A single molecule view of FEN1 remarkable substrate recognition,

perfect catalysis and regulation

Dissertation by

Manal S. Zaher

In Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy

King Abdullah University of Science and Technology, Thuwal,

Kingdom of Saudi Arabia

© May 2019

Manal S. Zaher

All rights reserved
The dissertation of Manal S. Zaher is approved by the examination committee.

Committee Chairperson: Prof. Samir M. Hamdan
Committee Members: Prof. Stefan T. Arold, Prof. Salim Al-Babili, Prof. Orlando Schärer (external examiner).
ABSTRACT

A single molecule view of FEN1 remarkable substrate recognition,
perfect catalysis and regulation

Manal S. Zaher

DNA replication is one of the most fundamental processes in all living organisms. Its semi-discontinuous nature dictates that the lagging strand is synthesized in short fragments called Okazaki fragments. In eukaryotes, each Okazaki fragment is initiated by an ~ 30-40 nucleotide-long RNA-DNA hybrid primer that is synthesized by Pol α-primase complex. To ensure genomic stability, the RNA primer has to be excised, any misincorporations by Pol α have to be corrected for and finally the resulting nick has to be sealed generating a contiguous strand. This feat is accomplished by a highly coordinated and regulated process called Okazaki fragment maturation. At the center of this process are 5’ nucleases, which are structure-specific nucleases that catalyze the incision of phosphodiester bonds one nucleotide into the 5’ end of ssDNA/dsDNA junctions.

Previous structural and biochemical studies have shed some light on the mechanism of FEN1 substrate recognition, its catalysis and regulation. However, many gaps in our understanding of this remarkable nuclease still persist. Moreover, the choice between the short- and long-flap pathways is still elusive. Finally, the mechanism of the coordination among the different enzymatic activities of the polymerase, the nuclease and the ligase during Okazaki fragment maturation is still debatable. In this work, we set out to study FEN1 substrate recognition, catalysis and regulation using single molecule techniques. We
show that FEN1 employs a sophisticated substrate recognition mechanism through which it actively distorts the DNA to ~100° bent angle. It also displays a remarkable selectivity towards its cognate substrate and avoids off-target substrate by a lock-down mechanism that commits the enzyme for catalysis on cognate substrates while promoting the dissociation of non-cognate substrates. We further characterized FEN1 reaction from substrate binding/bending to product handoff and built a comprehensive kinetic scheme that shows FEN1 releasing its product in two steps. Finally, we uncovered an unprecedented role of FEN1 in the choice between short- and long-flap pathways.
ACKNOWLEDGEMENTS

This dissertation presents the polished culmination of my research findings as a PhD candidate but it cannot in any way describe the journey of my scientific and personal growth that led me here. Allow me here to express my personal gratitude to those who supported and inspired me throughout the good, the bad and the ugly turns of this journey.

Firstly, I would like to express my heartfelt appreciation and gratefulness to my advisor Prof. Samir M. Hamdan for the opportunity to be part of his research group. This work would not have been possible without his support, guidance and inspiration, both on an academic and personal level. Secondly, to my fellow PhD student and best friend, Fahad Rashid, I cannot thank you enough for all your support, encouragement and motivation. You have been my solid rock through the ups and downs.

I would like to also extend my heartfelt appreciation to each and every one in Hamdan’s Lab group for being always there to encourage and challenge me to sharpen my scientific knowledge, to engage me in intellectual discussions or to simply brighten up my day with a smile. To our past members, Alia Al-Kilani, Mohamed Elshenawy and Luay I. Joudeh, you are not forgotten and you have been missed dearly.

My sincere gratitude goes to my collaborators here in KAUST, Prof. Satoshi Habuchi and his PhD student Paul D. Harris, and at Wesleyan University, Prof. Manju M. Hingorani and her PhD student Bo Song for their expert advice and contribution to the research presented in this dissertation.

I would also like to offer my warmest regards and heartfelt thanks to my friends Hanan Mahmood, Dina Abou Samra and Tahani Assi for their constant support and endless
patience especially when things went south. You ladies have provided me with a home far away from home.

On the note of home, I cannot express enough how grateful and appreciative I am to my family members, my parents, siblings, nephews and nieces. You have all in all different ways contributed to the person I am now. Especially to my mother and sister, I am forever indebted to you for your unconditional love and support and for always believing in me.
# TABLE OF CONTENTS

EXAMINATION COMMITTEE PAGE .................................................................................. 2  
ABSTRACT ..................................................................................................................... 3  
ACKNOWLEDGEMENTS ............................................................................................... 5  
TABLE OF CONTENTS ................................................................................................. 7  
LIST OF ABBREVIATIONS ............................................................................................ 11  
LIST OF ILLUSTRATIONS ............................................................................................. 14  
LIST OF TABLES ........................................................................................................... 16  

Chapter 1 ....................................................................................................................... 17  

1. DNA replication: how far have we come? ............................................................. 17  
   1.1 Preface ................................................................................................................. 17  
   1.2 Initiation ............................................................................................................... 19  
      1.2.1 Eukaryotic origins of replication: where does it all start? ....................... 19  
      1.2.2 Mechanisms of replication initiation ......................................................... 21  
      1.2.3 Spatio-temporal organization of replication initiation ......................... 25  
   1.3 Elongation ............................................................................................................ 27  
      1.3.1 Helicase unwinding and the RPC assembly ........................................... 29  
      1.3.2 DNA synthesis ......................................................................................... 34  
      1.3.3 Okazaki fragment maturation ................................................................. 43  
   1.4 Termination .......................................................................................................... 53  
      1.4.1 Where and when does it end? .................................................................. 54  
      1.4.2 Eukaryotic replication termination: what we know so far .................... 56  
   1.5 Looking ahead: Motivation of the work ............................................................. 61  

Chapter 2 ....................................................................................................................... 65  

2. Materials and Methods* ......................................................................................... 65  
   2.1 DNA substrates ................................................................................................... 65  
   2.2 Proteins expression and purification ................................................................. 66  
      2.2.1 Human FEN1 ............................................................................................ 66  
      2.2.2 Human RPA ............................................................................................. 67  
      2.2.3 Human PCNA ........................................................................................... 69  
      2.2.4 Human RFC ............................................................................................. 70
2.2.5 Human DNA Ligase 1 .......................................................... 71
2.3 TIRF-based smFRET.................................................................. 73
  2.3.1 FEN1 DNA bending assays .................................................. 77
  2.3.2 FEN1 DNA bending in the presence of PCNA and/or LIG1........... 78
  2.3.3 FEN1 single molecule cleavage assays................................. 78
2.4 Confocal-based smFRET ............................................................. 82
  2.4.1 Solution-phase smFRET experiments .................................... 83
  2.4.2 Surface-immobilized smFRET experiments ......................... 84
  2.4.3 Data processing .................................................................. 85
2.5 Bulk and single molecule PIFE experiments ............................... 86
2.6 Ensemble single turnover and steady state cleavage assays ........... 87
2.7 Steady state bulk cleavage in the presence of RPA ...................... 89
2.8 Surface Plasmon Resonance (SPR) binding ............................... 90
2.9 PCNA loading on nicked DNA via ΔN-RFC ................................. 90
  2.9.1 Bulk FRET assay ............................................................... 90
  2.9.2 smFRET assay .................................................................. 91
Chapter 3 ....................................................................................... 93
  3. Single-molecule FRET unveils induced-fit mechanism for substrate selectivity in flap endonuclease 1* .................................................. 93
  3.1 Abstract ............................................................................... 93
  3.2 Introduction .......................................................................... 94
    3.2.1 FEN1: a multi-functional small protein .............................. 95
    3.2.2 FEN1 cognate substrate .................................................... 96
    3.2.3 Substrate recognition ....................................................... 98
    3.2.4 FEN1 crystal structure ..................................................... 101
    3.2.5 FEN1 state of affairs ....................................................... 103
  3.3 Results ............................................................................... 104
    3.3.1 FEN1 actively bends the DNA ......................................... 104
    3.3.2 FEN1 never misses cleavage of its correct substrate ............ 108
    3.3.3 A comprehensive kinetic scheme of FEN1 cleavage reaction ... 112
    3.3.4 FEN1 verifies its substrate to avoid off-target DNA cleavage in the DNA lockdown step ..................................................... 122
Chapter 4 ....................................................................................... 129
4. Missed cleavage opportunities by FEN1 lead to Okazaki fragment maturation via the long-flap pathway* .................................................................................................................. 129
4.1 Abstract .................................................................................................................. 129
4.2 Introduction ........................................................................................................... 130
4.3 Results ................................................................................................................. 138
  4.3.1 Effect of 5’ flap length on the DNA bending activity of FEN1 ....................... 139
  4.3.2 Effect of Mg$^{2+}$ on DF substrates ................................................................. 141
  4.3.3 Effect of 5’ flap length on FEN1 kinetic mechanism ...................................... 144
  4.3.4 Coordination between RPA and FEN1 on short- and long-flap substrates .... 152
Chapter 5 ................................................................................................................. 159
A possible role for DNA-mediated and PCNA-coordinated product handoff between FEN1 and DNA Ligase 1* ................................................................................................................. 159
5.1 Abstract .............................................................................................................. 159
5.2 Introduction ......................................................................................................... 160
5.3 Results .................................................................................................................. 168
  5.3.1 PCNA stabilizes FEN1 bending of its product ................................................. 168
  5.3.2 LIG1 competes with FEN1 to the nicked product .......................................... 172
  5.3.3 PCNA-mediated product handoff ................................................................. 175
  5.3.4 PCNA loading via RFC in bulk and single molecule ..................................... 177
Chapter 6 ................................................................................................................. 183
6. Discussion and Future Directions* ....................................................................... 183
  6.1 Uncovering FEN1 mechanistic details ............................................................... 184
    6.1.1 FEN1 remarkable substrate recognition: an active multi-step vetting process .......................................................................................................................... 185
    6.1.2 A comprehensive kinetic scheme of FEN1 reaction .................................. 188
    6.1.3 Where we heading ....................................................................................... 190
  6.2 FEN1 in the context of Okazaki fragment maturation ....................................... 192
    6.2.1 FEN1 triggers long flap-pathway by missing cleavage of long flaps .......... 193
    6.2.2 FEN1 and RPA interplay on DF substrates .................................................. 194
    6.2.3. Follow-up projects .................................................................................... 195
  6.3 PCNA-mediated product handoff between FEN1 and LIG1 ............................ 196
    6.3.1 FEN1 acts as a product chaperone stabilized by PCNA .............................. 197
    6.3.2 LIG1 competes with FEN1 to the product .................................................. 198
6.3.3 PCNA seemingly favors FEN1 ................................................................. 199
6.3.4 Loading PCNA via RFC is a bottleneck: what comes next? ................. 200
6.4 Concluding Remarks .................................................................................. 202
BIBLIOGRAPHY ............................................................................................... 204
Statement of Author’s Contributions .............................................................. 244
ACCOMPLISHMENTS ...................................................................................... 245
# LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>dsDNA</td>
<td>double-stranded DNA</td>
</tr>
<tr>
<td>ssDNA</td>
<td>single-stranded DNA</td>
</tr>
<tr>
<td>E. coli</td>
<td><em>Escherichia coli</em></td>
</tr>
<tr>
<td>S. cerevisiae</td>
<td><em>Saccharomyces cerevisiae</em></td>
</tr>
<tr>
<td>S. pombe</td>
<td><em>Schizosaccharomyces pombe</em></td>
</tr>
<tr>
<td>ARS</td>
<td>autonomously replicating sequences</td>
</tr>
<tr>
<td>ORC</td>
<td>origin recognition complex</td>
</tr>
<tr>
<td>pre-RC</td>
<td>pre-replicative complex</td>
</tr>
<tr>
<td>MCM</td>
<td>minichromosome maintenance</td>
</tr>
<tr>
<td>CDK</td>
<td>cyclin dependent kinase</td>
</tr>
<tr>
<td>DDK</td>
<td>Dbf4-dependent kinase</td>
</tr>
<tr>
<td>pre-IC</td>
<td>pre-initiation complex</td>
</tr>
<tr>
<td>pre-LC</td>
<td>pre-loading complex</td>
</tr>
<tr>
<td>RPA</td>
<td>Replication Protein A</td>
</tr>
<tr>
<td>CMG</td>
<td>Cdc45-MCM2-7-GINS</td>
</tr>
<tr>
<td>SV40</td>
<td>simian virus 40</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>RPC</td>
<td>replisome progression complex</td>
</tr>
<tr>
<td>OB</td>
<td>oligonucleotide/oligosaccharide-binding</td>
</tr>
<tr>
<td>Ctf4</td>
<td>chromosome transmission fidelity 4</td>
</tr>
<tr>
<td>OF</td>
<td>Okazaki fragment</td>
</tr>
<tr>
<td>dNTPs</td>
<td>deoxynucleotides</td>
</tr>
<tr>
<td>PCNA</td>
<td>proliferating cell nuclear antigen</td>
</tr>
<tr>
<td>PIP</td>
<td>PCNA-interacting peptide</td>
</tr>
<tr>
<td>IDCL</td>
<td>interdomain connecting loop</td>
</tr>
<tr>
<td>RFC</td>
<td>replication factor C</td>
</tr>
<tr>
<td>FEN1</td>
<td>flap endonuclease 1</td>
</tr>
<tr>
<td>EXO</td>
<td>exonuclease 1</td>
</tr>
<tr>
<td>LIG1</td>
<td>human DNA ligase 1</td>
</tr>
<tr>
<td>bp</td>
<td>base pair</td>
</tr>
<tr>
<td>nt</td>
<td>nucleotide</td>
</tr>
<tr>
<td>LacO</td>
<td>lac operator</td>
</tr>
<tr>
<td>LacR</td>
<td>lac repressor</td>
</tr>
<tr>
<td>IPTG</td>
<td>Isopropyl β-D-1-thiogalactopyranoside</td>
</tr>
<tr>
<td>smFRET</td>
<td>single molecule Förster Resonance Energy Transfer</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>PAGE</td>
<td>polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>BME</td>
<td>beta-mercaptoethanol</td>
</tr>
<tr>
<td>PMSF</td>
<td>phenylmethane sulfonyl fluoride</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>DTT</td>
<td>Dithiothreitol</td>
</tr>
<tr>
<td>BSA</td>
<td>bovine albumin serum</td>
</tr>
<tr>
<td>2c-ALEX</td>
<td>two-color alternating excitation</td>
</tr>
<tr>
<td>CW</td>
<td>continuous wave</td>
</tr>
<tr>
<td>DF</td>
<td>double flap</td>
</tr>
<tr>
<td>SF</td>
<td>single flap</td>
</tr>
<tr>
<td>EQ</td>
<td>Equilibrated</td>
</tr>
<tr>
<td>Non-EQ</td>
<td>non-equilibrated</td>
</tr>
<tr>
<td>SPR</td>
<td>Surface Plasmon Resonance</td>
</tr>
<tr>
<td>TIFR</td>
<td>total internal reflection fluorescence</td>
</tr>
<tr>
<td>PIFE</td>
<td>protein-induced fluorescence enhancement</td>
</tr>
<tr>
<td>DBD</td>
<td>DNA-binding domain</td>
</tr>
<tr>
<td>AdD</td>
<td>adenylation domain</td>
</tr>
<tr>
<td>OBD</td>
<td>oligonucleotide/oligosaccharide-binding domain</td>
</tr>
</tbody>
</table>
LIST OF ILLUSTRATIONS

