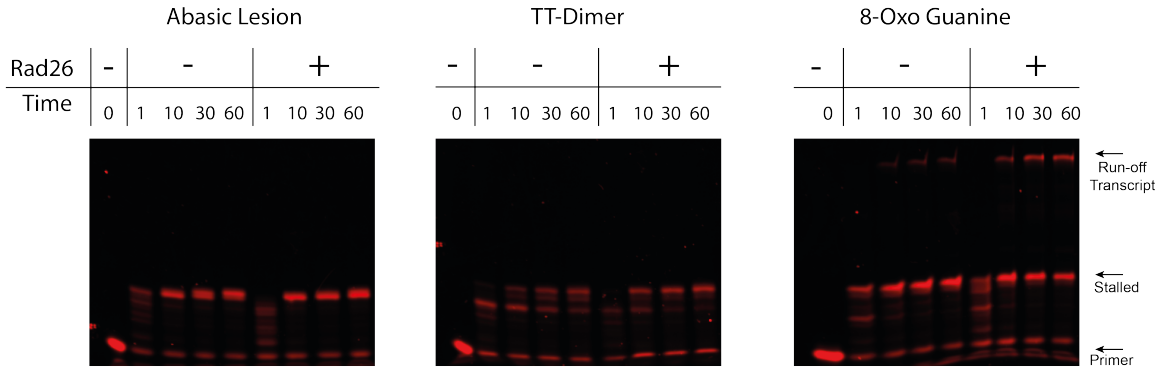


Figure 5.8: **Rad26 induces bypass of oxidative damage but not bulky or abasic lesions.** Denaturing PAGE showing RNA extension over time from primer (lower band, 12 nucleotides) to full length transcript (top bands, 69 nucleotides) or stalled position (mid arrow) at 10 µM [NTP] when RNAPII encounters a abasic lesion (left), TT-Dimer (middle) or 8-Oxo Guanine lesion (right), measured via Cy5 fluorescence.

In the presence of the 8-Oxo Guanine lesion, *in vitro* transcription assays showed that Rad26 increased the efficiency of bypass from 10%to 30% (Figure 5.8 right). Although no bypass of the TT-Dimer was detected upon addition of Rad26, a modest increase in the amount of RNAPII stalled at the site of damage was seen (Figure 5.8 middle). Minimal effect was detected upon RNAPII stalled at an abasic lesion, with no bypass or change in stalling was detected (Figure 5.8 left). Because of these effects seen *in vitro*, Rad26 was added to the smFRET assay where RNAPII was stalled at either the 8-Oxo Guanine or TT-Dimer lesions, to see if the observed FRET dynamics were altered.

At the single molecule level, Rad26 has a drastic effect on the stalled RNAPII dynamics at both lesions. Firstly, the original high FRET state (0.95) was no longer observed,

but instead Rad26 seems to shift the high FRET state to a decreased position of 0.85 (Figure 5.9A). The intermediate state previously observed at 0.6 did remain however in the position. The second major change showed that the low FRET state of 0.5 was also shifted to a decreased position of 0.4, consistent with a forward movement by approximately 1 – 2 nt based on the RNA calibration (Figure 5.9A and 4.12). These results are consistent with the structure of Rad26 bound to stalled RNAPII, which shows that Rad26 promotes forward motion of RNAPII (Xu et al., 2017). Furthermore, in line with the bulk data, the rate of bypass for the 8-Oxo Guanine lesion also increased at the single-molecule level but no bypass was observed when encountering the TT-Dimer.

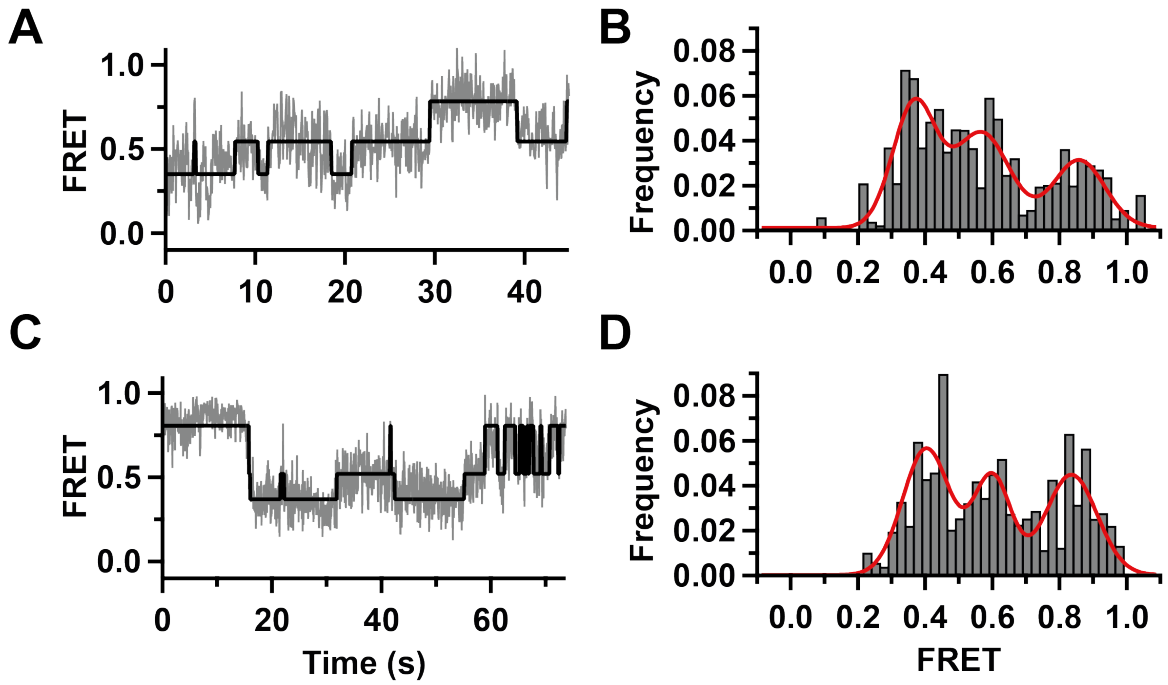


Figure 5.9: **Rad26 prevent long range backtracking of RNAPII.** (A) Example FRET trace of RNAPII in presence of Rad26 at 8-Oxo Guanine lesion showing dynamic movement of RNAPII. Grey trace shows raw data, black trace shows idealised FRET states detected via HMM analysis. NTPs were added and FRET was captured over a 50-minute time period with several 5-minute windows. (B) Histogram of idealised FRET of RNAPII stalled at site of 8-Oxo Guanine in presence of Rad26 showing peaks at 0.37, 0.57 and 0.86, peaks fitted to a Gaussian fit. (A-B) N = 127. (C) Example FRET trace of RNAPII in presence of Rad26 at a TT-Dimer showing dynamic movement of RNAPII. Grey trace shows raw data, black trace shows idealised FRET states detected via HMM analysis.(D) Histogram of idealised FRET of RNAPII stalled at site of TT-Dimer with excess NTPs washed away showing similar peaks at 0.40, 0.60 and 0.83, peaks fitted to a Gaussian fit. (C-D) N = 136

5.3 Discussion

Roadblocks such as DNA damage have been shown to perturb transcription leading to stalled polymerases, premature termination and eventually genome instability and cancer (Kouchakdjian et al., 1991; Kumar et al., 2022). Many examples of this have been shown both *in vitro* from structural and ensemble assays and *in vivo*. Although these roadblocks are often fixed by one of the multiple cellular damage repair pathways (e.g. TC-NER), RNAPII often encounters these damages and will be affected, however the precise mechanism of this process is not fully understood. Here, I demonstrate an assay to visualise this dynamic process and look at how different damages lead to heterogeneous mechanistic response of RNAPII (Figure 5.10).

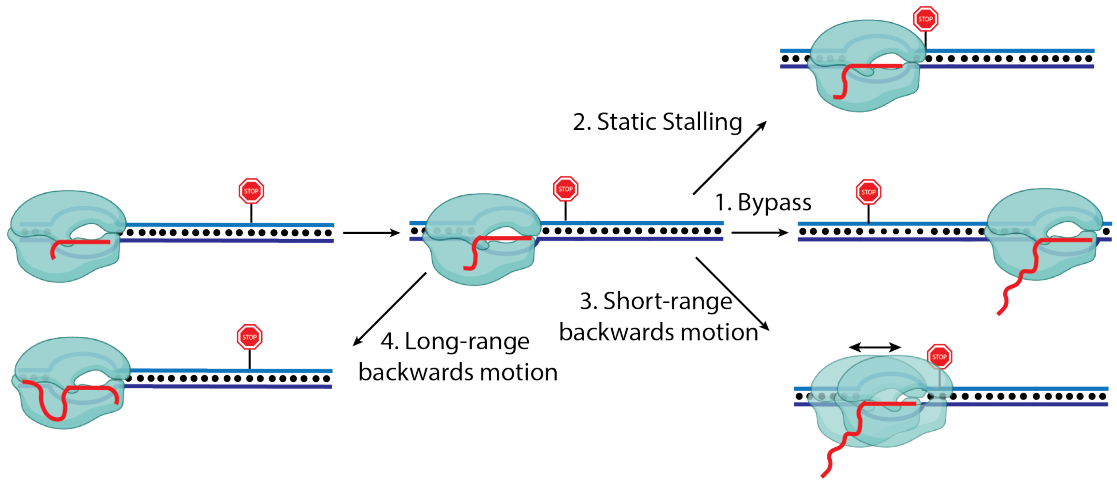


Figure 5.10: **RNAPII stalling mechanisms vary based on damage.** Schematic of RNAPII outcomes when encountering different damages. 4 main outcomes are visualised with the smFRET assay. 1. Bypass of the damage, 2. Static stalling, 3. short range dynamics and 4. long range dynamics.

RNAPII transcribed the undamaged DNA without significant stalling along the template when provided with all NTPs (Section 4.2.3). However, limiting the types of NTPs present severely impacted the RNAPII transcription pathway, inducing stalling when unable to insert the cognate NTP required by the TS (Figure 5.1). This is a commonly known

feature of RNAPII and has been observed in many biochemical assays (Brueckner et al., 2007; Wang et al., 2018; Oh et al., 2020). However, the mechanism of how RNAPII stalls at this position is not directly understood. When monitored using the smFRET assay, RNAPII exhibited a high level of dynamic behaviour, where RNAPII transitioned in both the forwards and backwards motion several times, transitioning through three main FRET states (0.95, 0.6 and 0.5). Two main types of transitions occurred, long range transitions, from 0.95 to 0.6, and short range transitions, from 0.6 to 0.5. Using an RNA calibration I confirmed these sites to be the pre-transcription conformation (0.95), 1-3 nt upstream of the damage site (0.6) and the site of damage (0.5). This suggests that when the RNAPII is unable to insert the subsequent NTP, it moves into a backtracked state, either close to the damage (0.6) or further upstream (0.95). This backtracking motion is likely to allow time for the correct NTP to be present and enter the active site of RNAPII. Based on dwell times analysis there was no preference for each state were similar dwell times were observed for all stalling positions (approx 3-4s s) (Figure 5.2 and Table 5.2).

The biochemical response of RNAPII to an 8-Oxo Guanine lesion was as expected and in line with previous data, where moderate stalling at the damage was observed but bypass of the lesion was the predominant outcome (Figure 5.4) (Oh et al., 2020). Interestingly, I observe that the predominant stalling position is at +13 nt from the start site, at the base (G) after the 8-Oxo Guanine lesion. This suggests that rather than pausing before the lesion, RNAPII will readily insert an NTP opposite the damage before becoming stalled once the inserted base has been moved +1 nt from the active site. Structural analysis has shown that an 8-Oxo Guanine lesion causes minimal changes in the DNA conformation both downstream and within the active site (Damsma and Cramer, 2009). Together this suggests that the initiating event for stalling is likely not the lesion itself

but the insertion of an NTP, and likely the non-cognate AMP, opposite 8-Oxo Guanine. Previous biochemical data has also suggested that RNAPII will backtrack at oxidative lesions including 8-Oxo Guanine, something that is also heavily observed in the smFRET results (Damsma and Cramer, 2009; Oh et al., 2020). The 8-Oxo Guanine modification caused almost identical dynamics and mean lifetimes to that seen with the artificially stalled, where RNAPII transitioned with roughly equal propensity in both the short and long range transitions (Figure 5.1 and 5.2). This ability to backtrack to a high extent is likely due to the structural similarities between the two conditions, where RNAPII is free to translocate in both the forwards and backwards direction without any impediment from the DNA backbone or nascent RNA. This is backed up further when analysing the structures of RNAPII at sites of abasic lesions and CPDs, which as discussed below are far more deforming than 8-Oxo Guanine.

The presence of an abasic lesion induced almost complete stalling at the site of damage, followed by slow insertion of AMP specifically opposite the abasic site, as has been previously observed for RNAPII (Wang et al., 2018). Abasic lesions caused a significantly different observed effect on RNAPII compared to the 8-Oxo Guanine lesion, where it stalled at the site of damage but did not exhibit the same dynamics. Instead, it remained stuck 1 nucleotide upstream of the abasic lesion (0.6) with minimal backtracking motion detected (Figure 5.3). Of the 10-15% of particles that did show dynamics, these were predominately short range transitions moving between a FRET state of 0.6 and 0.5 suggesting the RNAPII might be able to dynamically move around the site of damage but the presence of the abasic site within the active site blocks most of this (Wang et al., 2018). This high level of static stalling that I observe could be due to the increased flexibility of the 3' RNA. Recent structural studies have shown that due to the inability to base pair

with the template strand, the 3' end of the RNA can occupy the active site of RNAPII, blocking the insertion of a subsequent NTP (Wang et al., 2018). Furthermore, the template strand also showed an increased flexibility, due to lack of base stacking, where the abasic lesion was found above the bridge helix of the active site. It is hypothesised that the placement of this lesion and the increased flexibility would block the RNAPII from moving forwards and also backwards (Wang et al., 2018). Interestingly, the same study showed that RNAPII does backtrack significantly when encountering abasic lesions. Although I observe minimal dynamics in this assay when encountering abasic lesions, the dynamics that are observed are all in the 1-3 nt, which validates what was observed in the biochemical assay. One possible reason for the difference here could be that they observe backtracking in the absence of NTPs at an abasic site, whereas my measurements are in a solution of all NTPs, a more biological environment. Furthermore, backtracking is measured via TFIIS induce RNA cleavage, which is not present in my assay (Wang et al., 2018). This would emphasise any RNAPII backtracking as the RNA cleavage will further stimulate the ability to backtrack away from the lesion.

The TT-Dimer also induced stalling at the site of damage followed by slow insertion (several minutes) of a single NTP opposite the 5' thymine of the TT-Dimer (Figure 5.4). This is similar to what has been observed by other labs, although surprisingly insertion of 2 nucleotides (opposite both thymines) has been previously reported before complete stalling, something that I do not observe (Brueckner et al., 2007). However, recent structural advancements in visualising RNAPII stalled at CPDs has not observed this and so the number of nucleotides inserted is still debated (Lahiri et al., 2019; Wang et al., 2018). Structurally, a CPD lesion has been shown to cause similar DNA changes as with an abasic lesion where the damaged site sits on top of the bridge helix of the

trigger loop (Brueckner et al., 2007; Lahiri et al., 2019). Both of these damages cause structural changes in the template strand, often leading to an increased flexibility. This is due to the inability to base-stack with adjacent bases and the damage adopting a different conformation inducing the increased flexibility. Furthermore, it has been shown that the 3' end of the nascent RNA is more dynamic and can occupy the active site of the incoming NTP, preventing further translocation (Brueckner et al., 2007).

The smFRET data here showed that although RNAPII does stall at CPD lesions, it does so in a more dynamic manner where RNAPII is able to backtrack to a greater extent (50% compared to 10%). The polymerase transitioned through almost the same FRET intermediates (0.95, 0.6 and 0.5) as with the 8-Oxo Guanine and artificially stalled complex, but with varying rates and proportions. The TT-dimer led to the polymerase stalling predominantly at the site of damage and transitioning in the short-range region (between 0.6 and 0.5), while rarely moving backwards into the high FRET conformation (0.95). The mean lifetime for the stalled positions (0.5 and 0.6) were over 12 s, over twice as long as what was observed with an 8-Oxo guanine lesion or artificially stalled RNAPII (Figure 5.6 and table 5.2). This difference in dynamics could be addressed by this structural information, as although they are similar the structures do contain some differences. For example, the CPD induced less deformation on the downstream DNA than the abasic lesion caused and the 3' RNA showed slightly less flexibility as it does not contain the ability to base-pair with the CPD (Wang et al., 2018; Lahiri et al., 2019). This decreased 3' RNA flexibility backs up the increased dynamic backtracking movement that I observe here. As the active site is blocked to a lesser extent when encountering a CPD, RNAPII is able to backtrack to a higher degree than when encountering an abasic lesion. In fact, it is hypothesised that the stalling effect on RNAPII is primarily due to the

insertion of NTPs opposite the CPD, rather than the lesion itself. This is more similar to the case with 8-Oxo Guanine and likely accounts for their similarities in observed FRET states. These structural changes coupled with the smFRET data suggests that the reduced flexibility and deformation allows for more dynamic backtracking of RNAPII in attempts to recover from this lesion, highlighting a heterogeneous stalling mechanism of RNAPII at damages which biochemical data would otherwise suggest are similar in nature (Wang et al., 2018; Brueckner et al., 2007).

The increased time in the small range backtracked state could be to allow other repair proteins to detect the site of damage and fix it before allowing transcription to resume. Bulky lesions like CPDs often require excision of the damage site trough the TC-NER pathway and cannot progress until this has been fixed. In this pathway, RNAPII is the signalling factor to initiate TC-NER before other factors will bind to assist in the repair (Brueckner et al., 2007). Whereas small modifications like 8-Oxo Guanine can allow transcription to continue and in fact only stall in approximately 20-30% of cases. These small lesions are also fixed by the BER pathway and do not require stalled RNAPII to be recognised first, hence why RNAPII might be more dynamic at this lesion, compared to the CPD. The presence of the damage does decrease the rate of transcription significantly however, most likely due to the lesion adopting different conformations. 8-Oxo Guanine can adopt both an anti- and syn- form where the latter leads to mis-pairing with AMP causing mutations in the RNA and proteins possibly leading to cancer (Kouchakdjian et al., 1991; Kumar et al., 2022). RNAPII may move backwards even further to the high FRET state to allow other regulatory factors such as Rad26 to bind, and increase the efficiency of bypassing this damage. This bypass effect has not been shown to be true for bulky lesions like a TT-Dimer. One hypothesis for the presence of the longer range tran-

sitions (between 0.6 and 0.95) observed could be the polymerases exhibits longer range backtracking alongside the short range backtracking, where it backtracks up to 8-10nt resulting in the high FRET state as the 5' end of the RNA is reeled back through the protein. It has previously been reported using ensemble and single-molecule techniques that polymerases are able to backtrack by several nucleotides backing up this hypothesis(Shaevitz et al., 2003; Artsimovitch and Landick, 2000). This is further backed up the data from the Rad26 experiments as discussed below.

As was discussed in Ch4, it is important to note that again these results are from a single set of DNA constructs and so the phenotypes that we observe cannot be said to be fully representative of the damage tested yet. In order to back up our observations new experiments should be carried out. This would include moving the damage to different positions within the template, allowing us to see if the dynamic or static backtracking we observe is maintained or if the dynamics are due to where the damage is located within the DNA, and not the damage itself. Furthermore, these are only one type of damage within their given classes. There is a plethora of both oxidative and bulky lesions, which have been shown previously to give rise to different responses both in vitro and in vivo. Therefore, it is likely that the observed dynamics would shift dramatically based on the type of damage inserted, and this could also be further probed with this assay by replacing the damage site. Based on these caveats, while it is likely that these phenotypes would be conserved for a given damage, it could vary drastically based on the severity and size of the damage itself and should be further analysed to assess this. Furthermore, as was the case in the Chapter 4 assays, the phosphorylation state of RNAPII may still be a concern. We do observe a heterogeneous response to each type of damage, for example 80% of RNAPII molecules are dynamic while encountering an 8-Oxo guanine lesions

while 20% is static. This is almost the reverse for abasic sites, where 90% are static. It is therefore possible that if we had a heterogeneous mix of RNAPII molecules they may lead to different outcomes. However, if this were the case it is also likely that we would see a similar split of static vs. dynamic molecules between the damages if one of the RNAPII phosphorylation states was inactive or was not able to backtrack. As we do not see this, I feel this is unlikely the cause of the heterogeneity we observe and therefore should not be a driving factor in the conclusions of this thesis, although it is something that should be noted. Efforts to test any potential heterogeneity or remove this from the sample is discussed previously in Chapter 4.

It has been hypothesised that Rad26 acts by sliding on DNA and 'pushing' the polymerase forward past the site of damage and inhibiting the backtracked state (Le et al., 2018; Xu et al., 2020). Addition of Rad26 showed an increase in bypass efficiency in the ensemble assay (Figure 5.9) when encountering an 8-Oxo Guanine lesion but not a TT-Dimer. Interestingly, a shift in the observed FRET states with two main differences were observed in both conditions. Firstly, the high FRET state at 0.95 disappears and a new peak at 0.83 appears, and the lower FRET peak which correlated with the site of damage is also shifted lower to 0.37. The shift in the high FRET peak correlates with RNAPII occupying a position 4-5 nucleotides upstream of the DNA, rather than the 10 nucleotides upstream it was occupying in the absence of Rad26. And the new peak at 0.37 correlates to RNAPII moving further over the damage.

It was confirmed that the loss of the high FRET peak (0.95) was due to Rad26 acting on actively elongating RNAPII and not causing a structural change in the pre-transcription state. The pre-transcription FRET state remained at a high FRET position (0.95) in the presence of Rad26 causing minimal change to the condition without Rad26 suggesting the

two dyes are in a similar position(Figure 5.13). Only when NTPs are added does Rad26 act on RNAPII causing these FRET shifts. Secondly, the observed dwell times observed for the relative transitions were also altered, in particular for the reverse transition back to the fully backtracked state where Rad26 reduced this stalled intermediate (Figure 5.9C). Together this suggests that as expected, Rad26 induces RNAPII to shift further over the site of damage and remain bound there for a longer time than in its absence, possibly by 'pushing' RNAPII over the damage. Furthermore, RNAPII cannot backtrack as far away from the site of damage than in the absence of Rad26, especially when encountering a TT-Dimer. Surprisingly however, once RNAPII does backtrack slightly ($\sim$ 2nt), Rad26 actually amplifies the amount of long-range backtracking events observed away from the damage to the higher FRET state (0.8, 4-5 nt), although this distance is still less than what was observed in the absence of Rad26 (10 nt) (Figure 5.9C and Table 5.2). This could possibly allow for other repair factors to bind and increase the efficiency of damage bypass that I do not probe in this assay.

This result provides greater detail as to the mechanism of how Rad26 interacts with stalled RNAPII but also backs up the hypothesis that the longer range transitions I observed are in fact real backtracking events. It has been reported that Rad26 will bind directly upstream of the polymerase and prevent it from backtracking to a upstream position by physically blocking it (Xu et al., 2017). The disappearance of the high FRET state (0.95) and its corresponding long range transitions in these experiments suggests that the protein is unable to backtrack to such an extent when Rad26 is blocking its path. Whereas without Rad26, the protein was free to move freely and backtrack to the high FRET state. This is an important part of the function of Rad26/CSB in the TC-NER pathway as the stalled RNAPII is the trigger for this process and not the damage itself like for the 8-Oxo

Guanine lesion, which will be fixed via BER. Rad26 decreases the dynamics of RNAPII, holding it on top of the damage allowing for more efficient TC-NER initiation (Brueckner et al., 2007).

Overall, I have demonstrated that different damages can lead to severely different mechanistic outcomes on RNAPII. While some lead to static stalling (abasic lesions) others cause RNAPII to act very dynamically. Even within these dynamics RNAPII can behave distinctly, where it can backtrack to different degrees and with different rates. Furthermore, Rad26 can increase the bypass of these damages by altering the backtracking kinetics of RNAPII. This assay could be further used to probe the dynamics of other damages or roadblocks such as nucleosomes or other oxidative damage. It could also be aided by the addition of other regulatory or elongation factors could be added which I propose could increase our understanding of how RNAPII interacts with different damages and importantly how the proteins involved in TC-NER interact together.

5.4 Supplementary Figures

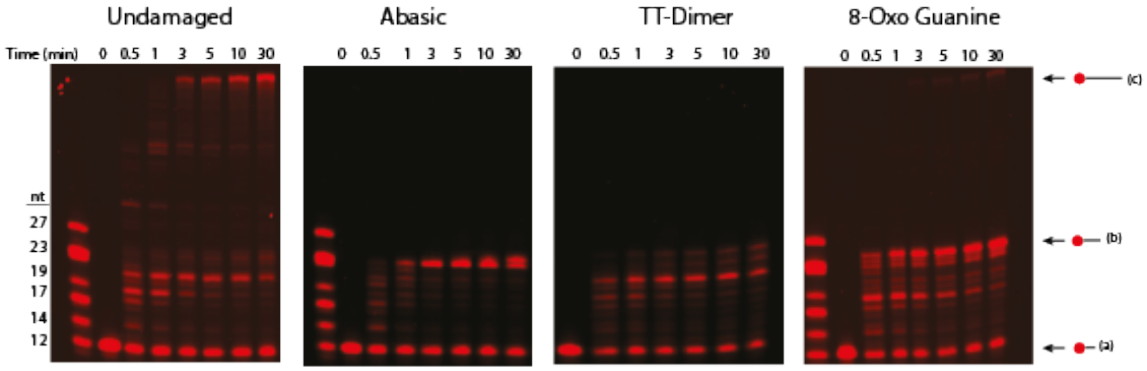


Figure 5.11: ***In vitro* transcription assay using 10 µM NTP showing NTP dependence.** Denaturing PAGE showing RNA extension over time from primer (a, 12 nucleotides) to full length transcript (c, 69 nucleotides) or various stalled positions (b) for the undamaged, abasic, TT-Dimer and 8-Oxo Guanine templates. 10 µM NTPs, measured via Cy5 fluorescence