Figure 1.1. A schematic depicting the current model for MCM loading and activation. 24
Figure 1.2. A model of the eukaryotic DNA replication fork................................. 28
Figure 1.3. Cryo-EM structure of CMG helicase suggests an oil rig-like pump jack model for DNA unwinding................................................................. 31
Figure 1.4. The catalytic subunits of eukaryotic replicative polymerases adopt a right-hand fold. ........................................................................................................ 36
Figure 1.5. A model of the eukaryotic priming mechanism........................................ 39
Figure 1.6. Okazaki fragment maturation regulation................................................ 50
Figure 1.7. Model for eukaryotic DNA replication termination.................................. 58
Figure 2.1. Schematic illustrating the construction of the microfluidic flow cell. .... 73
Figure 2.2. Objective-based TIRF-FRET setup.......................................................... 76
Figure 3.1. FEN1 substrate recognition...................................................................... 98
Figure 3.2. FEN1 actively bends its DF substrate...................................................... 106
Figure 3.3. Cleavage of cognate substrate by FEN1.................................................. 111
Figure 3.4. Internal-labeling scheme infers on FEN1 product release....................... 114
Figure 3.5. Building a kinetic scheme of FEN1 cognate substrate recognition and catalysis........................................................................................................ 119
Figure 3.6. smPIFE cleavage confirms FEN1 product release from an unbent state. .... 122
Figure 3.7. FEN1 multi-step vetting process of its substrate...................................... 126
Figure 4.1. FEN1 actions in Okazaki fragment maturation........................................ 135
Figure 4.2. Effect of 5’ flap length on FEN1 DNA bending activity............................ 140
Figure 4.3. Effect of Mg2+ on the conformational changes of DF substrates.............. 142
Figure 4.4. Effect of 5’ flap length on FEN1 cleavage activity.................................... 146
Figure 4.5. FEN1 misses cleavage, as the flap grows longer...................................... 150
Figure 4.6. Coordination between FEN1 and RPA actions on short- and long-flap substrates........................................................................................................ 155
Figure 4.7. RPA does not affect FEN1 single or multiple turnover kinetics on short flaps. ........................................................................................................ 157
Figure 5.1. The two proposed models for PCNA coordination of Okazaki fragment maturation. ................................................................. 164
Figure 5.2. Crystal structures of human PCNA-FEN1 complex and human LIG1 bound to DNA .................................................................................................................. 167
Figure 5.3. PCNA stabilizes FEN1 bending of the nick and product .................. 171
Figure 5.4. LIG1 competes FEN1 off the product DNA. .......................... 174
Figure 5.5. FEN1-LIG1 competition to the product in presence of PCNA. .... 176
Figure 5.6. ΔN-RFC purification and PCNA loading in bulk. ..................... 178
Figure 5.7. PCNA loading via ΔN-RFC at single molecule. ........................ 181
Figure 6.1. Model for control of catalytic selectivity by the DNA mutual-induced fit mechanism in FEN1 ........................................................................................................... 188
Figure 6.2. Comprehensive kinetic scheme of FEN1 reaction. .................... 190
LIST OF TABLES

Table 2.1. Substrates used for single molecule experiments in Chapter 3. .................. 81
Table 2.2. Substrates used for single molecule experiments in Chapter 4. .................. 81
Table 2.3. Substrates used in Chapter 5. ................................................................. 82
Table 2.4. Substrates used for ensemble single and multiple turnover kinetics. ........... 88
Table 2.5. Substrates used for steady state bulk cleavage in presence of RPA. ............ 89
Table 4.1. Kinetic parameters of FEN1 substrate bending and catalytic efficiency on DNA versus RNA DF substrates. ................................................................. 152
Chapter 1

1. DNA replication: how far have we come?

1.1 Preface

The requirement of cell division and genomic stability renders deoxyribonucleic acid (DNA) replication as one of the most vital processes in living organisms across all domains of life. Replication of the whole genome with high fidelity requires the coordinated action of many proteins and protein complexes (1). Errors made during DNA replication can be detrimental to genomic stability and can lead to many disease states including cancer, hence the tight association between DNA replication and DNA repair pathways. Early studies of DNA replication started with prokaryotic systems (2) soon after the DNA double helix structure was published in 1953 (3). However, the enormous eukaryotic genome sizes, along with the different physiological states of higher eukaryotes and their subsequent differential requirements, eluded early on that the understanding of eukaryotic DNA replication would take years of immense collaborative research.

Eukaryotic DNA replication is a complex, highly regulated and spatio-temporally organized process. It is also coupled with transcription, chromatin state and cell-cycle progression. The complexity of eukaryotic DNA replication further stems from the requirement of balancing a faithful complete duplication of the whole genome while
ensuring that replication occurs once and only once per each cell cycle. DNA replication as one of the most fundamental processes of life has attracted the attention of many molecular biologists, geneticists, biochemists and structural biologists. Their work over more than 30 years have shaped our current understanding of eukaryotic DNA replication, but we are only scratching the surface of this very sophisticated fundamental process. This chapter is dedicated to a description of the major breakthroughs in DNA replication research and outlining some gaps in our knowledge.

DNA replication is a semi-conservative (4), semi-discontinuous (5,6) duplication of the DNA. Its semi-conservative nature dictates the unwinding of the duplex parental DNA into two strands where each provides a template for the synthesis of the complementary strand resulting in two duplex DNA molecules each with one parental and one newly synthesized strand. On the other hand, the anti-parallel nature of the DNA double helix and the strict 5’-3’ directionality of most DNA polymerases impose that DNA is replicated semi-discontinuously. DNA replication occurs in three distinct stages: initiation, elongation and termination. Whereas the early steps of initiation start as soon as the cell emerges from mitosis and continue through G1 phase until the start of S phase, both elongation and termination occur during S phase. These stages are described below highlighting their protein components, protein-protein and protein-DNA interactions, as well as their regulation.
1.2 Initiation

1.2.1 Eukaryotic origins of replication: where does it all start?

DNA replication initiates at chromosomal sites called origins where double-stranded DNA (dsDNA) melts into two single-stranded DNAs (ssDNA) to provide templates for the successive DNA synthesis (7). It was evident early on that whereas prokaryotes mostly utilize a single origin of replication to initiate DNA replication on their relatively small circular chromosomes, eukaryotes require the initiation of replication from a multitude of origins to completely duplicate their longer linear genomic chromosomes within the cell cycle timing frame (8). However, one unifying feature for both prokaryotes and eukaryotes is that once replication is initiated, the fork progresses bidirectionally. The characterization of these origins is still elusive. While most prokaryotes initiate replication from specific DNA sequences such as OriC in Escherichia coli (E. coli) (9), the case is far more complex for eukaryotes. Early studies on budding yeast Saccharomyces cerevisiae (S. cerevisiae) showed that replication initiates at autonomously replicating sequences (ARSs) comprising an 11 bp consensus core sequence (10-14). It is worthy to note that although these sequence-specific sites in budding yeast mark potential initiation sites and favor the distribution of origins to ARSs, replication can putatively initiate from any DNA sequence (15-18). In the case of fission yeast Schizosaccharomyces pombe (S. pombe), the origins do not appear to possess consensus sequence but rather to have a rich AT content (19,20).
In higher eukaryotes, including mammalian cells, origins lack consensus sequences (21-24). However, mammalian origins have been shown, mostly by genome-wide studies, to exhibit some common sequence characteristics (25). These sequence characteristics include CpG island promoters, G-quadruplexes, the G-rich motifs called origin G-rich repeated elements, strand asymmetry, transcription start sites, and DNase-hypersensitive regions (26-31). Moreover, these origins generally correlate with a number of structural features on the DNA. Such structural features include transcription machinery, histone modifications, chromatin-packaging features and other DNA-DNA or DNA-protein interactions (23,32-35). It seems that the spatial distribution of origins in eukaryotes can be explained by a stochastic model with preferences to ARS elements in \textit{S. cerevisiae} and AT-rich regions in \textit{S. pombe} (15,36,37).

In conclusion, the choice of origin in eukaryotes seems to have mainly deviated from the high sequence-specificity adopted by prokaryotes. The reason for this deviation may very well reside in the requirement for the cell to adapt to different physiological states including, but not limited to, different differentiation states, different developmental stages and different tissue environments (15). These different physiological states confer various gene expression profiles and diverse composition and organization of the chromatin, which in turn could affect the availability of potential initiation sites due to their association with transcription and chromatin state during cell cycle progression. Hence, initiating replication at potentially any available DNA sequence could render eukaryotic cells adaptable to the greater complexity they face as compared to prokaryotic cells. This flexibility seems vital in the coordination of replication with transcription and chromatin de-/condensation as it makes it possible to initiate replication from different sites with
different cell types and with different cell cycles in response to the DNA packaging patterns (15).

1.2.2 Mechanisms of replication initiation

1.2.2.1 Origin licensing

The somewhat sequence-specificity of origins in budding yeast proved to be extremely beneficial for the initial identification of the major players of replication initiation and their mechanism of action (15,38-41). These factors were shown later to exist and be widely conserved among other eukaryotes (42). The first identified factor was the origin recognition complex (ORC), a hexameric complex with five subunits (Orc1-5) all related to AAA+ ATPase protein family (42-46). ATP binding to Orc1, the largest subunit, confers high affinity binding of ORC to S. cerevisiae ARS elements (15). These early studies in budding yeast and the succeeding studies in Xenopus egg extract helped shape our understanding of how replication is initiated. Initiation of replication commences by the binding of the ORC to the origin and the subsequent assembly of the pre-replicative complex (pre-RC) to properly load the core of the replicative helicase, that is the heterohexameric minichromosome maintenance (MCM2-7) with all of its subunits related to AAA+ ATPase superfamily (47-50). After ORC binds to the origin, it recruits Cdc6, an AAA+ protein, to the origin. Together, they indirectly recruit MCM2-7 through its binding partner Cdt1. This loading of MCM2-7 onto the origin is an ATP-hydrolysis dependent reaction (Figure 1.1).
The currently accepted model of MCM2-7 loading depicts the loading of two hexamers in two separate steps each dependent on two distinct sets of Cdc6 and Cdt1 molecules \((51, 52)\) (Figure 1.1). The two hexamers assemble on the origin in a head-to-head configuration where the dsDNA passes through the central channel of the hexamers \((53-55)\). The two hexamer rings are topologically linked to the DNA through the concerted action of ORC and Cdc6 in closing the gate between Mcm2 and Mcm5 subunits of each hexamer. The loaded double MCM2-7 hexamers in complex with ORC-Cdc6-Cdt1 form the pre-RC. However, the core replicative helicases are inactive at this stage. ORC binds to the origins emerging from mitosis and the pre-RC assembly occurs during G1 phase. The loading of the inactive replicative helicases during G1 phase is called replication licensing, and thus, Cdc6 and Cdt1 have been called licensing factors. Therefore, the loaded hexamers of MCM2-7 denote potential replication initiation sites and this licensing ensures that replication only occurs once per cell cycle and that it is tightly coupled to cell cycle progression \((15)\).

### 1.2.2.2 Helicase activation

Although the core of the replicative helicase is loaded in G1 phase, it only becomes activated in S phase in a highly regulated mechanism. This mechanism depends on the phosphorylation activity of two kinases, cyclin dependent kinase (CDK) and Dbf4-dependent kinase (DDK). Furthermore, it requires the association of a collection of protein factors, which constitutively associate with the replicative helicase and travel along the replication fork \((15)\). In budding yeast, ample research has identified an assembly of initiation protein factors that fall into two categories, the pre-initiation complex (Pre-IC)
(56) and the pre-loading complex (Pre-LC) (57). These factors include Cdc45, GINS heterotetramer complex (Sld5, Psf1, Psf2 and Psf3), Sld2, Sld3, Sld7, Dpb11, DNA polymerase ε (Pol ε), replication protein A (RPA) and MCM10 (8). The association of the Pre-IC and Pre-LC with the Pre-RC loaded onto the origin, along with the CDK and DDK phosphorylation activity, finally transform the inactive helicase into an active helicase complex composed of Cdc45, MCM2-7 and GINS (CMG complex) (58,59). This transformation is highly regulated and consists of two main steps: 1) MCM2-7 double hexamer splits into two single hexamers and 2) each MCM2-7 ring undergoes an opening-and-closing reconfiguration. The structure of CMG in the presence of ssDNA and non-hydrolysable ATP showed MCM2-7 AAA+ motor domains forming a cracked ring at the interface of MCM2 and MCM5 subunits (60). This structure showed CMG as a right-handed spiral with Cdc45 and GINS bridging the crack to form the closed topological toroid (60). This structure, along with CMG preference to ssDNA binding (61), supports a steric exclusion model where unwinding occurs by excluding the lagging strand from the central channel and the CMG encircles, and translocates along, the leading strand in 3’-5’ direction (61,62). Following the activation of the CMG helicase, other replication proteins are recruited to initiate DNA synthesis and form a complete replisome. The firing of the origin and consequently DNA unwinding where each CMG translocates further from the origin leaves ORC at the origin in what is termed as post-replicative complex (25) (Figure 1.1).
Figure 1.1. A schematic depicting the current model for MCM loading and activation. (A) During late M to G1 phase, the ORC recognizes and binds the origin. Double hexameric MCM2-7 are recruited to the ORC-bound origins via the concerted action of CDC6 and Cdt1 through a sequential assembly. This process results in licensed origins with loaded inactive core helicases and assembled pre-RC complexes. (B) As cells enter S phase, the loaded MCM2-7s are activated. This activation requires the association of GINS and Cdc45 with the MCM2-7 to form the CMG complex. Moreover, this process is highly regulated with the association of other protein factors, but most importantly by the phosphorylation with CDK and DDK kinases which are activated during S phase. To ensure that re-licensing does not occur, licensing and activating factors are regulated. For example, Cdt1 is regulated through a ubiquitin-mediated degradation as well as its specific inhibitor, geminin. CDC6 and ORC1 are also subject to regulation through CDK activity. Figure is adapted from (63).
1.2.3 Spatio-temporal organization of replication initiation

As mentioned earlier, there is high flexibility in the choice of what DNA site becomes a potential origin. The huge abundance of potential origins (8,34) ensures an expedited DNA synthesis of the large eukaryotic genomes. On the other hand, it posits a challenge for balancing the requirement that re-replication does not occur, especially considering their enormous number, with the requirement of complete genome duplication. Indeed, studying individual cells showed that origin choice is stochastic (36,64). Moreover, origin licensing does not guarantee the helicase activation, and consequently origin firing. In fact, only a subset of licensed origins fire while the rest stay dormant (27) presumably as a backup for stalled or collapsed replication forks. Genomic instability was observed with a marked reduction in licensed but dormant origins (22,26). The complexity is even further augmented by the timing of the firing. While some fire at the beginning of S phase, some fire mid or late S phase. Recent studies have shed some light towards understanding this spatio-temporal organization of replication initiation control (65,66).