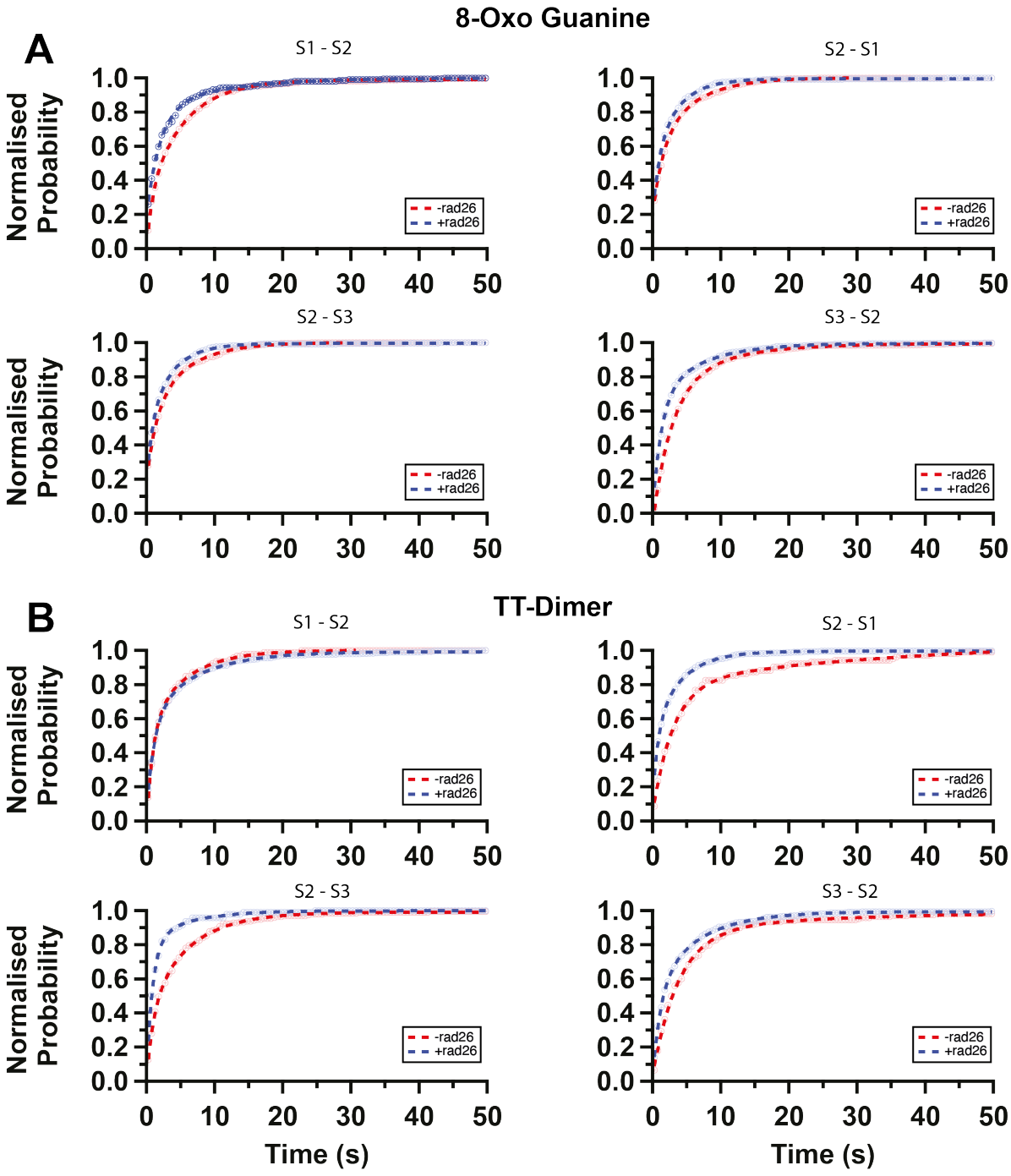


Figure 5.12: **Dwell time analysis of RNAPII transitions with and without Rad26.** Cumulative dwell time analysis of each transition observed for the 4 transitions. S1 - S2 (0.95 to 0.6) S2 - S3 (0.6 to 0.5) and their respective reverse transitions with and without Rad26 for (A) 8-Oxo Guanine and (B) TT-Dimer. Data was fitted to a double exponential. Red line = No Rad26, Blue line = With Rad26.

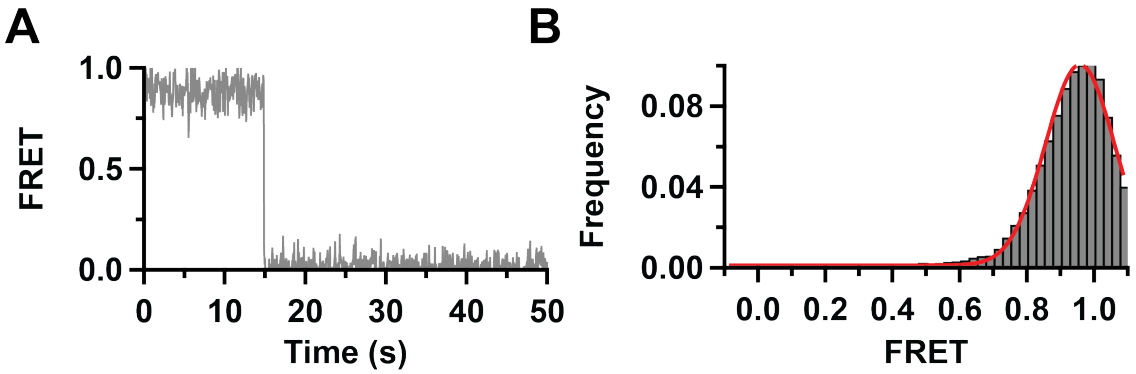


Figure 5.13: **Rad26 binding to RNAPII in absence of NTPs does not alter pre-transcribed FRET state.** (A) Example FRET trace of Rad26 bound to TEC in absence of NTPs. (B) Histogram of Rad26 bound to TEC showing a single high FRET state of 0.95, comparable to what was observed in the absence of Rad26 suggesting Rad26 has no effect on RNAPII in the absence of NTPs. N = 112, peak fitted to Gaussian fit.

Bibliography

I. Artsimovitch and R. Landick. Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proceedings of the National Academy of Sciences of the United States of America*, 97(13):7090–7095, 2000. ISSN 00278424. doi: 10.1073/pnas.97.13.7090.

F. Brueckner, U. Hennecke, T. Carell, and P. Cramer. CPD Damage Recognition by Transcribing RNA Polymerase II. *Science*, 315(February):859–862, 2007.

G. E. Damsma and P. Cramer. Molecular basis of transcriptional mutagenesis at 8-oxoguanine. *Journal of Biological Chemistry*, 284(46):31658–31663, 2009. ISSN 00219258. doi: 10.1074/jbc.M109.022764.

M. Kouchakdjian, V. Bodepudi, S. Shibutani, M. Eisenberg, F. Johnson, A. P. Grollman, and D. J. Patel. NMR Structural Studies of the Ionizing Radiation Adduct 7-Hydro-8-oxodeoxyguanosine (8-oxo-7/f-dG) opposite Deoxyadenosine in a DNA Duplex. %-H-dG(synydA(anti) Alignment at Lesion Site+. *Biochemistry*, 30:1403–1412, 1991. URL `https://pubs.acs.org/sharingguidelines`.

N. Kumar, A. F. Theil, V. Roginskaya, Y. Ali, M. Calderon, S. C. Watkins, R. P. Barnes, P. L. Opresko, A. Pines, H. Lans, W. Vermeulen, and B. Van Houten. Global and transcription-coupled repair of 8-oxoG is initiated by nucleotide excision repair proteins. *Nature Communications 2022 13:1*, 13(1):1–16, 2 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-28642-9. URL `https://www.nature.com/articles/s41467-022-28642-9`.

I. Lahiri, J. Xu, B. G. Han, J. Oh, D. Wang, F. DiMaio, and A. E. Leschziner. 3.1 Å structure of yeast RNA polymerase II elongation complex stalled at a cyclobutane pyrimidine dimer lesion solved using streptavidin affinity grids. *Journal of structural biology*, 207(3):270–278, 9 2019. ISSN 1095-8657. doi: 10.1016/J.JSB.2019.06.004. URL `https://pubmed.ncbi.nlm.nih.gov/31200019/`.

H. Lans, J. H. Hoeijmakers, W. Vermeulen, and J. A. Marteijn. The DNA damage response to transcription stress. *Nature Reviews Molecular Cell Biology*, 20(12):766–784, 2019. ISSN 14710080. doi: 10.1038/s41580-019-0169-4. URL `http://dx.doi.org/10.1038/s41580-019-0169-4`.

T. T. Le, Y. Yang, C. Tan, M. M. Suhanovsky, R. M. Fulbright, J. T. Inman, M. Li, J. Lee, S. Perelman, J. W. Roberts, A. M. Deaconescu, and M. D. Wang. Mfd Dynamically Regulates Transcription via a Release and Catch-Up Mechanism. *Cell*, 172(1-2):344–357, 1 2018. ISSN 10974172. doi: 10.1016/J.CELL.2017.11.017/ATTACHMENT/F59041D7-6C3D-450F-A8E2-F2ECC5D3CDF1/MMC1.PDF. URL http://www.cell.com/article/S0092867417313648/fulltexthttp://www.cell.com/article/S0092867417313648/abstracthttps://www.cell.com/cell/abstract/S0092-8674(17)31364-8.

J. Oh, J. Xu, J. Chong, and D. Wang. Structural and biochemical analysis of DNA lesion-induced RNA polymerase II arrest. *Methods*, (October 2018):0–1, 2019. ISSN 10462023. doi: 10.1016/j.ymeth.2019.02.019. URL https://linkinghub.elsevier.com/retrieve/pii/S1046202318303001.

J. Oh, A. M. Fleming, J. Xu, J. Chong, C. J. Burrows, and D. Wang. RNA polymerase II stalls on oxidative DNA damage via a torsion-latch mechanism involving lone pair–π and CH–π interactions. *Proceedings of the National Academy of Sciences of the United States of America*, 117(17):9338–9348, 2020. ISSN 10916490. doi: 10.1073/pnas.1919904117.

J. W. Shaevitz, E. A. Abbondanzieri, R. Landick, and S. M. Block. Backtracking by single RNA polymerase molecules observed at near-base-pair resolution. *Nature*, 426(6967):684, 12 2003. ISSN 00280836. doi: 10.1038/NATURE02191. URL /pmc/articles/PMC1483218//pmc/articles/PMC1483218/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC1483218/.

P. S. Thompson and D. Cortez. New Insights into Abasic Site Repair and Tolerance. *DNA repair*, 90:102866, 6 2020. ISSN 15687856. doi: 10.1016/J.DNAREP.2020.102866. URL /pmc/articles/PMC7299775//pmc/articles/PMC7299775/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC7299775/.

W. Wang, C. Walmacq, J. Chong, M. Kashlev, and D. Wang. Structural basis of transcriptional stalling and bypass of abasic DNA lesion by RNA polymerase II. *Proceedings of the National Academy of Sciences of the United States of America*, 115(11):E2538–E2545, 2018. ISSN 1091-6490. doi: 10.1073/pnas.

1722050115. URL http://www.ncbi.nlm.nih.gov/pubmed/29487211http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5856558.

J. Xu, I. Lahiri, W. Wang, A. Wier, M. A. Cianfrocco, J. Chong, A. A. Hare, P. B. Dervan, F. DiMaio, A. E. Leschziner, and D. Wang. Structural basis for the initiation of eukaryotic transcription-coupled DNA repair. *Nature*, 551(7682):653–657, 2017. ISSN 14764687. doi: 10.1038/nature24658.

J. Xu, W. Wang, L. Xu, J. Y. Chen, J. Chong, J. Oh, A. E. Leschziner, X. D. Fu, and D. Wang. Cockayne syndrome B protein acts as an ATP-dependent processivity factor that helps RNA polymerase II overcome nucleosome barriers. *Proceedings of the National Academy of Sciences of the United States of America*, 117(41):25486–25493, 2020. ISSN 10916490. doi: 10.1073/pnas.2013379117.

Chapter 6

Visualisation of RNAPII dynamics via optical tweezer based assays

6.1 Abstract

In order to visualise RNAPII transcription on different templates, longer DNA constructs must also be used. This allows tracking of RNAPII over thousands of base pairs rather than the 10-20 base pairs that can be detected via smFRET based assays. This can add greater mechanistic detail into the interaction with other DNA bound proteins such as repair proteins (Rad26, TFIIS), Cas complexes, DNA junctions and nucleosomes all of which can cause roadblocks to transcription. In order to study this, I built upon known DNA assays that can be optically trapped, and also developed novel assays for a similar purpose. Assays currently used often require biotinylation of RNAPII, but this is not always possible or desired (in the case of mammalian RNAPII). Therefore, new assays were developed combining optical trapping and confocal fluorescent microscopy to address this issue.

To this end two main assays were developed, one based on the well established ligation method where a TEC is ligated in between two biotinylated handles and the other a novel DNA construct that consists of both single-stranded and double-stranded DNA regions. Both constructs are still in initial phases of development and show binding of RNA and RNAPII to the desired positions. However neither as of yet permit transcription to be visualised. Despite this, surprising observations as to the DNA binding properties of RNAPII were observed. RNAPII showed significant binding propensity for ssDNA and DNA bubbles in an unregulated manner, shedding light onto the promiscuous nature of RNAPII and the mechanism of spurious initiation. Due to COVID I was unable to finish this project in time and therefore some of the results are incomplete. I will address in the discussion what steps would be taken next if there was more time to complete this project.

6.2 Results

6.2.1 DNA handle overhang length drives efficient TEC ligation

To study transcription using optical tweezers, long TEC constructs have in the past used a ligation method where short TECs (30-100 bp) are ligated in between two dsDNA handles of large length (>1-2 kb) (Chen et al., 2019). In order to trap these molecules, each side must contain a biotin modification either on the DNA handle ore the RNAPII molecule itself. Although this method has previously produced good results and is widely used, as described in Chapter 4 modification of proteins is not always desired or possible (Dangkulwanich et al., 2013; Chen et al., 2019). Instead, visualising transcription via fluorescence can be carried out using a correlative fluorescence optical trapping system (Lumicks C-Trap). In order to design such a construct, I first designed a TEC based on the ligation method commonly used, comprising of three main components. Two DNA handles (4 kb each) and a ssDNA oligonucleotide (30 nt) that ligates in between the two handles. This ssDNA region would then allow for both RNA and RNAPII binding to form a TEC.

To generate the DNA handles of large kb length, a 4.3 kb plasmid was used (pBR322) and linearised with two digestion enzymes producing 4 nt overhangs on each end. Handle 1 (H1) was digested using AatII and BmtI-HF, while handle 2 (H2) was digested with AvaI and StyI to produce linear DNA of 4.05 and 4.3 kb respectively. 100% digestion efficiency was detected yielding clean linear DNA of the expected length (Figure 6.1A). To form the 8 kb TEC, the intermediate oligonucleotide (8 kb TS, Table 6.1) was ligated in between H1 and H2, which was complimentary to the overhangs generated in the step above. An RNA primer complimentary to this intermediate oligonucleotide could be annealed also to allow

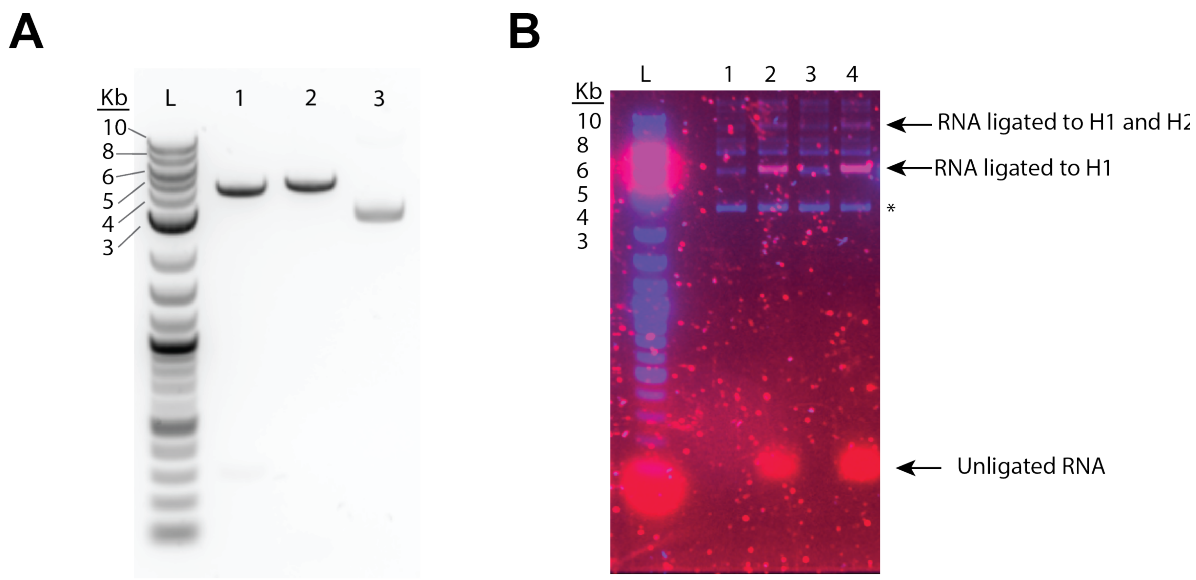


Figure 6.1: **Formation of 8 kb TEC via digestion and ligation.** (A) Digestion products generated from pBR322 plasmid. L = Ladder; 1 = H1 - 4Kb fragment digested using AatII and BmtI-HF; 2 = H2 - 4.2Kb fragment digested using AvaI and StyI; 3 = Undigested plasmid. Visualised via 0.8% TAE agarose and 1X SYBR gold staining. (B) Ligation of H1 and s and H2 with intermediate oligonucleotide to generate 8 kb TEC. L = Ladder, 1 = 1:1:1 ratio without RNA; 2 = 1:1:1 ratio with RNA; 3 = 1:2:1 ratio without RNA; 4 = 1:2:1 ratio with RNA. * = reformed plasmid. Visualised via 0.8% TAE agarose and 1X SYBR gold staining.

RNAPII to bind in a later step. It also validates whether the ligation has worked, as it can only be present in the 4-8 kb range if the intermediate oligonucleotide has also ligated. Although ligation of these components did show ligation of an 8 kb product containing the RNA primer as desired, it also showed various other products were formed during the ligation. Firstly, a large amount of the handles actually re-circularised reforming the original plasmid (Figure 6.1B). Secondly, some of the RNA can be seen co-localising with a band of approx 4 kb in length, which matches the size of the handles individually. The presence of RNA at this weight suggests that the intermediate oligonucleotide has ligated to H1 or H2 individually, but did not efficiently ligate together. In fact the total yield of expected product was only $\sim$ 1 ng (1% of the starting material) whereas the 4 kb band showed 5-times the amount of intermediate and RNA product. Finally, the presence of many other bands of longer length suggests that the handles are forming concatemers,

a process where the handles ligate to themselves several times without ligating to the intermediate oligonucleotide.

Alongside the concatemer issue, the low ligation efficiency was also likely due to the short overhangs produced (4 nucleotides) during plasmid digestion. Furthermore, due to the need to measure fluorescence on the optical trapping system it was deemed necessary to generate longer DNA handles so that accurate fluorescence measurements could be observed, given the resolution of confocal scanning is $\sim$ 250 nm(Lakowicz, 2006) . For all of these reasons stated above, I decided that using a plasmid based digestion method was not suitable. Instead, there was a need to generate long kb length DNA another way that would allow for a higher yield.

Table 6.1: DNA Sequences used for 8Kb TEC

Name	Sequence (5' to 3')	Modification
8Kb TS	CCGAT CCTGG TCGCG TCTGA TTACG CTGAG CTACG TAGCG CGTCG TACGT	5' Phosphate
RNA Primer	rArCrGrCrG rArCrCrArG rGrA	5'Amine Group

6.2.2 TEC with long DNA handles permits RNAPII transcription

This TEC required the same components as in the above method, two handles and an intermediate TEC. The main issues with the method above was that the ligation efficiency was too low due to the short overhang region (4 nucleotides) and the total length was too short to accurately measure fluorescence via optical trapping, which was limited due to the size of commercially available plasmids and digestion enzymes.

Instead, the handles for this construct were amplified using autosticky PCR from λ DNA using two sets of primers (Gál and Kálmán, 2002). This allowed for customisable control

Table 6.2: DNA Sequences used for 16Kb TEC

Name	Sequence (5' to 3')	Modification
6Kb	GCAGC TGTGA G**X**CGA ATGCT GCAGC	5' Phosphate
Primer F	GTCAC AACA	**X** = Abasic Site
9Kb	AGGTC GCCGC C**X**ACA GCGGG GAATA	5' Phosphate
Primer F	TCAGG TGCT	**X** = Abasic Site
10Kb	AGGTC GCCGC C**X**GGA TAAAT GGGCG	5' Phosphate
Primer F	CTGTA TGTC	**X** = Abasic Site
16Kb	AGGTC GCCGC C**X**GAG CCTGA AGATG	5' Phosphate
Primer F	ATGTG CTGA	**X** = Abasic Site
Biotin	**T**CA**T**C **T**GGCG GAATG GGTGT GGTAA	Biotin dT
Primer R	AAC	
16 kb TS	GGGCG GCGAC CTCAT GCGTA CACGC GTTTG TACAC CACAT GGTTG GACAC CATCC TGGTC GCGTC ACCAT CACCC ATAGT TCATT CCACT AGGTG TCCTT	5' Phosphate
16 kb NTS	ACTCA CAGCT GCAGG ACACC TAGTG GAATG AACTA TAGGG TGATG GTGAC GCGAC CAGGA TGGTG TCCAA CCATG TGGTG TACAA ACGCG TGTAC GCATG	5' Phosphate
RNA Primer	rArCrGrCrG rArCrCrArG rGrA	5'Amine Group

over the handles total length and overhang length. Furthermore, as it is an amplification based method rather than a digestion based one, the amount of starting material will be significantly increased allowing a higher amount of final product to be generated. The primers used here contained four biotin molecules on one primer and an abasic region on the second primer. When used in a PCR reaction, the polymerase stalls at the abasic lesion forming a ss-DNA overhang, while the remaining region of the DNA handle is fully double-stranded and also contains the required biotin molecules. Four main handles were amplified of different lengths, 6, 9, 10 and 16 kb. These handles were not only of a larger length than produced previously, but also contained overhangs of longer length (12nt) which would allow for better ligation efficiency. A single amplified band was detected for almost all handle lengths when analysed via agarose electrophoresis, with a total yield of 15-20µg (Figure 6.2A). The exception being the 16 kb amplified product, where two bands can be detected, one at the correct length and the other that remained stuck in the

well, suggesting some form on concatemer was also present (Figure 6.2A). For the purpose of the ligation, the 6 and 10 kb handles were utilised as they produced the highest and purest yield. The intermediate TEC contained overhangs complimentary to the overhangs generated via autosticky PCR. Ligation of the handles with the TEC produced a band at the desired length (16.1Kb) but was not completely efficient where a large amount of handles alone was also detected (Figure 6.2B). A 2:1 ratio of TEC:handles produced the highest ligation efficiency of 25%. This was calculated to be approximately 20 ng/µL. Although this is a low amount for traditional biochemical assays, it is sufficient for single-molecule based experiments.

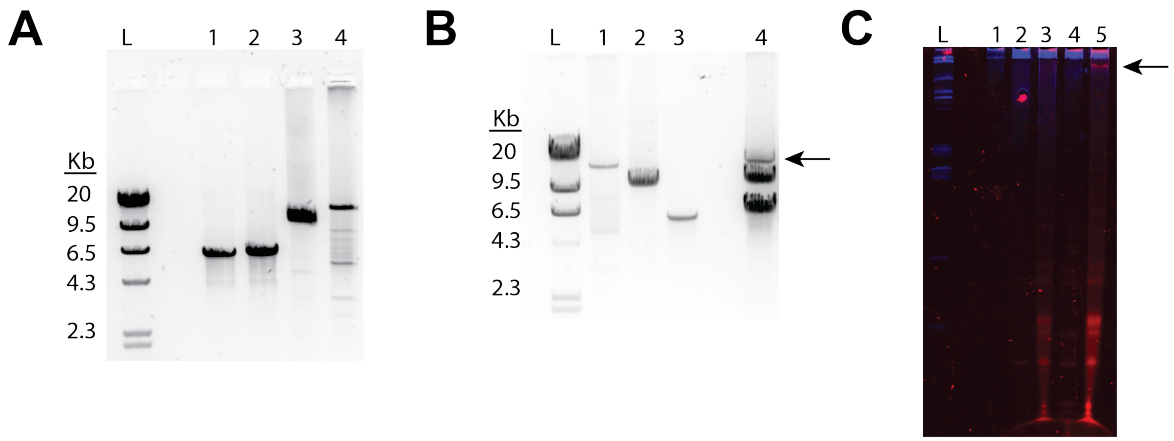


Figure 6.2: **Formation of 16Kb TEC via PCR and ligation.** (A) PCR products generated from λDNA. L = Ladder; 1 = 6Kb fragment; 2 = 6Kb fragment; 3 = 10Kb fragment; 4 = 16Kb fragment. Visualised via 0.8% TAE agarose and 1X SYBR gold staining. (B) Ligation of 6Kb and 10Kb handles and intermediate TEC to generate 16Kb TEC (black arrow). L = Ladder, 1 = 16Kb PCR fragment; 2 = 10Kb PCR fragment; 3 = 6Kb PCR fragment; 4 = Ligation product. Visualised via 0.8% TAE agarose and 1X SYBR gold staining. (C) Bulk Assay of 16Kb TEC. RNA extension can be seen upon addition of 1mM NTPs (black arrow). L = Ladder, 1 = 16Kb PCR product; 2 = 16Kb TEC no NTPs, 1:1:1 ratio; 3 = 16Kb TEC + 1mM NTPs, 1:1:1 ratio; 4 = 16Kb TEC no NTPs, 1:2:1 ratio; 5 = 16Kb TEC + 1mM NTPs, 1:2:1 ratio. Bands separated via 12% PAGE and visualised via Cy5 on RNA primer (red, RNA) and staining with 1X SYBR gold (blue, DNA).

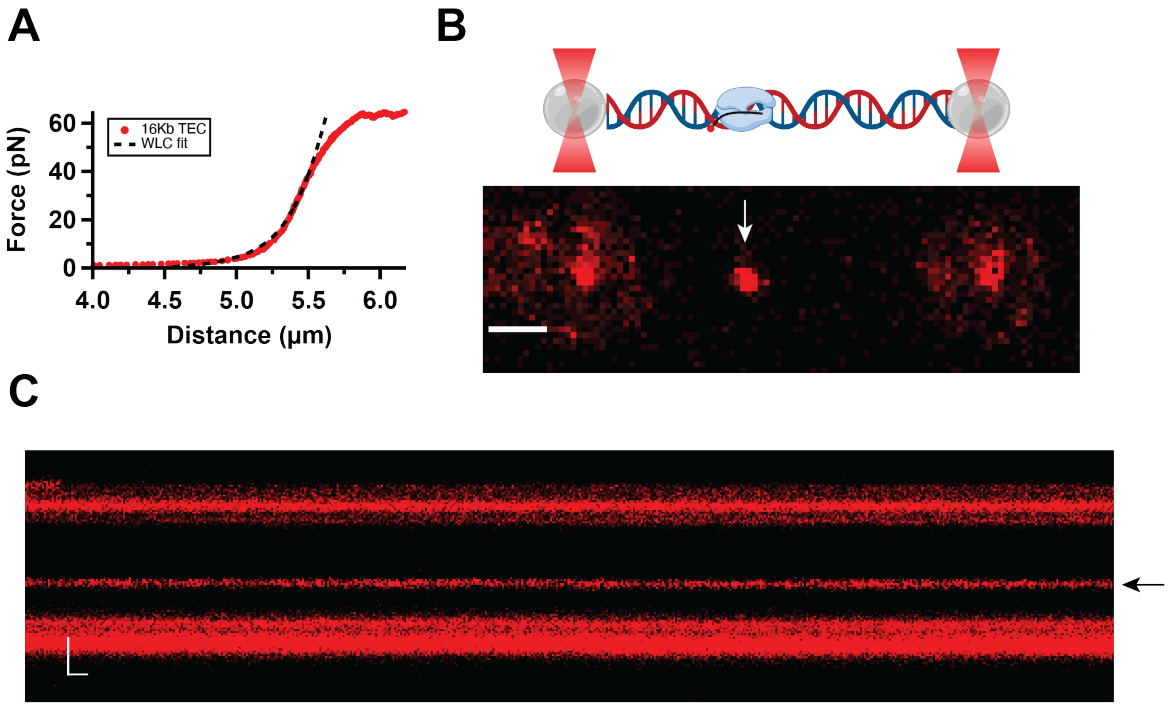


Figure 6.3: **16kb TEC held between optically trapped beads.** (A) Force-Distance curve of 16Kb TEC pulled from 0-65 pN. Red Curve = Extension; Black dashed line = WLC fit. (B) Confocal image of 16Kb TEC held at 5 pN between 1.78 µm beads with the Cy5 RNA primer visualised (arrow) at the expected position of ligation (6 kb from the left side, arrow) and example schematic of where RNAPII is bound.Scale bar = 1 µm (C) Kymograph of 16 kb TEC (as (B)) in presence of 1 mM NTPs showing no movement from the ligation site. Scale bars = 2 µm (vertical) and 1 second (horizontal).