To couple DNA replication with cell cycle progression, licensing is restricted to late mitosis and G1 phase. This restriction is imposed by the regulation of licensing factors expression and/or recycling and the activation/inactivation of mitotic-, G1- and S-CDKs (67-69). In late G1 phase, G1-CDKs are activated to prevent further licensing and the lack of S-CDK ensures that licensed origins do not fire at this stage (70). Once S-CDKs are activated at the beginning of S phase, origin firing can commence while the presence of
the inhibitory CDKs ensures that no new licensing occurs (8). This interplay between
CDKs activation/inactivation sets up a unique pattern for each cell cycle. Furthermore, the
chromatin state euchromatin versus heterochromatin influences the spatio-temporal
organization of origin firing. Most early firing origins are positioned within euchromatin
where chromatin is in a more-open state and firing factors can access the pre-RC
(34,71,72), whereas heterochromatin is less accessible and thus origins residing within
these regions tend to fire later (23,27,32). The delayed firing and replication of origins
residing in heterochromatin could be advantageous, if not needed, to maintain the integrity
of the nucleus. If heterochromatin replication were to co-occur with euchromatin
replication early in S-phase, then rapid and massive chromatin decondensation and re-
condensation is required, which in turn could overwhelm the cell and trigger DNA damage
response (73). This global control of origin firing is in competition with the local stochastic
origin choice; thus a ‘controlled stochastic’ model of origin choice has been proposed
(74,75).

Detailed understanding regarding the factors that regulate this choice of origins and
firing timing is still lacking. However, several regulators have been shown to influence
origin selection and firing time programme in budding yeast. These regulators include local
chromatin structure and epigenetic markers, chromosomal position of the origins, pre-RC
formation timing and maintenance, and the limiting firing timing factors recruitment
(reviewed in (8,23,35,76,77)). Moreover, the firing timing programme has been shown to
be influenced by transcription factors Fkh1 and Fkh2 which bias the early firing of a subset
of origins via ORC binding, clustering of early origins and association with Cdc45 (78). In
addition, there is a correlation between the presence of Rif1, a telomere-binding protein,
and replication-timing programme (79-82). When Rif1 is absent, the replication-timing programme is affected such that late origins at telomeres fire prematurely (83). In summary, replication initiation is highly regulated process to ensure that replication only occurs once per cell cycle, to set up the replication timing programme in accordance with cell cycle progression, and to ensure the proper loading of replicative helicases before elongation commences.

1.3 Elongation

Each origin firing, and the following DNA unwinding by CMG, creates two replication forks diverging away from the origin. At each replication fork, a set of proteins and protein complexes unwind the DNA into two parental strands, ensure ssDNA protection, synthesize the non-parental DNA strands and finally mature the synthesized DNA to generate two complete identical heritable dsDNA duplexes (Figure 1.2). This set of proteins and protein complexes is highly conserved among eukaryotes and collectively called the replisome (84). Much is known about the replisome components, their structure and function, owing to the wealth of information stemming from the early work with simian virus 40 (SV40) system (85) and the more recent in vitro reconstitutions of S. cerevisiae DNA replication on both naked and chromatin DNA templates (86,87). These recent reconstitutions confirmed the minimal set of proteins and stages required for eukaryotic DNA replication. Below is a summary of the current working model of the eukaryotic
replisome at the replication fork with brief description of each component structure and function and their spatially- and temporally-organized division of labor to achieve efficient and accurate DNA duplication. This elongation stage is centered around three themes; helicase unwinding and the assembly of the replisome progression complex (RPC), DNA synthesis, and maturation of the synthesized DNA (Figure 1.2).

**Figure 1.2. A model of the eukaryotic DNA replication fork.** The GMC replicative helicase composed of MCM2-7 motor protein, GINS and Cdc45 unwinds the parental DNA strands generating two ssDNA strands, which are protected by RPA. DNA synthesis on both strands is initiated by Pol α-primase complex, which generates an RNA/DNA hybrid primer though its dual polymerase activities. Pol α-primase complex is recruited to the CMG indirectly through its interactions with Ctf4 which bridges the eukaryotic helicase and primase activities. The primer is then handed off to the leading strand polymerase (Pol ε), which directly interacts with the CMG, and synthesis proceeds continuously. On the lagging strand, the primer is handed off to Pol δ, and the synthesis proceeds discontinuously in short Okazaki fragments (OFs). Both polymerases are tethered to the DNA through their interaction with the processivity-enhancing sliding clamp, PCNA, which is loaded on the DNA through its clamp loader, RFC. OF maturation is processed through the PCNA-coordinated action of Pol δ, flap endonuclease 1 (FEN1) and DNA Ligase 1. Pol δ strand displaces the previous OF primer creating a 5’ flap structure that is specifically recognized and cleaved by FEN1 to generate a nick that is finally sealed by DNA Ligase 1.
1.3.1 Helicase unwinding and the RPC assembly

As mentioned earlier, the CMG complex is composed of 11 subunits; 6 AAA+ ATPase subunits forming the MCM2-7 complex, 4 subunits comprising the GINS complex, and Cdc45. MCM2-7 is the catalytic core of the active CMG complex and the core of the inactive licensing complex. Crystal and EM structures from *Drosophila melanogaster* and *S. cerevisiae* have converged to the current understanding of the CMG assembly (60,88,89). The six subunits assemble to form a cracked ring with GINS and Cdc45 interacting mainly with subunits Mcm-2, Mcm-3 and Mcm-5. The MCM2-7 complex forms in two-tiers, the N- and C-terminal domains (Figure 1.3A). On one side, the NTD is comprised of the helical subdomains of all six subunits, a Zn-binding motif and oligonucleotide/oligosaccharide-binding (OB) motif, whereas on the other side, the CTD is assembled by the AAA+ motors of all subunits. The Cdc45 and GINS assemble mainly on the NTD side with Cdc45 interacting with the NTDs of Mcm2 and Mcm5 and the GINS complex interacting with the NTDs of Mcm3 and Mcm5 while both Cdc45 and GINS contacting each other (90) (Figure 1.3A).

1.3.1.1 CMG structure and its helicase mechanism

The cryo-EM structure of *S. cerevisiae* CMG (88) unveiled two major CMG conformers; an extended and a compact conformer. The extended conformer exhibits the CTD motor ring tilted by ~10° angle with respect to the NTD ring resulting in the motor AAA+ domains in the CTD ring being arranged in a spiral-like shape. On the other hand, the compact conformer shows the CMG to be in generally more compact state where the CTD
and NTD rings are more or less parallel to each other. Comparing the two conformers reveals that the NTD ring with Cdc45 and GINS seem to be a solid platform that is attached to the more flexible CTD ring containing the AAA+ motors (Figure 1.3B). The CTD motor ring presumably undergoes conformational changes between the extended and compact states with each ATP hydrolysis cycle (90). Taken together, a CMG mechanism resembling an oil rig pump jack attached to a stable platform was proposed (88). This pump jack movement between the two conformers allows the CMG to translocate along ssDNA 3’-5’ in a ratcheted inchworm motion while unwinding the dsDNA at the fork (Figure 1.3B). This mechanism is in sharp contrast to the sequential rotary ATP hydrolysis mechanism suggested to explain the unwinding of other homohexameric helicases including E. coli DnaB (91). While the sequential rotary ATP hydrolysis mechanism places equal importance to all the active ATPase sites, single point mutations of the ATPase sites in Drosophila (92) and yeast (93) Mcm2-7 subunits revealed nonequivalent roles of the different ATPase sites as only two of the six mutations cause major helicase defects.
Figure 1.3. Cryo-EM structure of CMG helicase suggests an oil rig-like pump jack model for DNA unwinding. (A) A top view of the cryo-EM structure of the CMG helicase highlighting all 11 subunits of the complex. MCM2-7 are shown in different shades of blue, GINS is shown in shades of green and Cdc45 is shown in fuchsia. The density map from cryo-EM is shown in semi-transparent surface rendering while the atomic model is superimposed and shown in cartoon (EMD: 6535 and PDB: 3JC5, respectively). The cryo-EM structure showed CMG in two conformers (extended and compact), only one shown here. (B) A schematic representing the oil rig-like pump jack model of CMG helicase mechanism of DNA unwinding. This is in contrast with the proposed sequential ATP hydrolysis mechanism of E. coli DnaB. Figure is adapted from (88).
1.3.1.2 CMG-associated proteins

Equally important to the CMG helicase unwinding activity of the parental DNA is the role it plays as a protein scaffold for other proteins and protein complexes. Some of them are essential for its activation such as Sld3/7, Sld2, Dpb11, Pol ε, DDK, CDK, and MCM10, while others are needed for the progression of the fork. Among these are the leading strand DNA polymerase, Pol ε (94), the initiator Pol α –primase complex (95) through Ctf4, the ssDNA-binding protein, RPA, and Topoisomerase I (Top1) (96). These proteins along with CMG form the RPC (90).

As the CMG helicase unwinds the parental DNA, intertwining of the parental strands ahead of the fork (positive supercoiling) accumulates which could pose a mechanical strain to the moving fork (97,98). This mechanical strain was proposed to be converted to intertwining of the daughter strands behind the fork as precatenates by a rotation at the fork (99). However, recent studies in budding yeast revealed that such rotation and precatenation are inhibited by Timeless/Tof1 and Tipin/Csm3 (100), two components of the replicosome. It was then proposed that topoisomerases of type I and II can relax the supercoiling generated during the elongation phase (98). This was further refined by assuming that only type I topoisomerases are needed to relax the positive supercoiling ahead of the fork which is in line with its association with the CMG (58), while the other type presumably functions behind the fork (101).

Nevertheless, the main outcome of the unwinding activity is the generation of two ssDNA strands which if left unprotected can be a hotspot for nucleases activity and
secondary structure formation. In all kingdoms, a ssDNA-binding protein is indispensable for DNA replication. The eukaryotic ssDNA-binding protein, RPA, is a heterotetramer complex (RPA70, RPA32 and RPA14) with multiple OB folds that assemble in a structure resembling a horseshoe (102) (Figure 1.2). RPA was the first cellular protein identified by the fractionation of the SV40 system (103-105), and since then, our knowledge of RPA protein interactions and posttranslational modifications has expanded to many cellular processes involving various DNA transactions such as DNA repair and recombination as well as the DNA damage response activation (104,106). Studies have shown that despite its reported high-affinity to ssDNA, its ssDNA binding is highly dynamic (107). This high dynamicity allows RPA to accommodate its versatile functionalities in different DNA transactions and sets up RPA as a first responder at ssDNA sites (104,106).

The recent discovery of a CMG-associating protein, namely chromosome transmission fidelity 4 (Ctf4) (58), has lastly shed some light on the long standing puzzle of how the helicase and primase activities in eukaryotes are coordinated, and consequently how leading and lagging strand synthesis are coordinated (95,108-110). In most prokaryotic systems, the helicase and primase activities are part of the same polypeptide or have direct physical interactions. Yet until recently, such association between the two eukaryotic activities had been missing. Ctf4, originally identified, as the name suggests, by searching for chromosome fidelity factors, is a homotrimer that self-assembles into a disk shape through its C-terminal domains (109). This homotrimer has been found to bind tightly to Sld5, one of GINS complex subunits, as well as the polymerase catalytic subunit of Pol α–primase complex (109) (Figure 1.2). It has been suggested that Ctf4-binding partners such as Sld5 and Pol α–primase complex contain a Ctf4-interacting peptide (CIP) box similar
to the PCNA-interacting peptide (PIP) box suggesting that Ctf4 can presumably act as a hub for different proteins (111). Other proteins have been shown to associate with Ctf4 including the helicase/nuclease Dna2, the rDNA-associated protein Tof2 and Chl1 helicase (111).

1.3.2 DNA synthesis

DNA synthesis is mainly accomplished by highly conserved enzymes, called DNA polymerases, which synthesize DNA in a restricted 5’-3’ directionality. Since DNA is an anti-parallel duplex, and the duplication of the two strands must be spatially and temporally coordinated, the polymerases’ restricted directionality poses a problem to the moving fork. To resolve this universal challenge, cells from all kingdoms have evolved a semi-discontinuous mechanism where one strand, the leading strand, is replicated continuously, and the other strand, the lagging strand, is replicated discontinuously (5,6) in short fragments of ~150-200 bases, called Okazaki fragments (OFs) (112). Most DNA polymerases necessitate a free 3’ hydroxyl group to instigate a nucleophilic attack leading to the phosphodiester bond formation, therefore, they cannot perform de novo DNA synthesis. DNA synthesis is thus initiated by a DNA-dependent RNA polymerase, called primase, which initiates the synthesis by polymerizing an RNA primer of species-specific unit length. This RNA primer provides the free 3’ hydroxyl group for further extension with deoxynucleotides (dNTPs) by the replicative polymerases.
1.3.2.1 Eukaryotic DNA polymerases: structure and division of labor

Eukaryotic replication employs primarily three different replicative polymerases; Pol α-primase complex, Pol δ and Pol ε in chronological order of their identification. With the exception of *S. cerevisiae* Pol δ, these three polymerases are heterotetramers composed of a catalytic polymerase subunit, a regulatory highly conserved B subunit and two additional accessory subunits. *S. cerevisiae* Pol δ is a heterotrimer with only one accessory subunit. The accessory subunits of the three replicative polymerases are diverse and are involved in different cellular functions (113). The catalytic subunits of the three eukaryotic polymerases host their DNA polymerase domains as well as 3’-5’ exonuclease domains (Figure 1.4A-C). This latter domain is active in both Pol δ and Pol ε but inactive in Pol α making it the least accurate among them. The polymerase core of the catalytic subunits exhibits the “right-hand” fold characterizing B family DNA polymerases with the classical palm, fingers, and thumb subdomains (114) (Figure 1.4A-C). Besides the polymerase core and the exonuclease domain, the catalytic subunits have N- and C- terminal extensions which vary in their lengths and degree of invariability from one polymerase to another (114). An important extension is the conserved C-terminal domain that is shared among the three polymerases and have two metal binding sites. In comparison to Pol α and Pol δ, Pol ε has the longest insertions (115). In fact, Pol ε’s catalytic subunit has duplicate polymerase and exonuclease domains with only one active set (113). The extensions of Pol ε’s catalytic subunit probably explains its higher intrinsic processivity compared to Pol α and Pol δ (116). From that perspective, Pol α has the lowest processivity among the three replicative polymerases.
Figure 1.4. The catalytic subunits of eukaryotic replicative polymerases adopt a right-hand fold. (A) The crystal structure of yeast Pol1349–1258 (core catalytic subunit of Pol α) bound to the primer-template and incoming nucleotide (PDB: 4B08 (117)). (B) The crystal structure of yeast Pol368–985 (core catalytic subunit of Pol δ) bound to the primer-template and incoming nucleotide (PDB: 3IAY (118)). (C) The crystal structure of the yeast Pol21–1228 (core catalytic subunit of Pol ε) bound to the primer-template and incoming nucleotide (to the left) (PDB: 4M8O (116)). To the right, 180° rotation showing the same structure from different orientation revealing the P-domain. For comparison, the four DNA polymerases are shown in similar orientations and similar subdomain coloring. The DNA polymerases all adopt a right-hand framework with the fingers (blue) contacting the incoming nucleotide (magenta) and the template strand, the palm containing the DNA polymerase (wheat) and the 3′–5′ exonuclease (cyan) domains, and the thumb (pale green) binding and directing the primer-template (orange) to the polymerase active sites. The N-terminal region is colored in yellow and the P-domain is shown in chocolate. Figure adapted from (114).