6.2.3 16 kb TEC permits transcription in bulk but not at single-molecule level

Before testing in the optical trapping system, NTPs were added to this 16 Kb TEC and transcription observed via gel electrophoresis utilising the Cy5 fluorophore on the RNA primer. 1 hour of incubation with 1 mM NTPs caused an increase in the length of RNA detected (Figure 6.2C). Three main populations were found. The first being at approx. 100 nt (corresponding to transcription of the TEC alone), the second being of high MW only just entering the gel (several Kb, black arrow) and the final being stuck in the well suggesting it is too large to enter the gel.

This 16 kb TEC was then captured between optically trapped beads (1.76µm) and the force-distance (FD) curve measured. This is done by pulling on DNA from low (0 pN) to high forces (65 pN) and the measured curve can be fit to a Worm-Like Chain (WLC) model to identify the contour length of DNA (L_{cDNA}) (Equation 2.10).

If ligated correctly, the DNA molecule should exhibit an L_c of $\sim$ 5.5 µm based on the base to base distance of dsDNA (0.34 nm/bp) (Wang et al., 1997). Captured DNA had a L_c of 5.6 µm, matching what would be expected for a dsDNA of this length (Figure 6.3A). Stably bound Cy5 signal at the expected position could be seen in over 80% of captured molecules suggesting a high ligation efficiency (Figure 6.3B).

However, incubation of these constructs in a channel containing 1 mM NTPs showed no change in the observed Cy5 signal, where it remained stably bound to the initial binding region over several minutes (Figure 6.3C).

6.2.4 Generation of hybrid ss-/ds-DNA constructs

Alongside the ligation method for generating long TEC constructs, a novel method was developed in collaboration with a PhD student in the lab of Simon Boulton (Ondrej Belan, Francis Crick institute) to generate a hybrid DNA construct that contained both ssDNA and dsDNA regions, published last year (Belan et al., 2021). This would in turn allow for an RNA primer and RNAPII to bind to the ssDNA region, using the dsDNA downstream region as the transcription template. Constructs were made as per Section 3.5.2 where hybrid DNA constructs of different gap sizes (200nt - 18 knt) were validated both by force and fluorescence based methods.

Briefly, this involved using tortionally constrained λ-DNA, generated via ligation of looped-DNA onto its cos-ends (Figure 3.3) (King et al., 2019). Addition of these end-caps causes

Table 6.3: DNA Sequences used for Hybrid DNA

Name	Sequence (5' to 3')	Modification
	rGrUrGrArU rArArGrUrG rGrArArUrG	5' Phosphate
λ 2	rCrCrArUrG rGrUrUrUrU rArGrGrArG	
	rCrUrArUrG rCrUrGrUrU rUrUrG	
	rCrArGrArT rArTrArGrC rCrTrGrGrT	5' Phosphate
λ 4	rGrGrTrTrC rGrUrUrUrU rArGrGrArG	
	rCrUrArUrG rCrUrGrUrU rUrUrG	
	rUrUrCrArU rUrArArUrC rCrCrUrGrA	5' Phosphate
λ 4.01	rArCrUrGrU rGrUrUrUrU rArGrArGrC	
	rUrArUrGrC rU	
	rArUrUrGrC rCrArGrGrC rUrUrArArA	5' Phosphate
λ 4.02	rUrGrArGrU rGrUrUrUrU rArGrArGrC	
	rUrArUrGrC rU	
	rGrGrCrCrA rUrGrUrArA rGrCrUrGrA	5' Phosphate
λ 4.1	rCrUrUrUrA rGrUrUrUrU rArGrArGrC	
	rUrArUrGrC rU	
	rGrCrCrArU rUrCrUrGrC rUrUrArUrC	5' Phosphate
λ 4.2	rArGrGrArA rGrUrUrUrU rArGrArGrC	
	rUrArUrGrC rU	
	rGrGrCrArA rUrGrCrCrG rArUrGrGrC	5' Phosphate
λ 5	rGrArUrArG rGrUrUrUrU rArGrGrArG	
	rCrUrArUrG rCrUrGrUrU rUrUrG	
RNA λ Primer	rArArUrGrC rGrCrCrGrC rC	5'Amine Group

the DNA molecule to become fully closed, an essential step in the production of hybrid DNA. Next, two sets of Cas9 nickase RNPs are utilised which induce nicks along the same strand of DNA, but in different positions (controlled by the guide RNA sequence). By varying the distance between the two guide RNAs, the size of the gap could be changed (Table 6.3). λ 4 guide RNA partnered with any of λ 4.01, 4.02, 4.1, 4.2 or 5 would produce ssDNA gaps of 0.1, 0.2, 1, 2 or 5 kb in length. To produce the gap size of 18 kb, λ2 and λ5 were utilised together. This produced a dsDNA containing two nicks at discrete positions, in order to remove the region of DNA between the nick sites, optical tweezers are used to pull on the DNA. By pulling on the DNA, the two strands of DNA will melt apart. Due to the molecule being tortionally constrained, the two strands will never completely melt apart, allowing me to keep hold of the DNA tether. Upon re-annealing of the dsDNA,

the region between the nick sites would have been removed, leaving a region of ssDNA exposed and producing the hybrid DNA construct (Figure 3.3). The size of the constructs were verified by the generation of an FD curve as above (Figure 6.4A).

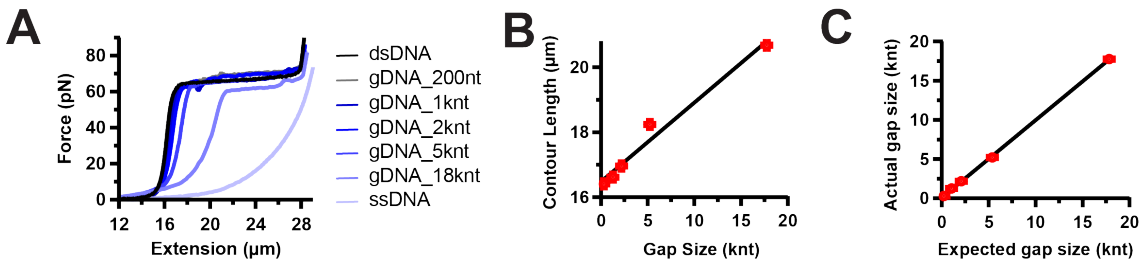


Figure 6.4: **Valdiation of gap sizes of hybrid DNA constructs.** (A) Force-distance curves of λ DNA,λ ssDNA and different hybrid DNA constructs of various gap sizes (200nt - 18Knt). Blue line shows WLC fit of each construct. (B) Contour length of the hybrid constructs calculated from WLC extracted from 2A. Error bars indicate standard deviation. Fit to a linear curve (black) (C) Calculation of expected gap size (from 2B and Equation 6.2) and comparison to the actual gap size. Error bars indicate standard deviation. Fit to a linear curve (black). Figure adapted from (Belan et al., 2021).

As the region of ssDNA increased from 200nt - 18Knt, the L_c of DNA also increased (Figure 6.4B) due to ssDNA being more flexible than dsDNA. The estimated gap size (L_{gap}) was then calculated as per the following equation which matched the actual gap size closely (known due to the specific guides used) (Figure 6.4C and Table 6.4):

$$L_{gap} = \frac{L_c(\lambda gDNA) - L_c(\lambda dsDNA)}{d(ssDNA) - d(dsDNA)} \tag{6.1}$$

Where $d(ssDNA)$ and $d(dsDNA)$ are the distances between nucleotides/base-pairs which are 0.6 nm.nt^{-1} and 0.34 nm.bp^{-1} respectively (Wang et al., 1997).

Table 6.4: Expected and Actual Gap Sizes

gDNA	Expected Gap (nt)	Actual Gap (nt)
gDNA 0.2Knt	227	298
gDNA 1Knt	1041	1262
gDNA 2Knt	2075	2188
gDNA 5Knt	5374	5217
gDNA 18Knt	17823	17756

Force based validation only detects if the gap is of the correct size, not where the gap is. In order to validate its location, fluorescently tagged RPA (eGFP) was used. RPA is a known ssDNA binding protein and should only bind to the ssDNA region, which can be imaged via confocal scanning. RPA was detected bound specifically to the ssDNA gap at the expected location within each construct (Figure 6.5).

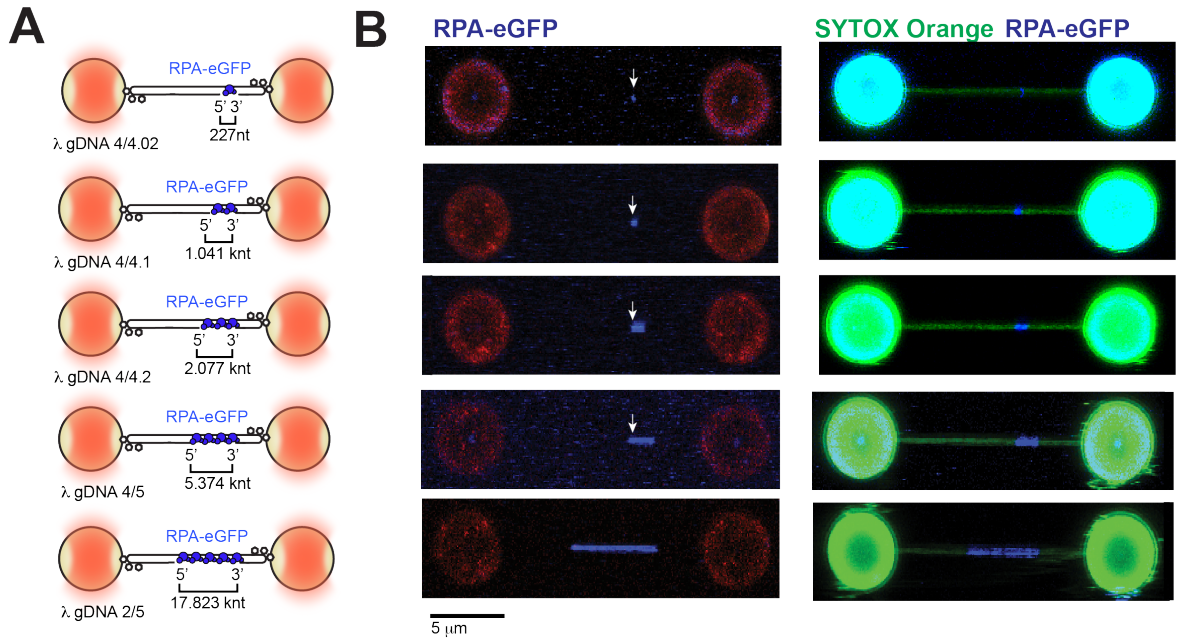


Figure 6.5: **Validation of gap sizes of hybrid DNA constructs determined via fluorescence.** (A) Schematics of the different hybrid constructs and the respective position of each gap (B) Representative confocal images of RPA-eGFP(5 nM) bound to each gapped construct (left panel) and dsDNA stained via 50 nM SYTOX Orange (right panel). Constructs held at constant 5 pN between 4.8 µm beads. Arrows show position one edge of the Cas9 induced cut site indicating the 5' end for reference. Figure adapted from (Belan et al., 2021).

6.2.5 Spurious binding of RNAPII driven by DNA accessibility

In order to use the gapped construct to measure transcription, both an RNA primer and RNAPII must bind first. A Cy5-RNA primer was designed to target specific positions within the gapped region (Table 6.3). RNA binding could not be detected at 10 nM, but was observed at an increased concentration range of 0.1 - 1 µM (Figure 6.6A-C).