The widely accepted working model for the three replicative polymerases division of labor is as follows. Pol α-primase complex initiates DNA synthesis at the origin both on
leading and lagging strands as well as primes the synthesis of every OF on the lagging strand. Pol ε is the major polymerase synthesizing the bulk of the leading strand, whereas Pol δ carries on the synthesis of OFs on the lagging strand (119,120) (Figure 1.2). A multitude of evidence supports this model, chiefly by analyzing replication errors (121), examining genomic rNMP incorporation patterns (122) and monitoring the localization of polymerases on replication forks (123). This division of labor is further supported by recent in vitro reconstitutions of budding yeast replication in the presence of CMG helicase, which seemed to enforce the proposed model where Pol ε’s activity is suppressed on the lagging strand and Pol δ’s synthesis is suppressed on the leading strand (86,87,124). However, a recent study disputed this division of labor and suggested that Pol δ replicates both leading and lagging strands (125), whereas another study proposed a limited involvement of Pol δ at the initiation stage of leading strand synthesis (87). Nevertheless, the involvement of Pol α-primase complex in priming DNA synthesis and that of Pol δ in lagging strand synthesis have not been rebutted.

1.3.2.2 Priming synthesis by Pol α-primase complex

In eukaryotic cells, the primase activity is carried out by the accessory subunits of Pol α-primase complex, also known as the eukaryotic primosome (120). These accessory subunits form a constitutive heterodimer primase complex composed of a catalytic subunit PriS/p49 (yeast/human) and a regulatory subunit PriL/p58, with the latter consisting of two domains connected via a flexible linker, an N-terminal domain (p58N) and a C-terminal
domain (p58C) containing a 4Fe-4S cluster (126). This heterodimer primes DNA synthesis by synthesizing an RNA primer of 8-12 nts which is intramolecularly handed-off to the polymerase catalytic subunit of the primosome (127). The polymerase catalytic subunit (p180 in human) in turn extends the RNA primer with dNTPs to ~30-40 nt long RNA/DNA hybrid primer (128). The primosome is recruited to the replication fork through its interactions with Ctf4 as described earlier. This indirect physical interaction between Pol α and the CMG presumably improves the priming efficiency of each OF on the lagging strand. Moreover, it was shown that the priming activity of the primosome on both strands is greatly reduced in the presence of RPA (129), and that this inhibition is rescued by the presence of CMG (124).

Until recently, the mechanism by which the primase maintains a species-specific unit length of RNA primer, or what is known as counting, and the details of the intramolecular switch between the two catalytic subunits of the primosome had puzzled researchers in the field. Recent advances with structural and biochemical studies of human and yeast primosomes have shed some light on these mechanisms (115,117,126,130). The more recent crystal structure of the full human primosome in the apo form exhibits a structure where the polymerase core and p58C of the primase are connected via flexible linkers to a stable platform composed of the rest of the primosome (131) (Figure 1.5). It is suggested that the sizeable flexibility of the polymerase core and p58C with respect to the platform drives the internal switching between the priming and extension activities of the primase catalytic subunit and the polymerase catalytic subunit, respectively. The study further suggests that the RNA primer length is dictated by steric clashes between p58N and p58C of the primase regulatory subunit as p58C maintains interactions with the 5’ terminus of
the growing RNA and rotates away from the primase catalytic subunit to accommodate the growing RNA primer (Figure 1.5). Thus, after a certain length, estimated at 10 nts by molecular modeling, the rotation of p58C is no longer tolerated (131). The inhibition of further extension of the RNA primer by the primase therefore leads to its transfer to the polymerase subunit (Figure 1.5). The structure predicts that it is p58C that delivers the primer terminus to the polymerase, since it makes more significant interactions with the template-RNA primer substrate than p49. Hence, the structure confirms the biochemical data that the polymerase cannot access the 3’ terminus of the RNA primer unless it is 9 nts long.

Figure 1.5. A model of the eukaryotic priming mechanism. This schematic is based on the crystal structure of the human apo-primosome with the subunits referred to as the human proteins’ nomenclature. The p49,
p58N, p180C and p70 form a stable platform upon which the switch between the RNA and DNA primer synthesis occurs. The p58C and p180 core are attached to this platform with flexible linkers that mediate the two polymerases activities and the switch between them. To initiate synthesis, p58C moves closer to p49. As the RNA primer is extended, p58C moves away from p49 towards p180 core. When the primer reaches a critical length (~9 nts), steric hindrance between p58C and p58N hinders further RNA primer extension. p58C then delivers the RNA primer to p180 core active site. At that stage, p180 core extends the RNA primer with dNTPs while p58C maintains interactions with the primer 5’ end. Finally, the primer is handed off to Pol ε or Pol δ for further extension. This figure is adapted from (131).

1.3.2.3 DNA extension by Pol ε and Pol δ

The 30-40 nucleotide RNA/DNA hybrid is then further extended by Pol ε on the leading strand and Pol δ on the lagging strand for more processive and accurate DNA synthesis. Whereas both Pol ε and Pol δ possess high-fidelity synthesis, their DNA synthesis processivity varies significantly. Pol ε is endowed with intrinsic processivity most likely due to its extensive contact with the DNA as a result of its overall extended structure (132) and the presence of a variant domain in its core polymerase subunit (domain P) that allows Pol ε to encircle the nascent dsDNA as it leaves the active site (116) (Figure 1.4C). Processivity is a pivotal characteristic of DNA synthesis and cannot be compromised if the DNA duplication time has to stay coupled to the cell division timing. However, this intrinsic variation in processivity between Pol ε and Pol δ is neutralized by tethering the polymerases to the DNA through a sliding clamp (Figure 1.2).

The eukaryotic sliding clamp, called proliferating cell nuclear antigen (PCNA), was originally identified as an antigen for an autoimmune-disease (133). PCNA is a ring-shaped homotrimer with pseudo sixfold symmetry that accommodates duplex DNA in its central
channel and slides along dsDNA (134). The homotrimer assembles in a head-to-tail arrangement with two asymmetric faces, the front face (consisting of C-termini) and the back face (consisting of N-termini). Each subunit consists of two domains (A and B) bridged together by the interdomain connecting loop (IDCL), which faces the front face and acts as a hub for binding of many partner proteins that possess the PIP box (135,136). Besides its role as a processivity factor for DNA polymerases, PCNA has been described as the maestro of the replication fork for its diverse and vast set of binding partners and its involvement in many DNA transactions including OF maturation, DNA repair and recombination, chromatin assembly and cell cycle control (137).

PCNA exists as a ring-shaped circular clamp in solution; thus, its autonomous efficient loading on dsDNA is topologically infeasible. The eukaryotic clamp loader, called replication factor C (RFC), was initially identified in SV40 system. RFC is a heteropentamer consisting of 5 subunits all containing an AAA+ ATPase related domain (138,139) (Figure 1.2). RFC loads PCNA onto dsDNA in a sophisticated process requiring ATP binding and subsequent hydrolysis, which leads to opening and reclosing of the ring structure. RFC binds the front face of PCNA inducing an opening at one of the subunits interfaces and binds to the 3’ terminus of a primer-template junction in an ATP-binding-dependent reaction (135). Consequently, ATP hydrolysis brings about a conformational change in RFC leading to its dissociation from PCNA and DNA leaving behind a closed PCNA ring encircling the dsDNA with its front face facing the 3’ terminus of the primer. The front face of the loaded PCNA is then free to interact with the polymerases in the correct orientation that supports processive DNA synthesis.
Nevertheless, the limited primer extension by Pol α and the switching of the RNA/DNA hybrid primer to the high-fidelity processive replicative polymerases are still not fully elucidated although several mechanisms have been proposed. Based on the differential binding affinity of Pol α to RNA/DNA duplexes with an A-form helix than to DNA/DNA B-form helices, it was suggested that perhaps the helical characteristics of the growing RNA/DNA duplex can be sensed by Pol α’s active site, and once the primer-template switches to a B-form, it induces Pol α’s dissociation from the DNA (117). However, the template used in this study was a poly (dT), a template that supports triplex structure formation and has been shown to inhibit processive synthesis by most DNA polymerases not just Pol α (140,141). Alternatively, it has been proposed that PCNA loading by RFC to the 3’ terminus of the primer mediates the switch from Pol α to Pol δ (142-144). This is supported by the high concentration RFC-induced inhibition of Pol α in the absence of PCNA (144) and rescuing of this inhibition in the presence of PCNA (142). Moreover, RPA presence at the primer-template junction has been shown to stimulate the polymerase switching from Pol α to Pol δ by directly binding RFC and allowing for higher specificity of PCNA loading and thus Pol α displacement (145,146). On the other hand, the mechanism of polymerase switching from Pol α to Pol ε on the leading strand is still elusive. Pol ε intrinsic processivity, even in the absence of RFC and PCNA, could possibly explain this switch with a processivity competition model favoring Pol ε’s access to the 3’ primer terminus after the less processive Pol α synthesizes a short stretch of DNA.

Regardless of the switching mechanism, Pol ε replicates the leading strand continuously and Pol δ replicates the lagging strand discontinuously. It is worthy to note that a recent low-resolution EM structure of the CMG with Pol ε strikingly shows Pol ε
interacting with the CTD motor tier of the CMG, in particular the CTDs of Mcm2 and Mcm5 ahead of the fork (147). This position was further verified by extensive cross-linking mass spectrometry analysis. With Pol α, via Ctf4, positioned at the NTD tier, this arrangement of Pol ε on the CTD tier opens the possibility to redefining the architecture of the replisome and the DNA path during helicase activity, especially whether the split point is internal within the CMG or just before entry to the CMG. Yet, it is still puzzling how the priming activity of Pol α and the extension by Pol ε are coordinated with this architecture placing them on either side of the CMG. However, with this arrangement, one might envision Pol ε being the first to interact with nucleosomes (120). On the other hand, the discontinuous nature of lagging strand synthesis has attracted many researchers to decipher the steps leading to a continuous heritable dsDNA in a process called OF maturation as detailed below.

1.3.3 Okazaki fragment maturation

The discontinuity of the lagging strand synthesis poses several challenges. The newly synthesized lagging strand would harbor multiple nicks between the OFs. The priming requirement of each of the OFs would lead to heterogeneity in the nucleic acid composition. Moreover, the DNA stretches laid down by the exonuclease-deficient Pol α might present a hotspot for inaccurately incorporated nucleotides. Therefore, for the generation of contiguous heritable dsDNA, it is of paramount importance for the RNA primers to be excised, at least some of the Pol α-synthesized stretches to be corrected for, and lastly the
nicks to be sealed. Each cell cycle generates ~50 million OFs. Failure to process these fragments efficiently and with high fidelity is detrimental to genomic stability and integrity.

1.3.3.1 Pol δ strand displacement and its regulation

Our current understanding of OF maturation stems mostly from biochemical, genetic and structural work on proteins involved in this process. As Pol δ extends the nascent OF, it is faced with the RNA/DNA hybrid primer of the previous OF (Figure 1.6). Pol δ continues to replicate using its limited strand displacement activity, thus gradually displacing the RNA/DNA hybrid primer creating a 5’ flap (148,149). This strand displacement action is counteracted by Pol δ’s exonuclease activity which cleaves the last nucleotide that Pol δ just synthesized releasing dNMPs. This competition between the forward strand displacement activity and the backward exonuclease activity is referred to as “polymerase idling” (Figure 1.6). This idling process, in the majority of sequence contexts, limits the flap length to about three nucleotides long (148,150,151). It is worthy to note that both strand displacement and polymerase idling seem to be exclusive to the lagging-strand polymerase. In fact, both activities are very weak in Pol ε (150,152), which is befitting considering their role in OF maturation.

On the other hand, as the RNA flap continues to grow, it acts as a “molecular brake” hindering the progress of Pol δ strand displacement. In other words, the rate of the forward strand displacement reaction decreases as the flap length increases (151). The efficiency of the forward strand displacement activity strikingly does not discriminate between RNA and DNA content of the previous OF, but is rather dependent upon the stability of the
duplex (148,151). Factors lowering the stability of the duplex including sequence contexts with AT-rich regions have been shown to enhance Pol δ’s strand displacement activity to extensive lengths. Therefore, the idling process, along with the inhibitory effect of the flap on Pol δ strand displacement, ensures that the 5’ flap length remains short and that extensive strand displacement is restricted (Figure 1.6). In fact, their cooperativity is essential as neither of them is sufficient to maintain Pol δ next to the nick. Hypothetically, if the flap length-dependent molecular brake did not exist, idling alone cannot catch up with the fast-forward reaction and flaps would grow even longer. Vice versa, if idling mechanism were to fail, the forward strand displacement would still generate longer flaps, albeit with slower rates (120). However, it is worthy to mention that the inhibitory effect of the flap length on strand displacement fails if the flap length grows beyond a “critical” length (153). The exact size of this “critical length” and the mechanisms by which Pol δ bypasses its regulatory mechanisms are still unknown. It is possible that Pol δ interacts with the 5’-end of the short flaps (154), but once the flap grows longer, Pol δ can no longer maintain these interactions and the regulatory mechanism fails resulting in Pol δ extensively strand displacing the previous OF with constant rates comparable to the very initial rates. Further work is needed to decipher how Pol δ decouples from its regulatory mechanisms and define the exact size of the “critical length.”

1.3.3.2 Short-flap pathway

The flap created by the strand displacement of Pol δ is recognized and endonucleolytically cleaved by a structure-specific nuclease called flap endonuclease 1
FEN1 employs a highly sophisticated substrate recognition mechanism that ensures efficient and specific incision of the flap to create a ligateable nick. An overview of the current literature status describing this substrate recognition mechanism is discussed in details in Chapter 3. This elaborate mechanism has also motivated our investigations as described in the same chapter. Briefly, FEN1 recognizes the flap substrate in an elegant mechanism involving the bending of the duplex DNA to ~100° angle, undergoes a disorder-to-order transition, and positions the scissile phosphate in the active site (155-158). FEN1 then catalyzes an incision reaction that is a metal ion-dependent nucleophilic attack involving a water molecule to hydrolyze the phosphodiester bond. The incision by FEN1 creates two products a short ssDNA 5’ flap and a nicked dsDNA product. The upstream region of the substrate corresponds to the nascent OF while the downstream region corresponds to the previous OF containing the 5’ flap.

Biochemical in vitro experiments sought to characterize the preferred substrate for FEN1 cleavage reaction. It was initially thought that single flap (SF) substrates comprising only a 5’ flap on the downstream region while maintaining a fully annealed upstream duplex was FEN1 optimal substrate. However, later studies showed that substrates containing double flap (DF) structures with variable length 5’ flaps and a strict 1 nt 3’ flap enhance FEN1-substrate affinity as well as cleavage efficiency and specificity (159,160). With these DF substrates, regardless of the flap length, FEN1 always incises one nucleotide inside the junction, which had not been the case with SF substrates (159). Therefore, biochemical studies moved to using static DF substrates with unpaired 1 nt 3’ flaps. Nonetheless, this does not quite describe the in vivo case. Both the priming of the OF by Pol α-primase complex and the strand displacement by Pol δ utilize the same template
strand. Thus, the more realistic \textit{in vivo} substrate is a DF substrate capable of equilibrating between a DF structure (where the 3’ nt is unpaired) or a SF substrate (where the 3’ nt is fully base-paired to the template). Using the equilibrated DFs, FEN1 incision always yields a ligateable nick.

There is a preponderance of evidence supporting that FEN1 cleaves the 5’ flap while it is short. In addition to Pol $\delta$’s limited and regulated strand displacement (discussed above), researchers have proposed that there is an active hand-off mechanism between Pol $\delta$ and FEN1 termed “nick translation” that ensures that FEN1 cleaves the 5’ flap as it is generated (Figure 1.6). Indeed, researchers have shown that the predominant product of FEN1 cleavage is a mono- or di-ribonucleotide (148,151). Therefore, Pol $\delta$ and FEN1 repeatedly cooperate to create (Pol $\delta$) and cleave (FEN1) the 5’ flap until the RNA/DNA hybrid primer is sufficiently processed generating a nick that can be sealed by DNA Ligase 1. This active hand-off mechanism between Pol $\delta$ and FEN1, and eventually to DNA Ligase 1, is believed to be coordinated via PCNA as discussed in more details in Chapter 5. This efficient nick translation in the absence of the ligase has been observed in yeast to persist almost endlessly until blocked by a DNA-binding protein such as nucleosomes (161). These efficient processes ensure that the vast majority of OFs are processed through the short flap pathway as described here. Yet, on rare incidences, FEN1 cleavage reaction becomes uncoupled from the nick translation process, and Pol $\delta$ proceeds with strand displacement to extensive lengths that require the action of a backup pathway, called the long flap pathway, to avert causing DNA damage (162,163).
1.3.3.3 Long-flap pathway

In those rare events where FEN1 cleavage decouples from nick translation and Pol δ’s strand displacement generates flaps longer than 20 nts long, these flaps become readily accessible by RPA. The sequence context could also affect Pol δ’s strand displacement synthesis rate, such as in AT-regions where the synthesis rate is expected to be faster (151). Nevertheless, once bound by RPA, these flaps become resistant to FEN1 cleavage. FEN1 cannot displace RPA from these RPA-coated flaps and its cleavage of such substrates is inhibited. These long flaps are detrimental to genomic stability and integrity as they can form secondary structures that might hinder the progression of DNA replication and repair or alternatively recombine at ectopic sites leading to duplication events, among other toxic effects (164). Therefore, cells are intolerable to such long flaps if left unprocessed.

The processing of long flaps requires the action of Dna2 (Figure 1.6). Dna2 is an essential protein in yeast with an ATPase-DNA helicase, 5’ flap endonuclease and cell-cycle checkpoint activities (165-167). In addition to its involvement in processing the long flaps during OF maturation, Dna2 is involved in end resection in double strand break repair pathway (168). On RPA-coated long flaps, Dna2 tracks down the 5’ end of the 5’flap thus displacing RPA (169,170) and progressively cleaving the flap. The predominant evidence suggests that Dna2 cleaves inside the flap leaving a short 5-8 nt long flap. Although one study reported that Dna2 can cleave at the base of the flap (171), the consensus is that further processing is required following Dna2 cleavage. Dna2 has been shown to have low affinity to short flaps, so it most likely dissociates (172). Alternatively, FEN1 has been reported to disengage Dna2 (173,174). In either case, the short flap is free again and can
be accessed and cleaved by FEN1.

Our current knowledge of Dna2 involvement in long flap processing stems mainly from genetic studies in yeast and biochemical studies on both yeast and human Dna2. It is worthy to mention that RAD27 (yeast FEN1) is not essential as its deletion mutation renders viable cells albeit with severe mutator phenotype. This was explained by the presence of other backup nucleases including the related exonuclease 1 (EXO1) and RNaseH2 which can substitute for FEN1 on short flaps (148,175). Genetic studies investigating the interplay between yeast FEN1 and Dna2 showed that overexpression of either protein suppresses the effect of deleting the other one; in other words, cells with conditional lethality mutation of Dna2 were rescued by overexpressing RAD27, whereas overexpressing Dna2 in cells with Rad27 deletion mutation suppressed their temperature sensitivity (176). Therefore, Dna2 was considered the principal nuclease in processing long flap pathways.

There is ample evidence suggesting the existence of long flaps both \textit{in vivo} and \textit{in vitro} (172,177-179). How these long flaps occur and how would they escape the perfect catalysis by FEN1 in the first place are questions that are very mechanistically intriguing. Yeast genetic studies with deletion mutations of a helicase, called Pif1, and the third subunit of Pol δ have implicated these two proteins in enhancing the strand displacement activity of Pol δ (148,180). Posttranslational modifications of lagging strand proteins also shed some light on the role of these enzymes in switching between the short and long flap pathways (164,181-183). Moreover, RPA was proposed to govern this switch (162). However, a complete picture of how and why the switch to the long flap pathway occurs is still missing.

In Chapter 4, we review the current literature status of this pathway and more importantly,
we provide an elegant piece of this puzzle where we show that FEN1 itself can play a role in switching to the long flap pathway.

**Figure 1.6. Okazaki fragment maturation regulation.** Pol δ strand displaces the RNA primer of the previous OF generating a flap structure. This flap length is maintained at short lengths by the combined effect of Pol δ 3’-5’ exonuclease activity and the negative effect of the growing 5’ flap acting as a “molecular brake” on strand displacement speed. The competition between the forward strand displacement and the backward exonuclease activity is referred to as idling. The repetitive action of Pol δ strand displacement and FEN1 cleaving the 1 nt flap in a tightly-coupled process called nick translation ensures that flap does not grow longer. In the rare event that the flap grows long, processing of this flap requires the action of the helicase/nuclease Dna2. This figure is adapted from (120).

### 1.3.3.4 DNA ligation

The successful processing of the flap structure leaves a nick in the newly synthesized strand, which upon ligation results in a contiguous dsDNA. This nick is sealed by the ATP- and metal ion-dependent action of DNA Ligase 1 (184). The classical DNA ligation reaction occurs in three steps; 1) the adenylation of the ligase at a conserved lysine, 2) the
activation of the 5’ phosphate group of the nick by transferring the AMP group from the protein to the DNA and 3) the metal ion-dependent phosphodiester bond formation catalyzed by the nucleophilic attack of the adenylated phosphate group onto the 3’ hydroxyl group and release of AMP (184). The human crystal structure of DNA Ligase 1 has been solved and it reveals unique features of the mammalian ligases and suggests possible mechanisms for their interactions with the DNA and PCNA (185). PCNA-Ligase interaction has been shown to be vital for successful OF maturation in mammalian cells (186,187). PCNA also stimulates the ligase activity in vitro (188). Nevertheless, the mechanistic details of PCNA coordination of the different protein partners involved in OF maturation are still not well understood. Two competing models, “toolbelt” and “sequential” models, have been proposed to explain PCNA coordination of all three activities (synthesis, primer removal and ligation). These models will be discussed in depth in Chapter 5 along with our preliminary data on the subject.

1.3.3.5 PCNA unloading

After the completion of DNA ligation of OFs, the fate of the loaded PCNA poses an interesting question. Just as PCNA cannot load itself on dsDNA and requires the action of the heteropentameric RFC complex, it cannot unload itself either. The largest subunit of the eukaryotic clamp loader RFC1 has three eukaryotic paralogues, RAD17, CTF18 and ELG1, that along with the other four subunits of RFC can form different loader complexes (189,190). RAD17-RFC complex loads the 9-1-1 clamp onto DNA damage sites and together they signal DNA damage to DNA repair and checkpoint-activation machineries (191), whereas CTF18-RFC loads PCNA onto the leading strand presumably through its
interaction with Pol ε (192). As for ELG1-RFC, its main function is unloading of PCNA from lagging strand after the completion of DNA ligation (189,193).

Evidence supporting ELG1-RFC role in unloading of PCNA comes from both in vivo and in vitro work. Yeast and human cells with depleted ELG1 experience an accumulation of DNA-bound PCNA (194). This accumulation is rescued with the expression of ELG1 in these cells. Alternatively, PCNA unloads from isolated chromatin in vitro as partially purified ELG1-RFC is added (193). Furthermore, studies in yeast showed that during replication fork stall on the leading strand, PCNA redistribute to the leading strand synthesis (123). Therefore, PCNA unloading and recycling is essential, and consequently, ELG1-RFC role in this unloading is pivotal. However, ELG1 is not essential in yeast suggesting that other clamp loaders can unload PCNA. RFC and CTF18-RFC have been shown to unload PCNA from nick- or gap-containing DNA, but not from fully ligated dsDNA (195,196). The ability to unload PCNA from fully ligated dsDNA seems to be exclusive to ELG1-RFC as it fails to unload PCNA with ligase-deficient cell extract but resumes its action if an exogenous ligase is added (197). Therefore, this property might be advantageous to distinguish the proper timing of PCNA unloading only after ligation is concluded. SUMOylation of PCNA has also been proposed to regulate its unloading as SUMOylated PCNA accumulates on the chromatin in ELG1-deficient cells (198).

The significance of unloading PCNA goes further than just its recycling and redistribution. Controlling the life-span of the loaded PCNA on DNA while its being replicated is crucial to the overall organization and coordination of the various enzymatic activities on the chromosome as PCNA is involved in several processes through its diverse
set of protein partners (135). A DNA-bound PCNA acts as the “memory” of different events that occurred on the newly synthesized DNA (135). Its presence might signal different messages to other DNA-transaction enzymes. It can signal de novo replicated DNA regions or incompletely replicated or repaired regions (199). For instance, during mismatch repair in Xenopus egg extracts, MutSα, a protein specialized in recognizing mismatched bases, interacts with the DNA-bound PCNA preventing it from being unloaded (200). By preventing PCNA from being removed off the DNA in a particular orientation, the site is marked for mismatch repair and PCNA can only unload after the repair has occurred. It seems that MutSα might limit ELG1-RFC access to DNA-bound PCNA, thus explaining how PCNA is retained on the DNA. Finally, if PCNA is envisioned to contain the memory of events on the replicated strands, then ELG1-RFC can be imagined as an eraser removing PCNA memory (135).

1.4 Termination

The sequence-specific single origin initiation of E. coli genome has simplified the research into its replication termination. Although termination does not occur at a single specific site in E. coli, it is localized to a 270 kb broad region containing specialized fork barriers, that is the region containing TER sites where Tus protein binding to any of these sites constitute a fork barrier (201). Ample research has elucidated the mechanism of the role
played by such fork barriers in termination events (202-205). However, for eukaryotic replication, replication does not initiate from a single sequence-specific origin (as described above), which in turn would translate to a complex picture for termination as well. For many years, the research into eukaryotic termination events focused at the limited existing loci that contained specialized fork barriers such as the rDNA locus of metazoan and yeast. However, these loci do not account for all the termination events that are required for complete eukaryotic genome duplication (101).

1.4.1 Where and when does it end?

Eukaryotic replication termination occurs when two neighboring replication forks approaching from either side converge. Their meeting point is roughly around the midpoint. This was shown using high-resolution replication profiling (206) and deep sequencing of OFs (207) in budding yeast, and was further verified by OF mapping in human cells (208). This is intuitive considering that both replisomes from either end travel at similar average rate. However, different origins have different firing times and variable efficiencies, which would affect the exact location of the point of convergence of the two approaching forks.

Contrary to intuition, termination does not occur at the end of S phase, but rather throughout the entire S phase. Replications forks arising from early firing origins will terminate earlier than those arising from late origins would (101). This means that at least
some of the forks originating at early S-phase firing origins will terminate in early S-phase as well. On average a replicon size of 31 kbp (209) and an average fork speed of 1.5 kbp/min (210) would translate to 10 min period before the neighboring forks meet each other. It follows then that more termination events probably occur mid S-phase than late S-phase. This is in line with the strict replication-timing programme for each cell, which demands that only origins in difficult-to-replicate regions are initiated and replicated during late S-phase (211).

As for the spatial organization of termination events, it is also dictated by the spatio-temporal origin firing programme. This programme allows active more open chromatin to be replicated from origins fired early in S-phase (211). It is not surprising that these regions of the chromatin contain active genes and that origins were found in between those genes (135). Hence, it follows that many early S-phase termination events coincide with highly transcribed genes. On the other hand, the relatively less accessible heterochromatin is replicated from origins firing late in S-phase. Since heterochromatin does not largely correlate with expressed genes, it is only fitting that many late S-phase termination events overlapped with large non-expressed regions of the DNA (208). Therefore, the more recent high-resolution research showed that the earlier research into termination events at natural replication fork pausing sites biases the termination picture to those sites with elevated probability of termination most likely due to the fact that their replication forks originate from early-firing efficient origins. Vice versa, altering the origin-firing spatio-temporal programme for natural replication pausing sites and other sites changes the spectra of termination both spatially and temporally. Therefore, the spatio-temporal organization of
origin firing dictates when and where termination occurs (206, 207). The question that still lingers is “how do two replication forks converge?”

1.4.2 Eukaryotic replication termination: what we know so far

When two replication forks emanating from neighboring origins converge halfway, there is a myriad of challenges facing them (135). The two approaching forks each with its huge replisome machinery are traveling with relatively high speed and heading for a head-on collision. The remaining DNA segment has to be free of all DNA-binding proteins that could hinder fork progression. This DNA segment has to be unwound which could pose a challenge as there is a spatial restriction for the presence of Top 1. Hence, the torsional stress ahead of the forks cannot be easily released and would require its transformation to precatenates that would accumulate behind the forks. Even after the converging of the forks, the challenge continues as the remaining DNA fragments have to fully duplicated while keeping in mind that the last OFs on the lagging strands have to be processed. Moreover, the replication machinery has to be evicted from the DNA and the chromatin has to be re-established. Finally, entwined sister chromatids should be resolved topologically, and cohesion has to be established to maintain the proximity of the sister chromatids for successful separation during mitosis.

Only recently, our understanding of these processes at termination events started to crystalize mainly through the elegant work of Dewar et al. (212). Their work elucidated some of the mechanisms involving converging forks and their resolution. The researchers
elegantly designed a plasmid construct with an array of lac operators (LacOs) bound by lac repressors (LacRs) that can be reversed by addition of Isopropyl β-D-1-thiogalactopyranoside (IPTG). In doing so, termination events can be synchronized and monitored as the fork progression is paused by LacRs, and upon addition of IPTG, the replisomes are allowed to converge towards replicating the array of LacOs. When these plasmids are allowed to replicate in cell-free Xenopus laevis egg extracts, blocked forks were witnessed at the ends of the array and their release was induced by IPTG addition. This system allowed for monitoring the steps involving unwinding of the array DNA as the forks converge, DNA synthesis of the array, processing of the last replicated DNA segment into ligatable dsDNA and resolution of the sister chromatids (Figure 1.7). The study surprisingly showed a linear DNA synthesis rate within the array after LacRs release (212). This rate was comparable to the fork speed in the same extract. These results indicate that, contrary to previous belief, when two forks converge, they pass each other with similar rates to the fork progression without substantial slowing down, stalling or crashing into each other (Figure 1.7). This would only be feasible if the notion that the approaching CMGs propagate through DNA by encircling the leading strand is accepted (61,89,213), and hence, when they converge, they are on opposite strands and can pass each other without a head-on collision.
Figure 1.7. Model for eukaryotic DNA replication termination. As two adjacent forks converge, CMGs pass each other without colliding or slowing down, each moving on the leading strand of the two forks. The steric stress ahead of the fork is transformed into catenanes behind the fork since Topo I cannot be accommodated ahead of the fork. The DNA stretches between the two CMGs are replicated with continuous leading strand synthesis. The CMGs then slide on dsDNA. The last OFs on the lagging strands of both DNA molecules have to be processed. However, little is known regarding which polymerase achieves this feat and their mechanism. The catenanes behind the fork are resolved by Topo II. Finally, the CMGs dissociate off.
the DNA and the intertwined sister chromatids are completely decatenated. This figure is based on the work by Dewar et al. (212).