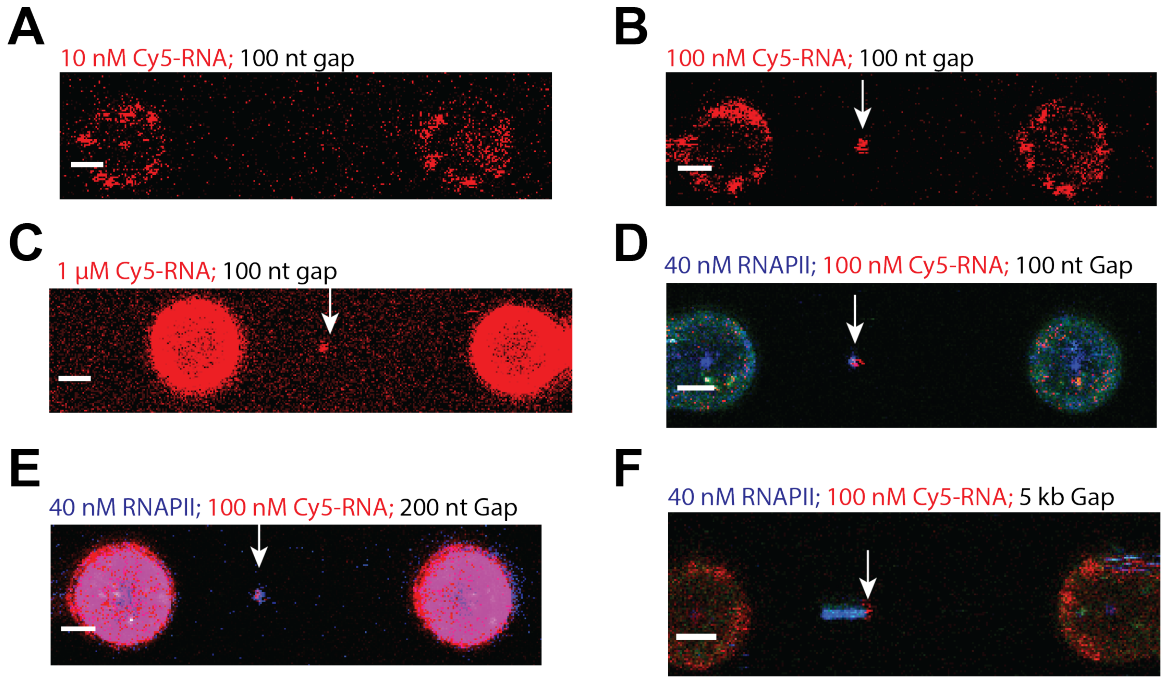


Figure 6.6: **RNA binds to ssDNA region within hybrid construct at expected location dependent on concentration.** (A - C) Cy5-RNA Primer binding to Hybrid DNA (100nt gap size) showing no binding at (A) 10 nM, but binding at expected position (white arrow) at (B) 100 nM and (C) 1µM. DNA tethered between 4.8 µm beads. Scale bar = 2 µm. (D - E) Alexa488-8WG16 RNAPII (40 nM) binding to Hybrid DNA at different sizes showing co-localisation with Cy5-RNA primer (100nM, white arrow) and also non-specifically binding to ssDNA gap. (A) 100 nucleotide gap (B) 200 nucleotide gap (C) 5 kb gap. DNA tethered between 4.8 µm beads. Scale bar = 2 µm

Once bound, it was then incubated in a channel containing RNAPII. RNAPII was imaged via and Alexa-488 labelled antibody targeting the CTD of RNAPII (8WG16), this is the same antibody used in the purification process. RNAPII was detected bound within the ssDNA gap co-localised with the RNA primer (Figure 6.6D-E). However, at low forces a significant amount of RNAPII was also detected bound to the DNA within the ssDNA region that was not co-localised with the RNA primer (5 pN). This was especially prevalent at longer gap sizes such as 5 kb (Figure 6.6F). Furthermore, when imaging in a channel containing RNAPII but no RNA primer, upon increasing the force from 5 - 60 pN the number of DNA bound RNAPII molecules increased dramatically to positions outside of the Cas9-induced gap (Figure 6.7A). This in turn reduced back to the original amount when the force was lowered back to 5 pN. Based on these results, this suggests that

RNAPII has a high propensity for ssDNA and DNA bubbles and will bind spuriously and promiscuously to them. These regions of DNA can occur naturally within the eukaryotic genome due to various reasons including transcription, replication, DNA repair, chromatin state and super-coiling (Nogales et al., 2017; Costa et al., 2013).

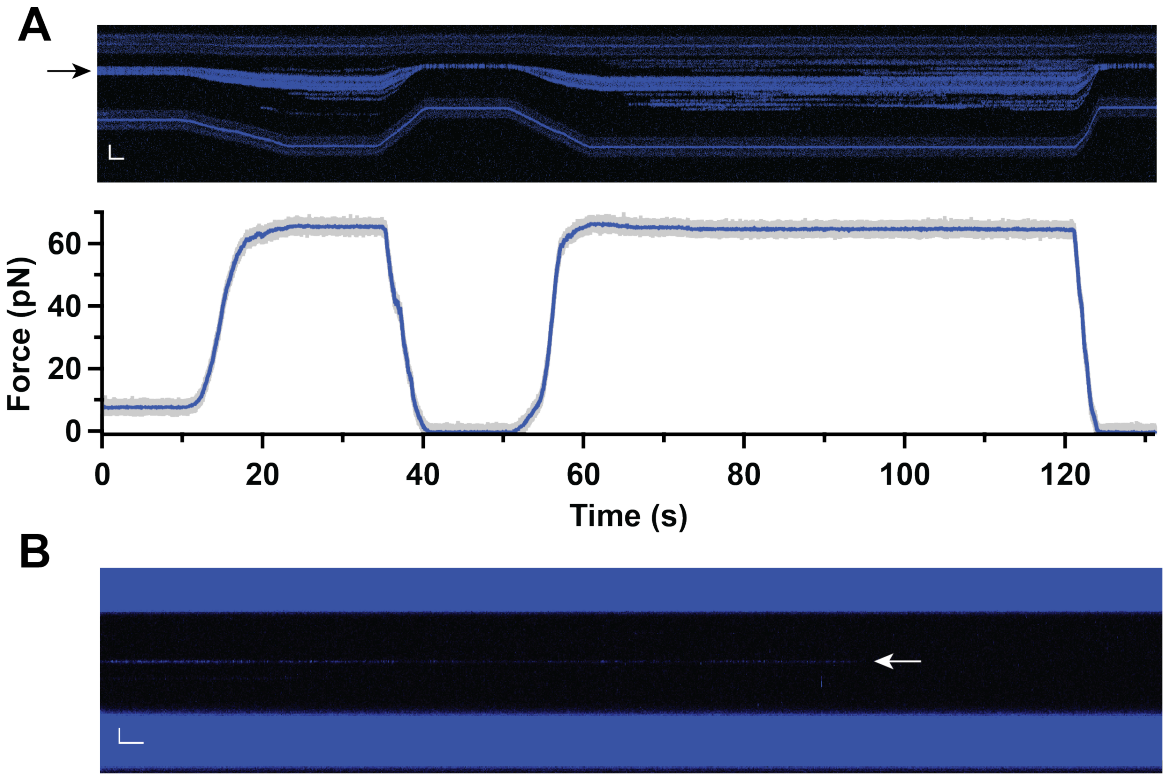


Figure 6.7: **RNAPII binding to ssDNA regions perturbing transcription elongation.** (A) Alexa488-8WG16 RNAPII binding to the ssDNA gap (arrow) at low forces (5 pN) but increased binding is seen upon higher forces (65 pN) where new RNAPII binds to bubbles within the dsDNA region. DNA tethered between 4.8 µm beads. Scale bar = 2 µm (vertical) and 2 seconds (horizontal). (B) Kymograph of 200nt hybrid DNA in presence of 1 mM NTPs showing no movement from the bound polymerase visualised via Alexa488-8WG16. Scale bars = 2 µm (vertical) and 1 second (horizontal).

The small gap sized DNA (200nt) with both the RNA primer and RNAPII was also placed in a channel containing NTPs to initiate transcription. Transcription was to be measured via Alexa488 fluorescence as above. However, no transcription was detected and the RNAPII and RNA remained stably bound over time (Figure 6.7B) as was the case with the other construct (Section 6.2.1).

6.3 Discussion

DNA constructs to study transcription using optical tweezers and fluorescence must meet certain requirements. Firstly, it must be of long enough length to be held in an optical trap and visualised via fluorescence, to give high spatial resolution. Secondly, it must be modified on each end to allow attachment to microspheres. And finally, for transcription it must contain a dsDNA region that permits transcription to occur, while having a TEC/single-stranded region present to allow RNA and RNAPII to bind.

Attempts to generate this construct via ligation based methods showed some promise but was not very efficient and generating handles of large kb length can be quite challenging. Initial attempts to do this via enzyme digestion showed little success. Although 4 kb handles were generated with a high purity (Figure 6.1A), the ligation efficiency was very low (1%) and the RNA showed predominantly ligation to only one handle, and not to both generating the 8 kb TEC as desired (Figure 6.1B). Furthermore, a big issue was concatenation of the handles to themselves. This occurs due to the overhangs having a palindromic sequence (e.g. ACGT) on one end which in turn allows for concatenation and circularisation of the handles (Ukai et al., 2002). This can inhibit the production of the linear 8 kb desired product and reduce the overall yield. There are ways to avoid this issue, mainly by purification of individual handles via gel extraction and de-phosphorylation of the handles to remove 5' phosphates. Although this would prevent ligation of the handles to one another (as they now lack the required terminal phosphate) it would also prevent the ligation of the intermediate oligonucleotide as the 3' 5' junction on one end. Although there are other ways to fix this issue, the length of overhangs on each handle also presented an issue. Digestion of plasmid DNA with restriction enzymes produces 4nt overhangs in almost all cases. This is very short and does not allow for great annealing or

ligation efficiencies *in vitro*. These short overhangs would also have presented an issue for the addition of biotin onto this construct. Ligation of short oligonucleotides containing biotins is one option, but as already discussed this leads to very poor efficiencies and yields. Another option is filling in the ends using Klenow Fragment (Section 3.5.1), but due to there being only 4nt the maximum number of biotin-NTPs that would be added would be 2-3 which is not always sufficient for successful tethering to optically trapped beads. Instead, longer overhangs on the handles were desired in order to increase ligation efficiency. This is why a new approach was developed, utilising autosticky PCR (Section 6.2.2) (Gál and Kálmán, 2002).

Autosticky PCR allowed the generation of new handles that were not only longer in length (tp to 10 kb each) but also for longer and more customisable overhangs (Figure 6.2A. This is simply controlled by the addition of a known sequence to the PCR primer set on one end. Utilising these handles permitted much higher ligation efficiency of our desired substrate (25%) and prevented any other concatemer products from forming (Figure 6.2B). It also allowed biotin groups to be added easier as they were pre-incorporated onto the other primer. Once generated, this construct was successfully visualised within the optical tweezer system (Figure 6.3A). Although DNA of the correct length was captured, and both the RNA primer and RNAPII showed stable binding over time, no transcription was detected on this construct (Figure 6.3). This was unexpected as transcription was detected in the ensemble assay where RNA of several kb in length was visualised 6.2C). There could be several possibilities as to why no transcription was detected despite all necessary factors being present. Firstly, it could be due to the handles not ligating to the TEC fully and therefore effectively leaving a nick in the DNA backbone on one or both strands inhibiting processive transcription. The break would be only 50-60 nucleotides

downstream of the polymerase and so if the polymerase were transcribing only a small distance then this would not be detected by confocal scanning which is diffraction limited in its resolution (250 nm). This is backed up by the fact that I do see bands of such sizes appearing in the ensemble assay suggesting there could be a blockage here. However, there will also be some free TEC in the ensemble assay which could also explain the presence of this band.

Another possible cause could be a force-based issue. When using optical tweezers there are several options that can be utilised to vary the force acting on the DNA. One option is to pull the DNA to the desired force and then visualise the fluorescence, in this mode you are allowing the force to fluctuate from this desired position. In a low force regime (<5 pN) the DNA would be in a relaxed and flexible structure, while higher forces would induce straightening of the DNA but at too high a forces would induce bubbles and nicks to appear (Section 2.3). The other option is applying a force clamp - where the force is held constant at a desired force and changes in the force are offset by changing the distance between the two beads. These experiments were carried out in both modes utilising a force of 5 pN predominately, although slightly higher and lower forces were tried (Figure 6.3). This is in line, although perhaps slightly lower, with the forces used in previous published work on RNA Polymerase II (Dangkulwanich et al., 2013; Chen et al., 2019).

However, neither of these issues would solely explain why transcription was observed in the ensemble assay, but not the optical tweezer setup. Another issue I encountered was the lifetime of the fluorescent dyes bound to the RNA or RNAPII. In > 80% of cases the fluorophores bleached within 1-2 minutes (Figure 6.3). This bleaching time is quick and may limit the ability to visualise transcription. The ensemble assay showed transcription after 1 hour, and other single-molecule assays have shown that complete transcription

may take several minutes even at high NTP concentrations. If the dyes have bleached by this time, I will not see any protein translocation even if transcription is permitted. In order to increase the lifetime of these dyes, different oxygen scavenging systems could be utilised (Glucose Oxidase/Catalase) that may enhance the lifetimes of certain fluorophores. Furthermore, the exposure time of the confocal scanning microscope could be altered as to increase the lifetime of the dyes. Most experiments were carried out using exposure times of 40-100 ms, other studies that have captured diffusing molecules have used times in the seconds range and this will allow for longer observation of fluorescent signal (Anand et al., 2021; Losito et al., 2021). Different dyes could also be tested to see if they react differently in this assay under the above mentioned conditions (Cy3B, Alexa based dyes etc). Another possible way around the lifetime issue would be to utilise an RNA selective dye (SYTO RNA Select, ThermoFisher). This dye selectively binds to ssRNA and as the length of the RNA increases the dye would bind to it as the polymerase transcribes along the template. Because fresh dye would continuously bind to the newly synthesised region of RNA, fluorescence bleaching should not be an issue as I will always visualise the nascent RNA adjacent to the polymerase. Although this is a cell-based stain and not designed for *in vitro* assays, it has been utilised before in unpublished work in the lab looking at T7 polymerase on λDNA where transcription was visualised (Figure 6.8).