However, the study with its current synthetic design neither addresses the role of DNA-protein barriers ahead of the forks, nor fully elucidates the torsional stress building up ahead of the fork. As for the torsional stress ahead of the fork, in the context of this plasmid construct, Dewar et al. (212) showed that it does not slow down the fork progression, but these results might not apply to other constructs. As for the DNA-protein barriers, it has been reported that large barriers on the lagging strand can slow down fork progression (214,215). From the perspective of one of the moving replisomes on the leading strand, the other approaching replisome can be considered a barrier on the lagging strand, which should have slowed down the fork progression but was not the case in Dewar et al.’s study (212). More work to reconcile this seeming paradox is required. As for nucleosomes removal ahead of the fork, the study by Dewar et al. (212) cannot address such mechanism, since the synthetic design of the plasmid clears up the chromatin. The in vitro reconstitution of budding yeast DNA replication with purified proteins shed some light into the effect of nucleosomal packaging on DNA replication termination (86). The results of this reconstituted system suggest that termination but not elongation could be ultra-sensitive to chromatin remodeling.

At this stage, our knowledge concerning the DNA synthesis of the remaining DNA segment after the forks pass each other is limited to the data provided by Dewar et al. (212). DNA synthesis seemed to proceed upto few bases before the last OF with no noticeable
persistent gaps (Figure 1.7). Nevertheless, neither the question regarding which polymerase replicates the last fragment, nor the mechanism of processing the last OF were addressed. On the other hand, our knowledge of the replisome dissolution concluding the replication fork termination is relatively better advanced at this point. The replisome dissolution seems to be a highly conserved process among eukaryotes with work in budding yeast and *Xenopus laevis* egg extracts converging to similar models for this dissolution in both organisms (216,217). At the center of this process is the polyubiquitination of the MCM7 subunit of the CMG complex. For budding yeast, the ubiquitin ligase responsible for ubiquitylating MCM7 is SCF\textsuperscript{Dia2} (216), while for higher eukaryotes this role is played by CRL2\textsuperscript{Lrr1} ubiquitin Ligase (218). In either case, ubiquitin chains of MCM7 linked through lysine 48 are created while the replisome is bound to the DNA, but no degradation on the chromatin occurs. The replisome rather dissembles through binding to a protein remodeler called p97 (or VCP) in metazoans, Ter94 in insects, Cdc48 in yeast and CDC-48 in *Caenorhabditis elegans* (101). This protein remodeler is a hexameric ubiquitin-dependent AAA+ ATPase. Upon binding to ubiquitylated MCM7, p97 hydrolyzes ATP releasing energy that is required for its conformational changes in a motion called interprotomer transmission (219). With this motion, p97 is able to actively sequester the ubiquitylated CMG complex from the chromatin (216,217). Whether the ubiquitylated MCM7 subunit is destined for degradation or de-ubiquitylation is not clear at this point.
1.5 Looking ahead: Motivation of the work

We are still at an early stage of understanding the organization of the DNA replication machinery in its entirety. Nevertheless, the elegant and increasingly sophisticated *in vitro* reconstitution experiments and the enhanced structural approaches, especially the most recent improvements of cryo-EM, have led to major breakthroughs in DNA replication field. The reconstitution of the initiation of *S. cerevisiae* DNA replication (220) and the following reconstitutions of DNA replication of both naked DNA and chromatin templates (86,87,221) with minimal set of proteins have confirmed some major features of DNA replication that have long been expected as well as provided some ground-breaking features. In short, eukaryotic DNA replication of naked DNA requires a minimal set of over 40 polypeptides for the various stages. 1) ORC, Cdc6 and Cdt1 are required to load the core replicative helicase MCM2-7. 2) Cdc45, GINS, Sld3/7, Sld2, Dpb11, Pol ε, DDK, CDK, and MCM10 are essential for the assembly and activation of the CMG helicase. 3) RPA and Pol α-primase complex are required for the initiation of DNA synthesis. 4) Additionally, RFC, PCNA, topoisomerase I or II, and Pol δ are crucial for the efficient elongation of leading and lagging strand. 5) Moreover, FEN1 and DNA Ligase 1 are required for the efficient OF maturation. The reconstitution of eukaryotic DNA replication on chromatin further highlighted the function of other proteins involved in dismantling the nucleosomes ahead of the fork and their re-establishment behind the fork (161). Other DNA-binding proteins/protein complexes need to be removed for the progression of the replication forks such as loaded MCM2-7 on dormant origins (101). Finally, cohesion ring
complexes have to be established during DNA replication to maintain sister chromatids being held together until mitosis (222).

However, despite the paramount effort to decipher the mechanisms of the proteins and protein complexes involved in DNA replication, we are still a long way from painting a full picture of this complex process. The major breakthroughs with the reconstitutions of yeast DNA replication are still lacking in vivo confirmation. Furthermore, major mechanistic details regarding each and every piece of this tremendous undertaking are still missing. The pioneering genetic and biochemical approaches, as well as the structural experiments drew the broad contours of our understanding. The recent advances in cryo-EM are adding some colors to this overall picture. Nevertheless, with every technique comes a set of limitations and challenges. Although biochemical techniques can deliver some quantitative analyses of protein kinetics and general characterization, and biophysical techniques attempt to explain proteins binding and physical interactions, they both fall into the oversight of ensemble averaging where synchronization of individual molecules is a must and often challenging. Not only does ensemble averaging overlooks the diversity between different molecules, but it also averages the molecular behavior of each individual molecule leaving out vital information regarding the intermediary steps describing the dynamic behavior of each molecule. On the other hand, structural analyses have overcome the inaccuracy of ensemble averaging especially with single particle electron microscopy; but nonetheless, fail to draw the dynamic mechanistic behavior, as they merely provide static snapshots. The outgrowing scientific necessity to characterize the biological processes at the molecular and sub-molecular levels has led to the development of single molecule imaging technologies. Single molecule imaging techniques meet the needs and
fill the gaps left by other techniques unveiling long-standing mysteries of many biological processes that were left unanswered by the conventional techniques. Single molecule imaging techniques unravel those mysteries by attending to the details provided by each single molecule where synchronization is no longer a requirement. They also build a continuous timing mechanism of the process by accessing the intermediary steps of a reaction and recording the events and conformational changes in real time. In fact, the ability to manipulate and monitor single biological molecules is revolutionizing modern biological inquiry, allowing us to look at elementary biochemical reactions with unprecedented precision and clarity (223-226).

Taking one-step at a time, we focus our attention on OF maturation, with a special interest in the mechanistic details of FEN1 in the short- and long-flap pathways. FEN1, as a structure-specific nuclease, presents an intriguing platform for studying protein enzymology. Its sophisticated substrate recognition relying on DNA structure rather than sequence is in itself a testament to the fascination of how far evolution has come. This is only outshined by its perfect catalysis that reaches substrate-protein encounter and its high precision and selectivity. Moreover, the choice between the short- and long-flap pathways highlights several underlying regulatory mechanisms. Additionally, one cannot mention FEN1 regulation without referring to PCNA-mediated coordination of OF maturation. In this dissertation, we attempt to answer questions related to these three aspects in Chapters 3-5. We principally use single molecule Förster Resonance Energy Transfer (smFRET) complimented with ensemble kinetics and binding assays.
Briefly, we show, in Chapter 3, that FEN1 actively recognizes its substrate in an induced-fit mechanism under diffusion-limited kinetics. We then build a comprehensive kinetic scheme of FEN1 reaction from substrate recognition to product release. Additionally, we offer insight into how FEN1 achieves its high specificity by avoiding off-target substrates. In Chapter 4, we focus on the role played by FEN1 in the choice between the two OF maturation pathways. We discovered that FEN1 could trigger the long-flap pathway by missing the cleavage of long flaps. This observation is inaccessible and could be overlooked in ensemble assays. Finally, in Chapter 4, we shed some light on the kinetics of PCNA-mediated FEN1 product handoff to DNA Ligase 1. While Chapters 3 and 4 report on already published work, Chapter 5 presents preliminary results.
Chapter 2

2. Materials and Methods*

2.1 DNA substrates

DNA oligos were synthesized and HPLC purified by Integrated DNA technologies (IDT) or Sigma-Aldrich. Substrates were annealed by mixing template: 5’ flap: 3’ flap strands in 1: 3: 5 molar ratios in TE-100 buffer (50 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0, 100 mM NaCl), and heating at 95˚C for 5 mins followed by slow cooling to room temperature. For longer flaps used in cleavage assays, 10 mM MgCl$_2$ was added to the annealing buffer. Substrates were purified to >80% purity by non-denaturing polyacrylamide gel electrophoresis (PAGE) and eluted using the crush and soak method in TE-100 buffer, shaking at 16˚C for 30 mins. Eluted substrates were passed through 0.2 μm filters, aliquoted and stored at -20˚C. DF substrates were either equilibrating (EQ) or non-equilibrating (Non-EQ).


2.2 Proteins expression and purification

2.2.1 Human FEN1

Human FEN1 (amino acids: 2−380) was cloned into a pE-SumoPro expression vector (Lifesensors), which encodes an N-terminal 6xHis-Tag followed by SUMO protein. To create the R47A mutation, arginine 47 was mutated into alanine by site-directed mutagenesis kit following manufacturer’s instructions (Stratagene). Both proteins were expressed and purified following the same protocol described below. (Note: FEN1 used in ensemble cleavage kinetics assays (Section 2.6) was purified following (229)).

The clone expressing either protein was transformed into BL21 (DE3) E. coli strain for recombinant expression. The transformed cells were selectively grown on LB agar plates containing kanamycin (50 μg/mL). Positive transformants were inoculated in 2L 2xYT media and grown at 37°C to an OD$_{600}$ of 0.8, at which point the culture was shifted to 18°C, induced with 0.1 mM IPTG, and grown for an additional 12 hrs. The cell culture was harvested by centrifugation and lysed by sonication in lysis buffer-N (50 mM Tris-HCl pH 7.5, 10% (v/v) glycerol, 250 mM NaCl, 10 mM beta-mercaptoethanol (BME) and 25 mM imidazole). The lysate was cleared by centrifugation in ultra-centrifuge at 35,000 rpm at 4°C for 1 hr. The cleared lysate was loaded onto Ni-NTA column, washed extensively with buffer-N and eluted with linear gradient against buffer-N+250 mM imidazole. Fractions containing FEN1 were pooled and N-6xHis-tag SUMO protease (Ulp1) was added to cleave off the 6xHis-SUMO tag residues while being dialyzed against buffer-N at 4°C.
overnight. The cleaved 6xHis-SUMO tag and 6xHis-SUMO protease were removed over Ni-NTA chromatography. FEN1 protein, which went into flow through, was collected and diluted to a final salt concentration of 150 mM NaCl with buffer-H (50 mM Tris-HCl pH 7.5, 10% (v/v) glycerol and 10 mM BME). FEN1 was then loaded onto heparin column with buffer-H+100 mM NaCl, washed extensively with buffer-H+100 mM NaCl and eluted with a linear gradient against Buffer-H+1 M NaCl. Fractions containing FEN1 were pooled, concentrated and loaded onto HiLoad 16/60 Superdex 75 column (GE Healthcare) for further purification and buffer exchange into Buffer-S (50 mM Tris-HCl pH 7.5, 10% (v/v) glycerol, 300 mM NaCl and 10 mM BME). Fractions containing FEN1 were dialyzed against buffer-D (50 mM Tris-HCl pH 7.5, 50% (v/v) glycerol, 150 mM NaCl and 10 mM BME), collected, flash frozen and stored at -80°C.

2.2.2 Human RPA

The vector encoding full length human 3-subunit RPA (pET11d-tRPA) was a generous gift of Professor Marc S. Wold, and the protein was purified as previously described (230). Briefly, the plasmid was transformed into BL21 (DE3) cells and positive transformants selected on ampicillin-containing plates were grown in 3L TB media at 37°C overnight without shaking. The next morning shaking was started at 180 rpm and the culture continued to grow at 37°C until OD$_{600}$ reached 0.75. At that point, the culture was induced with 0.3 mM IPTG and continued to grow at 37°C (with shaking) for another 2 hrs. Cells were collected and resuspended in lysis buffer, HI buffer + 1 mM phenylmethane sulfonyl
fluoride (PMSF). HI buffer contained 30 mM HEPES-KOH pH=7.8, 0.25% (w/v) myo-inositol, 0.25 mM ethylenediaminetetraacetic acid (EDTA)-pH=8.0, 1 mM Dithiothreitol (DTT) and 0.01% (v/v) igepal (NP-40). Cell pellet was flash frozen in liquid nitrogen and stored at -80°C.

The purification proceeded by thawing the cell pellet, and then the cells were lysed at 4°C using 2 mg/mL lysozyme and sonication. The lysate was cleared with a single step centrifugation of 14,000 rpm at 4°C for 1:20 hr. The cleared lysate was loaded on a pre-equilibrated HiTrap Blue HP 5mL column (GE-healthcare) with HI buffer + 50 mM KCl. Extensive washing was performed in three steps with HI buffer containing increasing salt concentrations, 50 mM KCl, 800 mM KCl and 0.5 M NaSCN. The protein complex was then eluted in a linear gradient against HI buffer + 1.5 M NaSCN. Fractions containing the 3-subunit RPA were pooled and loaded onto home-packaged Hydroxyapatite (Bio-Rad) column pre-equilibrated with HI buffer. The protein was then eluted using HI buffer + 50 mM potassium phosphate (pH=7.8). The final step of purification was performed over MonoQ column (GE healthcare), pre-equilibrated with HI buffer + 100 mM KCl. After extensive washing, the protein was eluted with linear gradient against HI buffer + 400 mM KCl. Fractions containing the 3-subunit RPA complex were collected, concentrated and flash frozen in liquid nitrogen then stored at -80°C.
2.2.3 Human PCNA

2.2.3.1 Cloning, expression and purification

PCNA ORF was cloned in pET-Duet1 vector (Novagen) in MCS1 to encode PCNA with N-terminal 6xHis tag. This clone was then transformed into BL21 (DE3) *E. coli* cells and selectively grown over LB agar plates containing ampicillin. Positive transformants were inoculated into 2L LB media and grown at 37°C until OD<sub>600</sub> reached 0.7, at which point the culture was induced with 0.5 mM IPTG and shifted to grow further at 20°C. PCNA was purified, using standard methodologies, over sequential chromatographic columns that included HisTrap HP, Q-sepharose and finally Superdex 16/600-75 pg size exclusion using gel filtration buffer (50 mM HEPES-KOH pH=7.5, 100 mM NaCl and 0.5 mM Tris(2-carboxyethyl)phosphine hydrochloride (TCEP)). The cell lysis and purification over HisTrap HP column followed the protocol described above for human FEN1. The purification over Q-sepharose column was performed through a linear gradient against buffer containing 1 M NaCl. Unlabeled PCNA was purified over the same size exclusion column with gel filtration buffer 2 (50 mM Tris-HCl pH=7.5, 150 mM NaCl and 5 mM BME).

2.2.3.2 PCNA labelling

PCNA was non-specifically labelled with Cy5-Maleimide (GE Healthcare) following manufacturer’s instructions. Briefly, to label PCNA, 5-fold molar excess of Cy5-Maleimide was added to PCNA and incubated for 6 hours at 4°C. The labelling reaction
was quenched by adding DTT to a final concentration of 10 mM. The excess free dye was separated from Cy5-labelled PCNA using Superdex 75 10/300-GL gel filtration column. Fractions contacting Cy5-PCNA were collected, dialyzed against Storage Buffer (25 mM Tris–HCl pH=7.5, 50% (v/v) glycerol, 50 mM NaCl, 1 mM EDTA and 1 mM DTT), flash frozen in liquid nitrogen and stored at -80°C.

2.2.4 Human RFC

Human full-length RFC expression vectors pGBM-RFC1, pET-RFC4/2 and pCDFK-RFC5/3 were generous gifts of Dr. Yuji Masuda (231). An N-terminal truncated version of RFC1 was created by deleting the coding region for the first 550 amino acids and was replaced with strepII–tag coding sequence to create pGBM-ΔN-RFC1strepII. pCDFK-RFC5/3 was also modified to include 6xHis tag at the N-terminal of RFC3. These three plasmids were co-transformed into BL21 (DE3) cells and colonies were selected on agar plates containing three antibiotics (Kan+Amp+Str). RFC was overproduced by growing the transformed cells in 8L TB media containing the three antibiotics. Cells were grown at 25°C to OD$_{600}$ of 0.8 and then induced with 0.2 mM IPTG and incubated further for 24 hours at 16°C. Cells were collected by centrifugation and resuspended in buffer A+600 mM NaCl (Buffer A: 50 mM HEPES-KOH pH=7.5, 10 mM BME and 1 mM PMSF). All further steps were performed at 4°C. Cells were lysed enzymatically by adding 2 mg/ml lysozyme and mechanically by sonication. The lysate was clarified by centrifugation at 35,000 rpm for 50 mins. The clarified supernatant was adjusted to 10 mM imidazole and
loaded onto HisTrap HP 5ml column (GE Healthcare) and eluted with linear gradient between Buffer A and Buffer A+300 mM imidazole. Fractions containing all the subunits of RFC were combined and loaded onto StrepTrap HP 1ml column (GE healthcare). Only the incorrect stoichiometric complexes bound to the column, whereas correct stoichiometric complex went into the flow-through. Flow-through was collected and diluted to 300 mM NaCl with Buffer A, then loaded onto HiTrap heparin HP 1ml column (GE healthcare) and eluted with linear gradient between Buffer A+300 mM NaCl and Buffer A+1 M NaCl. Fractions that contained RFC subunits were collected, concentrated and loaded onto HiLoad Superdex 16/600-200 pg gel filtration column (GE healthcare) with storage buffer (25 mM Tris–HCl, pH=8.0, 250 mM NaCl, 0.01% NP-40, 1 mM DTT, 0.5 mM EDTA and 10% (v/v) glycerol). Fractions containing RFC with correct stoichiometry co-eluted and were collected, concentrated, flash frozen and stored at -80°C.

2.2.5 Human DNA Ligase 1

2.2.5.1 Full Length LIG1 (FL-LIG1)

_E. coli_ optimized sequence of full-length human DNA Ligase 1 (LIG1) was cloned untagged into pRSF1-b vector. Expression and purification followed the protocols published previously (232). This plasmid was transformed into BL21 (DE3)-RP cells (Stratagene) and grown over kanamycin and chloramphenicol containing LB agar plates. Positive transformants were inoculated into 2L 2xYT media containing the same
antibiotics, and grown at 37˚C until OD$_{600}$ reached 0.7. At this point, the culture was induced with 0.1 mM IPTG and shifted to grow further at 16˚C for 18 hrs. Cells were harvested and then lysed in lysis buffer A (50 mM Tris-HCl pH 7.5, 1 mM EDTA, 1 mM DTT, 10% (v/v) glycerol, 50 mM NaCl, 1 mM Benzamidine, 1 mM PMSF, 1% NP-40, 1 µg/ml pepstatin, and 1 µg/ml Aprotinin). Cells were lysed using 2 mg/mL lysozyme and cell disruptor. Lysate was cleared by centrifugation at 35,000 rpm, 4˚C for 50 mins. Cleared lysate was then loaded onto phosphocellulose (Whatman) column pre-equilibrated with buffer A. Extensive washing with buffer A was performed followed by linear gradient elution against Buffer A + 750 mM NaCl. Fractions containing LIG1 were pooled and 5-fold diluted with buffer A. The diluted protein fractions were then loaded onto Q Sepharose 5 mL column (GE Healthcare) pre-equilibrated with buffer A and eluted with a linear gradient against buffer A containing 350 mM NaCl. Protein fractions containing LIG1 are again pooled and loaded onto a HiTrap Blue HP 5mL column (GE healthcare) pre-equilibrated with buffer A containing 150 mM NaC and the protein was eluted with a linear gradient against buffer A containing 1 M NaCl after extensive washing. The final step of purification proceeded by concentrating the fractions containing LIG1 and loading the sample onto HiLoad 16/60 Superdex 200 pg size exclusion column (GE healthcare) using buffer A containing 150 mM NaCl. The fractions containing LIG1 are then pooled, dialyzed against storage buffer (25 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1 mM EDTA and 1 mM DTT), concentrated, flash frozen in liquid nitrogen at stored at -80˚C.
2.2.5.2 ΔN-LIG1

The sequence of truncated LIG1 lacking the N-terminal domain (1-232 a.a) was cloned into pE-SUMOpro vector. The expression and purification of ΔN-LIG1 followed the protocol described above for human FEN1 (Section 2.2.1).

2.3 TIRF-based smFRET

Glass coverslips were functionalized and passivated by 1:100 molar ratio of biotinylated polyethylene glycol (Biotin-PEG-SVA MW 5,000) and polyethylene glycol (mPEG-SVA MW 5000) (Laysan Bio Inc.). These functionalized coverslips, along with a quartz slide, were used to construct an airtight microfluidic flow cell with a 3 mm-wide channel of polyethylene double-sided tape SA-S-1L (100 µm thick) (Secure-Seal, Grace BioLabs) sandwiched in between. This flow cell had inlet and outlet tubings attached for exchange of buffer (Figure 2.1).

Figure 2.1. Schematic illustrating the construction of the microfluidic flow cell.
For immobilization of the double-labeled biotinylated DNA substrates, just prior to any experiment, the flow cell was incubated with 0.2 mg/mL NeutrAvidin for 10-15 mins. This treatment was followed by excessive washing with reaction buffer to remove excess NeutrAvidin and to block extra binding sites. The reaction buffer contained 50 mM HEPES-KOH pH 7.5, 5% (v/v) glycerol, 1 mM DTT, 0.1 mg/mL bovine albumin serum (BSA), 100 mM KCl and 10 mM salt containing divalent cations (CaCl$_2$ or MgCl$_2$ to assess bending or cleavage efficiency, respectively). The DNA substrates were then immobilized on the surface (~100-200 pM) until optimal coverage was reached, followed by excessive washing. Flowing DNA substrates and any subsequent steps occurred in imaging buffer (reaction buffer + oxygen scavenging solution).

To enhance the photostability of the fluorophores under our imaging conditions, we used an oxygen scavenging solution as described earlier (233), which relies on the enzymatic elimination of oxygen through a 6 mM proto-catechuic acid (PCA) (Sigma-Aldrich, P5630) and 60 nM protocatechuate-3,4-dioxygenase (PCD) system. To reduce the photo-blinking effect, we included 2 mM of the triplet-state quencher, Trolox (Sigma-Aldrich), in our imaging buffer.

Unless otherwise specified, single molecule experiments were performed using a custom-built TIRF-FRET setup (234) (Figure 2.2). Several movies on different fields of view were recorded for the DNA substrates and different protein additions using two-color alternating excitation (2c-ALEX) between green and red laser with a 160 ms time resolution and/or continuous wave (CW) excitation through green laser with a 50 or 100 ms time resolution. For each experiment, a transformation matrix file was constructed by
imaging a diluted sample of fluorescent beads (FluoSpheres, F8810 Invitrogen) in TIRF mode, then linking each particle in the green channel to its corresponding pair in the red channel. This transformation matrix was then used through twotone software (235) to map the donor and acceptor positions. In this process, certain restrictions concerning the brightness of the donor and acceptor, the distance between the centers of two adjacent particles, as well as the clustering distance are applied to ensure the particles are not too dim or too bright and are well-separated. The software then extracts donor and acceptor intensities by measuring the photon counts from the 2-dimensionally Gaussian-fitted point spread functions in both channels. These intensities are used by the software to calculate apparent FRET efficiency.
Figure 2.2. **Objective-based TIRF-FRET setup.** (A) A flow cell constructed with a functionalized coverslip and a quartz slide as in Figure 2.1. The doubly labeled DNA is tethered to the surface through biotin/NeutrAvidin interaction on a PEG-passivated coverslip. The flow cell is placed on top of the objective with oil immersion. (B) Schematic showing objective-based TIRF setup. Excitation is generated by a laser beam that passes through a half waveplate (λ/2) then a polarizing beam splitter (PBS) and gets expanded by a beam expander (BE). The laser beam is then focused through lens L1 and reflected on a dichroic mirror (DM1). The beam hits the periphery of the objective back focal plane leading to total internal reflection at the coverslip water-glass interface. The totally reflected beam passes through the objective leaving behind an evanescent excitation wave restricted to ~200 nm at the interface. The emission fluorescence of the labeled tethered molecules is then collected using the objective and it passes through a longpass (LP) filter and a second lens (L2). A dichroic mirror (DM2) then splits the beam according to the wavelength where the donor emission is reflected and the acceptor emission is allowed to pass through. The split emission beams pass through L3, then get focused on Mirror 3 and their emission is collected on EM-CCD camera. (C) The emission is recorded with 50-100 ms/frame resolution generating two-channel movies, one for the donor (green) and one for the acceptor (red). These movies
are processed to determine the intensity of single molecules donor and acceptor intensities (D) and consequently calculating FRET efficiency. Figure adapted from (236).

### 2.3.1 FEN1 DNA bending assays

These experiments were performed under the conditions described above in the presence of CaCl$_2$. For each condition, DNA-only and subsequent protein titrations, at least three 2c-ALEX and three CW movies were recorded. The histograms of apparent FRET efficiencies were constructed based on 2c-ALEX movies as described (237). These histograms were analyzed and fitted using OriginPro software to determine the center of FRET peaks and to integrate the percentage of each peak. The percentages of the bent-state peak versus FEN1 concentration were plotted and fitted to a one-site binding Model using GraphPad Prism software, and $K_{d\text{-bending}}$ was estimated using the constraint $B_{\text{max}} \leq 100$. For dwell time analysis of the time traces, CW movies were analyzed using the vbFRET package implemented in Matlab (238), where the time traces were idealized and fitted to two FRET states (bent and unbent). The dwell times spent in each state were plotted in histograms and fitted with a single exponential decay yielding $k_{\text{bending}} (1/\tau_{\text{bending}})$ and $k_{\text{unbending}} (1/\tau_{\text{unbending}})$. $k_{\text{bending}}$ was plotted versus different FEN1 concentrations to obtain $k_{\text{on\text{-bending}}}$ from the slope of the linear fit. $k_{\text{off\text{-bending}}}$ was obtained from the average of $k_{\text{unbending}}$ at different FEN1 concentrations.
2.3.2 FEN1 DNA bending in the presence of PCNA and/or LIG1

FEN1 DNA bending of respective substrates was assayed in the presence of 500 nM PCNA and/or 500 nM LIG1 following the methods described above. In those experiments, after substrate immobilization and movie recording, PCNA and/or LIG1 were injected into the flow cell and 2c-ALEX and CW movies were recorded. Excess proteins were washed by FEN1 buffer containing 250 mM NaCl followed by injection of imaging buffer. FEN1 titrations proceeded as above with each titration including 500 nM PCNA and/or LIG1. Each titration was followed by two washing steps, a 250 mM NaCl-containing FEN1 buffer followed by imaging buffer. Data acquisition and processing followed the methods described above for FEN1 alone experiments.

2.3.3 FEN1 single molecule cleavage assays

smFRET cleavage assays were performed using two labeling schemes, flap and internal. In both cases, DNA substrate immobilization and all subsequent steps were performed with imaging buffer containing 10 mM MgCl₂. For experiments with RPA, the DNA-immobilized surface was pre-incubated with sufficient RPA before co-injection of RPA and FEN1. Then 250 nM FEN1 was injected into the chamber with or without RPA. In all experiments, recording movies with single excitation by green laser at 50 ms temporal resolution was started before the proteins reached the microfluidic flow cell. In both labeling schemes, time traces were manually screened for cleavage events.
For the flap-labeling scheme, a cleavage event was identified as a transition from the unbent to bent FRET state (with clear anti-correlation between the donor and acceptor intensities), followed by a single-step loss of signal of both donor and acceptor. As a control, the acceptor was directly excited at the end of the experiment to ensure the selective loss of the donor signal due to cleavage. To differentiate between cleavage and photo-bleaching, we placed the more photo-stable Cy3 donor on the 5’flap. We further quantified the percentage of loss of Cy3 molecules due to photo-bleaching in the presence of CaCl₂ and FEN1, which was estimated to constitute ~15% while the signal loss due to FEN1 cleavage in the presence of MgCl₂ was ~80%, resulting in a signal loss of ~65% due to 5’flap cleavage. Among these cleaved events, the molecules with acceptor photo-bleaching were not used for further calculation. For viscosity measurements, falling balls viscometer (Gilmont) was used to calculate absolute viscosity. The density of a solution was determined by measuring the mass of 1 ml of same solution as used to calculate the viscosity.

For the internal-labeling scheme, a cleavage event was identified by following the transition from the unbent substrate FRET to that of the bent substrate and then the unbent nicked product. The unbent substrate and product FRET states showed a difference of ~0.05 as evident in the DNA-only histograms of the DF substrates and nicked duplex in CaCl₂ experiments. The dwell time of the substrate spent in bent state before loss of signal (in the flap-labeling scheme), and the dwell times spent in bent state before transitioning to the FRET state of the nick (in the internal-labeling scheme) were calculated by manually counting the frames in the bent state. We opted for manual counting of frames rather than automated fitting as these dwell times were too short to be picked accurately by most
available tools. In both cases, distributions of the dwell times were plotted and fitted to
gamma distributions using Matlab dfittool and the mean and the standard error of the mean
were reported.

The internal-labeling scheme was also used to assess missed cleavage events. An event
where bending occurred but the FRET state shifted back to that of the unbent DF substrate
(\sim 0.3) rather than the unbent nicked duplex (\sim 0.25) was considered a missed event. Dwell
times of the missed events were calculated as described above for cleavage events. The
percentage of missed cleavage was calculated as the percentage of particles (rather than
events) that showed at least one missed event.

The substrates used for single molecule experiments are listed in Tables 2.1-2.3
according to their appearance in Chapters 3-5.
Table 2.1. Substrates used for single molecule experiments in Chapter 3.

<table>
<thead>
<tr>
<th>Substrate Name</th>
<th>Substrate Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>NonEQ DF-4.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Non EQ DF-4.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>EQ DF-4.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>EQ DF-4.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Nick</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Product</td>
<td></td>
</tr>
</tbody>
</table>

Table 2.2. Substrates used for single molecule experiments in Chapter 4.