The generation of transcription complexes using the hybrid DNA construct also showed some promise, but had similar issues when attempting to visualise transcription. The constructs I was using have been validated previously and so the generation of these was not an issue (Belan et al., 2021). The first issue arose attempting to anneal the RNA primer to the ssDNA region, where a high concentration was needed to see efficient

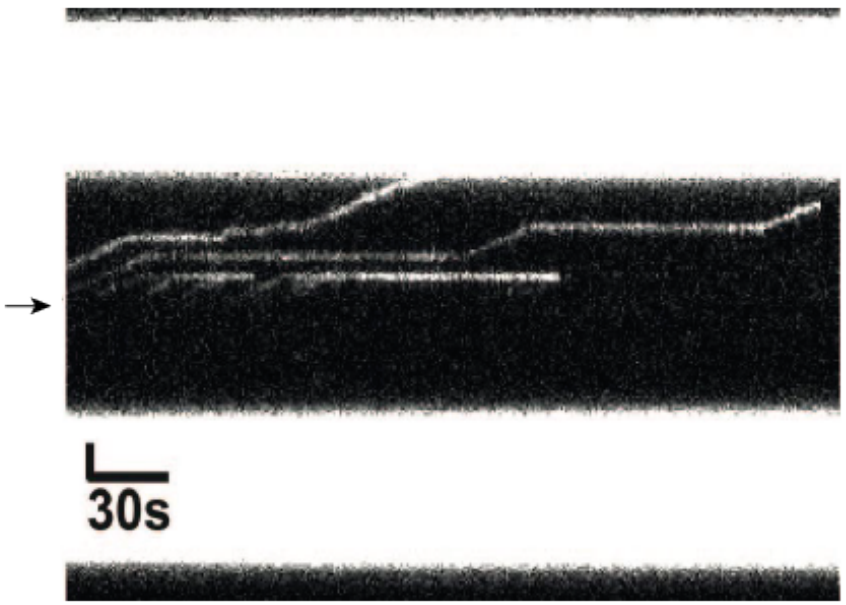


Figure 6.8: **Visualising T7 mediated transcription of λDNA via SYTO RNA select dye.** T7 polymerase was bound to promoter site (black arrow) and incubated with 1mM NTPs permitting transcription. 5 nM SYTO RNA select dye bound to newly synthesised RNA showing multiple transcription sites along the DNA often pausing ate random sites. Unpublished work carried out by Matt Newton

binding (0.1-1 µM) (Figure 6.6). This meant that the background fluorescent signal of the primer (Cy5) was very high and made imaging quite difficult. This could be easily overcome by mixing the Cy5-labelled RNA with a non-labelled version so that although the overall concentration is high enough, the level of Cy5 signal will not be too high. Only the molecules showing a fluorescently bound molecule would then be used for later experiments.

Secondly, as was the case in the ligation method no transcription was detected despite RNAPII being bound to the template (Figure 6.7). Some of the possibilities for this could be the same as with the ligation method, including force-based issues and fluorescent life-time issues. In the case of the latter, like I discussed above a different approach utilising the RNA selective dye could be utilised here also. Another possibility is that the polymerase is getting stalled at the junction of single-stranded and double-stranded region. If this were the case the polymerase would not progress past it and due to the distance

travelled being under the spatial resolution (250 nm) I would not see this. The RNA primer was placed adjacent to the ds-DNA region as to best match the desired substrate. One way to possibly resolve this issue would be to anneal the primer (and hence RNAPII) in the middle of the ssDNA gap and then anneal a second DNA strand complimentary to the gap (NTS) to form a transcription bubble. This should then allow the polymerase to function on the template as expected. One other possibility is that multiple polymerases could be bound to the ssDNA region at once in turn blocking potential elongation (Figure 6.7). A high amount of RNAPII signal was observed in the gap region, especially at higher forces, and it has been previously reported that converging RNAPII can cause transcriptional stress (Hobson et al., 2012). This could be remedied by using smaller gap sizes and lower protein concentration to only have a single polymerase bound permitting transcription. The original assay was carried out using 40-50 nM of RNAPII, which although suitable for ensemble assays is high for optical tweezer assays and is most likely why I observed multiple binding events. To prevent this issue the concentration would be lowered to 5-10 nM as is the case with other single-molecule experiments (Belan et al., 2021; Anand et al., 2021).

Although the assay did not permit transcription, the presence of RNAPII binding non-specifically to ssDNA and the induced binding at high forces was surprising. As the gap size increased as did the amount of RNAPII binding, despite not co-localising with an RNA primer (Figure 6.6). This is further illustrated when increasing the gap size artificially via force. Increasing the force not only stretches the Cas9-induced gap, increasing the distance between bp and hence its total length, but can also induce bubbles into the DNA template other positions. Upon increasing the force to 65 pN, the amount of RNAPII bound to the gap increased, and more binding was detected to other regions along

the DNA, likely binding to these induced bubbles. This is backed up by the fact that when the force is lowered (back to 5 pN) these binding events disappear as the bubbles disappear and only the RNAPII bound to the original Cas9-induced gap can be detected. DNA bubbles and ssDNA regions appear heavily throughout the eukaryotic cell including regions of transcription, replication and DNA super-coiling (Nogales et al., 2017; Fosado et al., 2021).

It has been previously reported that RNAPII is a promiscuous enzyme and if not regulated, will bind non-specifically to DNA (Wade and Grainger, 2018). This spurious binding binding property could be very detrimental for the cell, causing RNAPII to bind non-specifically to regions within the genome inducing spurious initiation. Spurious initiation has been shown to occur in both prokaryotic and eukaryotic organisms often leading to developmental diseases (Scandaglia and Barco, 2019; Sen et al., 2022). As to what drives this spurious initiation in eukaryotes is not fully understood, but current hypotheses suggest epigenetic markers including H3K36 methylation are important. This marker is down-regulated at spurious initiation sites, preventing chromatin compaction and increasing the accessibility of the DNA to proteins such as RNAPII (Wade and Grainger, 2018; Pan et al., 2018). The data here shows that our RNAPII also follows this trend and has a propensity to bind unregulated to ssDNA and force-induced bubbles, which is further promoted at regions where DNA is naturally unwound. Our data here is consistent with previous hypothesis and together this suggests that spurious initiation might be driven by the accessibility of the surrounding DNA in turn promoting RNAPII promiscuous nature. To control this within the cell, elongation and regulatory factors and epigentic markers will regulate this process to limit spurious binding. Addition of these regulatory factors including DSIF, NELF and TFIIF to this assay could further shed light onto the

spurious binding mechanism of RNAPII, which should inhibit this activity.

Both of these assembly methods have potential for being utilised to monitor transcription in the future. The hybrid-DNA construct would also allow to investigate the effect of other factors on transcription quite easily. This could involve regulatory factors (such as Rad26, TFIIS, RPA and elongation factors) or factors which bind to DNA and cause roadblocks (Cas9 or other polymerases). Furthermore, a new NTS could also be annealed into the generated gap to produce a realistic bubble or introduce mismatches/damage into the DNA template that could act as roadblocks to transcription, to better understand this effect and compliment the smFRET assay developed. This would be harder to introduce in the ligation method as inserting a specific site of damage into a 6-10 kb region is complex. However, it would be the better method if nucleosomes wanted to be introduced intro the construct. An extra region of DNA could be ligated in which contains a pre-assembled nucleosome or a handle that contains the Widom 601 sequence that would allow nucleosomes to specifically bind to that region once already assembled.

Although there have been several papers published using the ligation based method, they ligate the protein directly to one bead and measure transcription via force based measurements(Chen et al., 2019; Dangkulwanich et al., 2013). The hybrid-DNA methods is very new but has already been shown to be useful in probing other DNA processing enzymes (Belan et al., 2021; Anand et al., 2021). This would suggest both of these approaches are viable and therefore would require some more troubleshooting to optimise it but transcription should be able to be visualised with the above modifications made. Due to time restrictions this was not able to be carried out by the submission date but could be carried out utilising the work already done in this thesis.

Bibliography

R. Anand, E. Buechelmaier, O. Belan, M. Newton, A. Vancevska, A. Kaczmarczyk, T. Takaki, D. S. Rueda, S. N. Powell, and S. J. Boulton. HELQ is a dual-function DSB repair enzyme modulated by RPA and RAD51. *Nature 2021 601:7892*, 601 (7892):268–273, 12 2021. ISSN 1476-4687. doi: 10.1038/s41586-021-04261-0. URL `https://www.nature.com/articles/s41586-021-04261-0`.

O. Belan, G. Moore, A. Kaczmarczyk, M. D. Newton, R. Anand, S. J. Boulton, and D. S. Rueda. Generation of versatile ss-dsDNA hybrid substrates for single-molecule analysis. *STAR Protocols*, 2(2):100588, 6 2021. ISSN 2666-1667. doi: 10.1016/J.XPRO.2021.100588.

Z. Chen, R. Gabizon, A. I. Brown, A. Lee, A. Song, C. Díaz-Celis, C. D. Kaplan, E. F. Koslover, T. Yao, and C. Bustamante. High-resolution and high-accuracy topographic and transcriptional maps of the nucleosome barrier. *eLife*, 8:1–38, 2019. ISSN 2050-084X. doi: 10.7554/elife.48281.

A. Costa, I. V. Hood, and J. M. Berger. Mechanisms for Initiating Cellular DNA Replication. *http://dx.doi.org/10.1146/annurev-biochem-052610-094414*, 82:25–54, 6 2013. ISSN 00664154. doi: 10.1146/ANNUREV-BIOCHEM-052610-094414. URL `https://www.annualreviews.org/doi/abs/10.1146/annurev-biochem-052610-094414`.

M. Dangkulwanich, T. Ishibashi, S. Liu, M. L. Kireeva, L. Lubkowska, M. Kashlev, and C. J. Bustamante. Complete dissection of transcription elongation reveals slow translocation of RNA polymerase II in a linear ratchet mechanism. *eLife*, 2013(2):1–22, 2013. ISSN 2050084X. doi: 10.7554/eLife.00971.

Y. A. G. Fosado, D. Michieletto, C. A. Brackley, and D. Marenduzzo. Nonequilibrium dynamics and action at a distance in transcriptionally driven DNA supercoiling. *Proceedings of the National Academy of Sciences*, 118(10):1–6, 2021. doi: 10.1073/pnas.1905215118/-/DCSupplemental.y.

J. Gál and M. Kálmán. Autosticky PCR. *Methods in molecular biology (Clifton, N.J.)*, 192:141–151, 2002. ISSN 10643745. doi: 10.1385/1-59259-177-9:141. URL `https://link.springer.com/protocol/10.1385/1-59259-177-9:141`.

D. J. Hobson, W. Wei, L. M. Steinmetz, and J. Q. Svejstrup. RNA Polymerase II Collision Interrupts Convergent Transcription. *Molecular Cell*, 48(3):365–374, 2012. ISSN 10972765. doi: 10.1016/j.molcel.2012.08.027. URL http://dx.doi.org/10.1016/j.molcel.2012.08.027.

G. A. King, F. Burla, E. J. Peterman, and G. J. Wuite. Supercoiling DNA optically. *Proceedings of the National Academy of Sciences of the United States of America*, 116(52):26534–26539, 12 2019. ISSN 10916490. doi: 10.1073/PNAS.1908826116/-/DCSUPPLEMENTAL. URL /pmc/articles/PMC6936703//pmc/articles/PMC6936703/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC6936703/.

J. R. Lakowicz. *Principles of fluorescence spectroscopy*. Springer, 2006. ISBN 0387312781. doi: 10.1007/978-0-387-46312-4.

M. Losito, Q. M. Smith, M. D. Newton, M. E. Cuomo, and D. S. Rueda. Cas12a target search and cleavage on force-stretched DNA. *Physical Chemistry Chemical Physics*, 23(47):26640–26644, 12 2021. ISSN 1463-9084. doi: 10.1039/D1CP03408A. URL https://pubs.rsc.org/en/content/articlehtml/2021/cp/d1cp03408ahttps://pubs.rsc.org/en/content/articlelanding/2021/cp/d1cp03408a.

E. Nogales, R. K. Louder, and Y. He. Structural Insights into the Eukaryotic Transcription Initiation Machinery. *Annual Review of Biophysics*, 46(1):59–83, 2017. ISSN 1936-122X. doi: 10.1146/annurev-biophys-070816-033751. URL http://www.annualreviews.org/doi/10.1146/annurev-biophys-070816-033751.

M. R. Pan, M. C. Hsu, L. T. Chen, and W. C. Hung. Orchestration of H3K27 methylation: mechanisms and therapeutic implication. *Cellular and Molecular Life Sciences*, 75(2):209, 1 2018. ISSN 14209071. doi: 10.1007/S00018-017-2596-8. URL /pmc/articles/PMC5756243//pmc/articles/PMC5756243/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC5756243/.

M. Scandaglia and A. Barco. Contribution of spurious transcription to intellectual disability disorders. *Journal of Medical Genetics*, 56(8):491–498, 8 2019. ISSN 0022-2593. doi: 10.1136/JMEDGENET-2018-105668. URL https://jmg.bmj.com/content/56/8/491https://jmg.bmj.com/content/56/8/491.abstract.

P. Sen, G. Donahue, C. Li, Y. Lan, G. Egervari, N. Robertson, P. P. Shah, E. Kerkhoven, D. C. Schultz, P. D. Adams, and S. L. Berger. Spurious intragenic transcription is a hallmark of mammalian cellular senescence and tissue aging. *bioRxiv*, page 2022.05.20.492816, 5 2022. doi: 10.1101/2022.05.20.492816. URL `https://www.biorxiv.org/content/10.1101/2022.05.20.492816v1https://www.biorxiv.org/content/10.1101/2022.05.20.492816v1.abstract`.

H. Ukai, M. Ukai-Tadenuma, T. Ogiu, and H. Tsuji. A new technique to prevent self-ligation of DNA. *Journal of Biotechnology*, 97(3):233–242, 8 2002. ISSN 0168-1656. doi: 10.1016/S0168-1656(02)00107-4.

J. T. Wade and D. C. Grainger. Spurious transcription and its impact on cell function. *Transcription*, 9(3):182, 5 2018. ISSN 21541272. doi: 10.1080/21541264.2017.1381794. URL `/pmc/articles/PMC5927700//pmc/articles/PMC5927700/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC5927700/`.

M. D. Wang, H. Yin, R. Landick, J. Gelles, and S. M. Block. Stretching DNA with optical tweezers. *Biophysical Journal*, 72(3):1335–1346, 1997. ISSN 00063495. doi: 10.1016/S0006-3495(97)78780-0.

Chapter 7

Conclusion

7.1 Thesis Discussion

Eukaryotic transcription is a highly complex process and must efficiently and accurately transcribe protein coding genes. This is made even more complex due to the fact that DNA is prone to a high amount of damage causing transcriptional stress. In this thesis I have developed new single-molecule assays that have allowed me to probe the dynamics of eukaryotic RNAPII transcription elongation at such damages. In Chapter 4, I developed a new single-molecule FRET based assay that required no modification to the RNA polymerase while still achieving high resolution visualisation of RNAPII dynamics along a DNA template. Furthermore, I illustrated the effect the DNA sequence can have on this process, where AT-rich DNA induced transcriptional pauses and premature termination. This work furthers our ability to monitor such complex processes while maintaining the ability to use higher eukaryotic samples, such as mammalian RNAPII.

This assay was then further utilised in Chapter 5, to monitor the effect DNA damage has on transcription elongation. Previous biochemical and structural data has shown that RNAPII stalls at damages including abasic sites, 8-Oxo Guanine lesions and CPDs, but has not highlighted exactly how they effect RNAPII and if they all cause the same stalling effect. The results presented here clearly show that different damages induce heterogeneous mechanistic outcomes of RNAPII dynamics. While abasic lesions induced complete stalling of RNAPII, 8-Oxo Guanine and CPDs led to a more dynamic behaviour in RNAPII, which exhibited different levels of backtracking. It was also shown that these dynamics were not NTP dependent and were in fact due to different intrinsic properties of RNAPII.

These single-molecule results can be explained in more detail due to the structural information available, where each damage also induces changes in the RNAPII elongation complex. My results suggest the abasic lesion is the most severe damage tested, inducing complete stalling and no backtracking. This is validated by the biochemical and structural data available, which shows no bypass of this lesion, and severe structural changes in the RNAPII active site (Wang et al., 2018). The lack of base-pairing and stacking causes the DNA and RNA to block forward translocation, and based on this single-molecule data may also explain why long-range backwards translocation is also hindered. CPDs showed the next biggest effect on RNAPII where although it was more dynamic than the abasic lesion (50% vs 10%), these dynamics were predominantly found in the short range transitions. These short range dynamics are likely due to the CPD also blocking the polymerase forward motion past the damage, but allowing for mild fluctuations in the backwards motion as the main impediment here is insertion opposite the CPD, not the damage itself (Brueckner et al., 2007). This was in contrast to the transitions observed when encountering an 8-Oxo Guanine lesion, where both short-distance and long-distance events were observed. This high amount of dynamics is likely due to the 8-Oxo Guanine group causing almost no structural changes within the RNAPII active site and so only induces pausing and fast backtracking unlike the other damages (Oh et al., 2020).

It is also important to consider how these damages are treated within the cell, which might add further insight into how RNAPII interacts with them in a heterogeneous way. Small lesions like abasic sites and 8-Oxo Guanine are dealt with by the BER pathway, while larger damages such as CPDs use the TC-NER pathway. Both of these have different initiation events, where in BER the damage signals for the repair while in TC-NER, the stalled polymerase is thought to be the main initiation signal. To test the effect NER

factors have on stalled RNAPII, I added Rad26, a CSB ortholog, and saw a change in the FRET dynamics when binding to RNAPII stalled at a CPD and 8-Oxo Guanine lesion. Now, RNAPII could not backtrack to the high FRET position that was observed in its absence, instead limiting backtracking to 4-5 nt. Furthermore, lower FRET states were observed in the presence of Rad26, suggesting RNAPII has been 'pushed' further along the template over the damage itself. This is likely due to the forward translocation effect Rad26 has been hypothesised to posses (Xu et al., 2017). This is functionally important as Rad26 reduces the dynamic behaviour of naturally stalled RNAPII and holds it on top of the damage site, a vital first step in initiating TC-NER (Brueckner et al., 2007). These change in dynamics were also seen when RNAPII interacted with the 8-Oxo Guanine lesion, where backtracking to the high FRET state was no longer detected. While I did not see an increase in bypass efficiency when encountering a CPD, increase in bypass of 8-Oxo Guanine lesions were observed highlighting the effect Rad26 can have on stalled RNAPII. While Rad26 is not a component of the BER pathway, recent evidence has suggested that NER factors may also be involved in stalled RNAPII at smaller oxidative lesions, and so these results are still of importance to better understand the link between stalled RNAPII and damage repair factors (Kumar et al., 2022).

The final chapter of my thesis, Chapter 6, dealt with monitoring transcription using optical tweezers, currently more commonly used to study RNAPII elongation. This technique is powerful and recently has been shown to identify stalled RNAPII with high resolution, but again required the modification of RNAPII. To be able to utilise this powerful technique I combined it with fluorescence and generated two major constructs, one using a ligation based method which has been utilised in the past and a second novel assay that allows the visualisation of RNAPII on both ds- and ss-DNA simultaneously (Belan

et al., 2021). The main benefit of the second novel assay is that it would allow monitoring of transcription at DNA junctions, ssDNA gaps and forks that are often encountered during transcription. Both of these assays would be suitable to study transcription as the use of longer DNA constructs allows the use of multiple other factors, such as repair factors, to be added and monitored alongside RNAPII transcription, something that is more difficult to achieve using smFRET TIRF based assays. Although I was not able to directly visualise transcription with these constructs within the time I had available to me, I was able to see RNA and RNAPII binding in both instances. Furthermore, I fond that RNAPII will bind to the ss-DNA regions without the need for an RNA primer or other transcription factors as previously thought for eukaryotic RNAPII, while no binding to dsDNA at all was visualised. An increase in the force exacerbated this effect, where new RNAPII events were detected as bubbles within the DNA were formed. This may aid in elucidating how RNAPII binds to the eukaryotic genome, where it is unable to bind to dsDNA without the presence of its regulatory factors preventing spurious transcription. But at single-stranded regions (such as forks, nicks and bubbles) RNAPII would be able to bind and possibly disrupt further processes, highlighting the need for extra regulation within the cell.

For all the experiments carried out in this thesis it is important to address the differences between the *in vitro* situation I use here and the *in vivo* situation where RNAPII would be naturally found. As this is a simplified construct to observe the underlying RNAPII mechanism there are clearly some salient differences that will affect how RNAPII will perform *in vivo* to our system. Firstly, as discussed earlier the phosphorylation state of RNAPII in our system is not changeable and will contain very few post-translational modifications. In the *in vivo* setting, RNAPII phosphorylation is very important for

the regulation of transcription elongation, where S2 phosphorylation often encodes for elongating RNAPII (Ahn and al., 2004). As this system does not contain any kinases, no further phosphorylation is possible. Secondly, RNAPII will usually be found interacting to a multitude of both negative and positive elongation factors that help to regulate transcription speed and processivity. Again, in our simplified setting we do not have such proteins present, this in turn could emphasise any pause states that are observed in the FRET assay. While they may also be happening *in vivo* it is likely that the presence of elongation factors will reduce both the time and frequency of such pauses. Finally, the NTP concentration used in the FRET assay is much lower than the concentration found within the cell. The normal intracellular [NTP] within the nucleus is in the mM range, while we use low µM range in our experiment in order to slow down the normal transcription process. In turn we are likely exaggerating the pause states during the transcription pathway that are observed in the FRET assay. As the [NTP] is much higher *in vivo* these pause states would not likely impact RNAPII as we observe. However, it should be added that the intracellular concentration of proteins that use these NTPs is also much higher than in our setting and so while the absolute concentration is higher than our system the amount of free NTPs for RNAPII elongation might be lower than these levels. Furthermore, the [ATP] is much higher than for other NTPs present as this is used as an energy source also within the cell. We do not account for this in our system, although *in vivo* it may account for a favouring of ATP incorporation and may be a driving force for why ATP is preferentially inserted opposite an abasic lesion.

While these differences and more make our system quite divergent from the real-life *in vivo* system, it can still provide useful information as to the underlying mechanism of how RNAPII behaves in a minimal way. In order to understand how TFs, phosphorylation

state, altered [NTP] affect transcription and more specifically pausing it is first important to understand the underlying mechanism of RNAPII. Once this has been probed, which this assay is a step towards, we can then add complexity such as the addition of TF or phosphorylation state to determine specifically their effect on transcription. This is something that could be carried out next using this assay.

7.2 Future Work

The development of the smFRET assay laid out here has allowed for a better understanding of the effect DNA damages have on RNA Polymerase II. One of the benefits of such an assay is the ability to change the roadblock in question with relative ease, whether that be another form of damage such as single-stranded breaks, epigentic marks such as DNA methylation or a nucleosome, which as discussed in Chapter 1, is a common roadblock RNAPII must deal with. Moving the damage from the template strand to the non-template strand could also be of interest, as the results from Chapter 4 highlight that the NTS may play a bigger part in RNAPII pausing then previously thought. Another roadblock that would be interesting to investigate would be the presence of other cellular proteins that are present along DNA during transcription. This could be members of the replisome or CRISPR/Cas proteins such as Cas9 which have been shown to be affected by supercoiled DNA states, often imposed by RNA polymerases (Newton et al., 2019). Addition of these proteins downstream of the RNAPII binding site would allow us to determine how RNAPII stalls, backtracks or bypasses such proteins, where in some cases it has been shown to evict the other protein while in others cause complete stalling. Due to the fact that RNAPII does not need to be modified, the development of this assay also opens it up to be used with higher eukaryotic RNAPII samples such as from mammalian

tissue. This could be important as mammalian RNAPII (from calf or bovine sources) is closer genetically to human RNAPII and could respond differently in regards to its kinetics hence yielding greater insight into this process.

Although this work has provided great insight into RNAPII dynamics, the eukaryotic DNA damage response system is highly complex involving multiple repair proteins. As discussed, our understanding of this process is still expanding with new interactions between stalled polymerases and repair factors being hypothesised, and single-molecule assays would allow a deeper understanding of such interactions. The next step of this project would be to add other repair factors to this assay, as was done with Rad26, to see how they interact with stalled RNAPII and in combination with Rad26. For instance, TFIIH has been hypothesised to induce backtracking of RNAPII at bulky lesions, counteracting the effect of CSB (Rad26 homolog). Addition of this factor may induce the return of the dynamics that were observed without the addition of Rad26 and help to better understand its mechanism and allow for better targeting of this pathway to reduce the prevalence of genetic diseases.

This investigation could be aided even more by the use of the optical tweezer assay laid out in Chapter 6. The next step for these assays would be to finalise optimisation so RNAPII transcription can be visualised fully with an optical tweezers and fluorescence set-up, once achieved addition of these repair factors will provide better understanding of how TC-NER factors interact with stalled RNAPII. For example, the bacterial homolog of Rad26 (Mfd) has been shown to exhibit a diffusive property on dsDNA where it translocates behind the RNAPII and actively pushes it forward. This is not something that has yet to be probed fully with eukaryotic proteins and using such an assay would allow probing of this mechanism even further.

Bibliography

S. Ahn and e. al. Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing. *Mol. Cell*, 13:67–76, 2004.

O. Belan, G. Moore, A. Kaczmarczyk, M. D. Newton, R. Anand, S. J. Boulton, and D. S. Rueda. Generation of versatile ss-dsDNA hybrid substrates for single-molecule analysis. *STAR Protocols*, 2(2):100588, 6 2021. ISSN 2666-1667. doi: 10.1016/J.XPRO.2021.100588.

F. Brueckner, U. Hennecke, T. Carell, and P. Cramer. CPD Damage Recognition by Transcribing RNA Polymerase II. *Science*, 315(February):859–862, 2007.

N. Kumar, A. F. Theil, V. Roginskaya, Y. Ali, M. Calderon, S. C. Watkins, R. P. Barnes, P. L. Opresko, A. Pines, H. Lans, W. Vermeulen, and B. Van Houten. Global and transcription-coupled repair of 8-oxoG is initiated by nucleotide excision repair proteins. *Nature Communications 2022 13:1*, 13(1):1–16, 2 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-28642-9. URL `https://www.nature.com/articles/s41467-022-28642-9`.

M. D. Newton, B. J. Taylor, R. P. Driessen, L. Roos, N. Cvetesic, S. Allyjaun, B. Lenhard, M. E. Cuomo, and D. S. Rueda. DNA stretching induces Cas9 off-target activity. *Nature Structural & Molecular Biology 2019 26:3*, 26(3):185–192, 2 2019. ISSN 1545-9985. doi: 10.1038/s41594-019-0188-z. URL `https://www.nature.com/articles/s41594-019-0188-z`.

J. Oh, A. M. Fleming, J. Xu, J. Chong, C. J. Burrows, and D. Wang. RNA polymerase II stalls on oxidative DNA damage via a torsion-latch mechanism involving lone pair–π and CH–π interactions. *Proceedings of the National Academy of Sciences of the United States of America*, 117(17):9338–9348, 2020. ISSN 10916490. doi: 10.1073/pnas.1919904117.

W. Wang, C. Walmacq, J. Chong, M. Kashlev, and D. Wang. Structural basis of transcriptional stalling and bypass of abasic DNA lesion by RNA polymerase II. *Proceedings of the National Academy of Sciences of the United States of America*, 115(11):E2538–E2545, 2018. ISSN 1091-6490. doi: 10.1073/pnas.

1722050115. URL http://www.ncbi.nlm.nih.gov/pubmed/29487211http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5856558.

J. Xu, I. Lahiri, W. Wang, A. Wier, M. A. Cianfrocco, J. Chong, A. A. Hare, P. B. Dervan, F. DiMaio, A. E. Leschziner, and D. Wang. Structural basis for the initiation of eukaryotic transcription-coupled DNA repair. *Nature*, 551(7682):653–657, 2017. ISSN 14764687. doi: 10.1038/nature24658.