<table>
<thead>
<tr>
<th>Substrate Name</th>
<th>Substrate Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>EQ DF-4.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>EQ DF-4.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Non EQ DF-4.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>NonEQ DF-4.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>NonEQ DF-12.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>NonEQ DF-29.1</td>
<td></td>
</tr>
</tbody>
</table>

Legend: iCy3, Alexa647N, Biotin, Phosphate
2.4 Confocal-based smFRET

These set of experiments were performed on a custom-built confocal epifluorescence microscope setup (239). Fluorophores were excited with a 532 nm line of a pulsed laser diode operating at 20 MHz (100 ps pulse width, LDH-P-FA-530L, PicoQuant) or a 50 mW 532 nm Cobolt Samba laser through a microscope objective. A water immersion objective (UPLSAPO60XW NA 1.2, Olympus) and an oil immersion objective (UPLSAPO100XO NA 1.4, Olympus) were used for solution-phase smFRET and smFRET on surface-immobilized molecules, respectively. A circularly polarized beam was obtained by...
inserting a Berek compensator (Mo. no. 5540, Newport) in the excitation beam path. The laser beam was made Gaussian and expanded to fill the back aperture of the objective lens before introducing it into the microscope using a spatial filter with a 30 µm pinhole. The laser beam was reflected off the surface of a longpass dichroic Di02 R532-25x36 (Semrock Inc.) into the objective. The excitation power at the sample plane was set to 400 Wcm⁻² or 255 Wcm⁻² for solution-phase smFRET or smFRET on surface-immobilized molecules, respectively. In the detection path, emitted fluorescence passed through the dichroic and was focused onto a 100 µm pinhole by the tube lens of the microscope (IX71, Olympus) and re-collimated using a lens. A longpass BLP01-532R-25 filter (Semrock Inc.) was used to remove scattered laser light, and then the beam was split into donor and acceptor channels using a dichroic FF635-Di01-25x36 (Semrock Inc.). The donor and acceptor paths were equipped with a band-pass FF01-580/60-25-D (Semrock Inc.) and a longpass LP02-664RU-25 filter (Semrock Inc.), respectively, before being focused onto single-photon avalanche diodes (τSPAD, PicoQuant). Fluorescence intensity trajectories were recorded by a time-correlated single-photon counting (TCSPC) module (HydraHarp 400, PicoQuant) in the time-tagged time-resolved (TTTR) mode, which allowed for recording the arrival time of each photon emitted by the fluorophores. The SymPhoTime software (PicoQuant) was used for the data acquisition as well as for controlling the excitation lasers and TCSPC module.

2.4.1 Solution-phase smFRET experiments

The solution-phase smFRET experiments were performed in a home-made flow-through chamber by sandwiching a paraffin film spacer (0.13 mm thick, Bemis Inc.) between two
glass coverslips. The glass coverslips were functionalized and passivated with polyethylene glycol (mPEG-SCA, MW5000; Laysan Bio Inc.) prior to the construction of the flow-through chamber. Samples were prepared by mixing imaging buffer with appropriate dilutions of the stock enzyme and stock DNA solutions. The solution was allowed to flow onto the flow cell by pipetting the solution into one side of the chamber while applying suction to the opposite end. Then, the flow cell was placed on the microscope. In all solution-phase smFRET experiments, the reaction buffer was used with the addition of 10 mM CaCl$_2$ and 2 mM Trolox. In the solution-phase smFRET experiments, the excitation laser was focused approximately 40 µm above the surface of the bottom cover slip. The fluorescence intensity trajectories on the donor and acceptor channels were recorded for 15 minutes to obtain between 3500〜8000 bursts from individual DNA molecules. SymPhoTime script was used to analyze the bursts and generate burst histograms. The intensity trajectories were first binned to 0.5 ms, and bursts above 35 total counts were considered for the analysis. The FRET efficiency was calculated by the integrated intensity of each burst in the donor and acceptor channels. OriginPro was used to fit the histograms of the FRET efficiency to Gaussian peaks.

2.4.2 Surface-immobilized smFRET experiments

The smFRET experiments on surface-immobilized molecules were performed using either the microfluidic flow chamber used in the TIRF-based FRET experiments or pre-made sticky-Slide VI0.4 microfluidic chambers (ibidi GmbH), with cover slips identical to those
used in the TIRF-based FRET experiments. The DNA substrates were immobilized on the glass cover slips according to the procedures described above. The smFRET experiments were performed in the presence of the oxygen scavenger and the triplet quencher used in the TIRF-based FRET experiments. The excitation laser was focused on the surface of the cover slip using back reflection. Fluorescence intensity trajectories of individual molecules were acquired by first scanning a 10x10 µm section of the coverslip using a scanning piezo stage. Then, individual molecules were manually chosen from the image and the trajectories were sequentially acquired, with the laser focus dwelling on each point for 10 s. The SymPhoTime software was used for the image acquisition and stage positioning.

2.4.3 Data processing

Once fluorescence intensity trajectories were acquired, SymPhoTime was used to generate traces by binning the data to either 2, 5 or 10 ms and then exporting the donor and acceptor counts. A custom-written MATLAB script was used to generate traces from data exported from SymPhoTime and subsequently to select regions before photobleaching (https://github.com/harripd/ConfocalFret). Then, the FRET efficiency trace was calculated using the intensity trajectories of the donor and acceptor, and the histograms of the FRET efficiency were generated from the selected regions of the traces. Aberrant traces were excluded for further analysis. The selected regions were exported by the MATLAB script into files readable by HaMMy, a software used for analysis of single-molecule FRET trajectories using hidden Markov modeling (240). The FRET trajectories were analyzed
by a two-state model using HaMMy. Another custom-written MATLAB script was then used to collate the results from HaMMy and generate lists of dwell times for low and high FRET states. These lists were imported into OriginPro, histogrammed and fitted to a single exponential decay.

2.5 Bulk and single molecule PIFE experiments

For bulk time-resolved fluorescence lifetime measurements, a QuantaMaster 800 spectrofluorometer (Photon Technology International Inc.), coupled with a supercontinuum fiber laser source, was used. Cy3 fluorescence lifetime measurements of Non-EQ DF-6,1\text{PIFE} in the absence and presence of 1 \text{µM} FEN1 were performed at room temperature in TCSPC mode. Cy3 was excited at 532 nm and emission was collected at 568 nm with 5 nm-wide slits for both excitation and emission. To reduce collection of scattered light, a longpass filter of 550 nm was placed on the emission side. A suspension of colloidal silica was utilized to estimate the instrument response function (IRF). In all measurements, 10,000 counts were acquired. Cy3 fluorescence lifetime in both cases was then determined by a two-exponential decay fit using FluoFit software package (PicoQuant) that implements the IRF reconvolution. The best fit was achieved with a reduced chi-square and randomness of the residuals.
Single molecule PIFE experiments followed the standard conditions used for smFRET cleavage reactions with Non-EQ DF-6,1\textsubscript{PIFE} as the substrate. The data was analyzed in a similar fashion and the number of frames a molecule spent in the enhanced-fluorescence state were counted manually and fitted to a gamma distribution.

2.6 Ensemble single turnover and steady state cleavage assays

FEN1 cleavage activity on EQ DF substrates of varying lengths (DF-6,1, DF-30,1, DF-50,1, DF-60,1) was measured by bulk single turnover experiments using a rapid quench-flow apparatus (RQF-3; KinTek Corp.). The 5' end of the 5' flap strand was modified with FAM dye, and substrates were prepared by mixing 5' flap: template: 3' flap strands in 1:1.5: 2.5 ratios, heating the mixture at 80 °C for 10 minutes followed by spinning down and cooling O/N to 25°C for a yield of 75-85 % (annealing buffer: 50 mM HEPES-KOH, pH 7.5, 100 mM KCl, 10 mM MgCl\textsubscript{2}). A list of these substrates is included in Table 2.4.

In each experiment, 15 µl of FEN1 was mixed with 15 µl of DNA, incubated at 37°C for varying times and then quenched with 76 µl of 200 mM EDTA (final reaction conditions: 1.2 µM FEN1, 0.035 µM DNA in buffer containing 50 mM HEPES-KOH, pH 7.5, 40 or 100 mM KCl, 10 mM MgCl\textsubscript{2}, 0.1 mg/ml BSA, 5% glycerol, 1 mM DTT). The quenched reactions were placed on ice until analysis by denaturing PAGE. A 25 µl aliquot of each reaction was mixed with an equal volume of denaturing dye (0.3% bromophenol
blue, 12 mM EDTA in formamide), heated at 95°C for 1 min and run on an 18% denaturing urea polyacrylamide gel (3.25x3x0.15 cm) for 40 minutes at 12 W. The substrate and 5’-flap product were quantified on a Typhoon scanner (λ_{ex} = 488 nm). The fraction of product formed was plotted versus time, and the data were fit to a single exponential equation by GraphPad Prism to obtain the cleavage rate (\( k_{STO} \)).

In steady state experiments, 1 nM FEN1 was mixed with 800 nM DNA substrate (10x\( K_M \); (241)) at 37 °C, and 15 µl aliquots of the reaction were mixed with 4 µl of 100 mM EDTA at varying times to quench the reaction and determine initial velocity. The substrates and products were resolved and quantified as described above, and the data fit to a linear equation yielded the \( k_{cat} \) (slope/[FEN1]).

Table 2.4. Substrates used for ensemble single and multiple turnover kinetics.
2.7 Steady state bulk cleavage in the presence of RPA

EQ DF substrates of varying flap lengths (DF-2,1, DF-6,1, and DF-30,1) with Atto647 dye at the 3’ end of the 5’ flap oligo were used to assess FEN1 cleavage efficiency in the absence and presence of RPA. A list of these substrates is included in Table 2.5. For direct comparison, the assay was performed as described previously for yeast proteins (242). Briefly, 0.1 nM FEN1 and increasing RPA (0, 0.75, 1.25, 2.5, or 5 nM) were pre-mixed in buffer containing 30 mM HEPES-KOH, 0.5% inositol, 40 mM KCl, 4 mM MgCl₂, 0.01% Nonidet P-40, 0.1 mg/mL BSA, 1 mM DTT, and 5% (v/v) glycerol. DNA substrates were added at 0.25 nM concentration to initiate the reaction. The 20 µL reactions were incubated at 37°C for 10 mins before quenching with an equal volume of 2X buffer (90% deionized formamide, 100 mM EDTA). The samples were then heated at 95°C for 10 mins, cooled immediately on ice, and the products were resolved by 20% denaturing urea PAGE and quantified on a Typhoon TRIO Variable Mode Imager (GE Healthcare, Life Sciences).

Table 2.5. Substrates used for steady state bulk cleavage in presence of RPA.

<table>
<thead>
<tr>
<th>Substrate Name</th>
<th>Substrate Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>EQ DF-2,1</td>
<td>![SEQ_EQ_DF_2_1]</td>
</tr>
<tr>
<td>EQ DF-6,1</td>
<td>![SEQ_EQ_DF_6_1]</td>
</tr>
<tr>
<td>EQ DF-30,1</td>
<td>![SEQ_EQ_DF_30_1]</td>
</tr>
</tbody>
</table>
2.8 Surface Plasmon Resonance (SPR) binding

SPR binding was performed on a Biacore T100 (GE Healthcare Inc.). The corresponding biotinylated DNA substrates (DF-6,1, nick and product) were immobilized on S-series streptavidin sensor chips in HBS-EP buffer according to the manufacturer’s recommendations. FL-LIG1 was dialyzed overnight at 4°C against FEN1 buffer (50 mM HEPES-KOH pH=7.5, 100 mM KCl, 1 mM DTT, 5% (v/v) glycerol and 10 mM CaCl$_2$). Increasing concentrations of FL-LIG1 (100 nM, 500 nM and 1000 nM) were injected at a flow rate of 20 µL/min for 100 s. For each concentration, the run started with a surface-regeneration injection of reaction buffer+1 M NaCl at a flow rate of 100 µL/min for 120 s, followed by the protein injection. The sensorgrams were corrected for bulk refractive index and residual nonspecific binding with the surface using a blank flow cell. The sensorgrams were processed using Biacore T100 Evaluation Software (GE Healthcare Inc.).

2.9 PCNA loading on nicked DNA via ΔN-RFC

2.9.1 Bulk FRET assay

PCNA loading on nick DNA via ΔN-RFC in bulk FRET assay, followed the general methods described in (243). PCNA was loaded on a 5 nM Cy3-labeled nick DNA that
contained a biotin at the downstream side and a Digoxigenin at the upstream side. The nick DNA substrate was blocked from both ends by incubation with 100 nM NeutrAvidin and 40 μg/mL anti-Digoxigenin antibodies at room temperature for 10 mins. Steady state fluorescence measurements were conducted at room temperature using Fluoromax-4 (HORIBA JOBIN YVON). The fluorescence intensity of the nick substrate was measured in loading buffer (25 mM Tris-HCl pH=7.5, 10 mM magnesium acetate, 125 mM KCl, 0.1 mg/mL BSA, 5 mM ATP and 1 mM DTT). For measurements in the presence of PCNA, 10 nM Cy5-PCNA was added to the mixture of labeled substrates in loading buffer and allowed to reach equilibrium state after 5 mins incubation at 37°C. For ΔN-RFC-mediated PCNA loading, 10 nM ΔN-RFC from P1 or P2 peaks was incubated with the mixture for additional 5 mins at 37°C before measurements were recorded. In all cases, Cy3 was excited at 535 nm and emission spectra were collected between 550 and 700 nm. Both excitation and emission slit widths were set to 5 nm. Measurements were recorded with an integration time of 0.1 s. The emission spectra were corrected by subtracting the background emission of a blank solution comprised of loading buffer.

2.9.2 smFRET assay

For PCNA loading on nick DNA via ΔN-RFC in smFRET assay, a similar doubly blocked substrate from both ends was used, however, this substrate was dually labeled with Cy3 donor and Alexa Fluor 647 acceptor in the same positions used for internal-labeling scheme. For blocking the substrate, it was pre-incubated with 10 μg/mL anti-Dig antibodies
at RT for 10 mins before immobilizing on the functionalized coverslips. The surface was then washed with 1xPBS solution containing 250 mM NaCl followed by 1xPBS wash. For loading PCNA, 20 nM PCNA was mixed with 20 nM ΔN-RFC (P2 Peak) in loading buffer (50 mM Tris pH=7.5, 100 mM KCl, 10 mM magnesium acetate, 0.1 mg/mL BSA, 1 mM DTT and 5 mM ATP) and pre-incubated at 37°C for 10 mins. The protein mixture was then injected into the flow cell at lower flow rate and incubated for further 10 mins at RT. Excess proteins were washed with FEN1 buffer containing 250 mM NaCl followed by FEN1 buffer wash. FEN1 titrations then proceeded and DNA bending was characterized as described above (Section 2.3.1).
Chapter 3

3. Single-molecule FRET unveils induced-fit mechanism for substrate selectivity in flap endonuclease 1*

3.1 Abstract

Human flap endonuclease 1 (FEN1) and related structure-specific 5’nucleases precisely identify and incise aberrant DNA structures during replication, repair and recombination to avoid genomic instability. Yet, it is unclear how the mechanism of these nucleases in DNA distortion and protein ordering robustly mediate efficient and accurate substrate recognition and catalytic selectivity. Here, single-molecule sub-millisecond and millisecond analyses of FEN1 reveal a protein-DNA induced-fit mechanism that efficiently verifies the substrate and suppresses off-target cleavage. FEN1 sculpts DNA with diffusion-limited kinetics to test and verify its substrate. This DNA distortion mutually ‘locks’ the protein and DNA conformation and enables substrate verification with extreme precision. Strikingly, FEN1 never misses cleavage of its cognate substrate whereas it blocks the probable formation of catalytically competent interactions with noncognate substrates, thus fostering their pre-incision dissociation. These findings establish a comprehensive kinetic scheme of FEN1 cleavage reaction of its cognate substrate.
3.2 Introduction

Biologically-critical, structure-specific 5’ nucleases are highly conserved endo- or exo-nucleases that hydrolyze phosphodiester bonds that are one nucleotide into the 5’end of ssDNA/dsDNA junctions (Figure 3.1A), including nicks, gaps, flaps, bubbles and four-way junctions (155,160,244,245). These structure-specific 5’ nucleases are vital in many DNA metabolic pathways including DNA replication, repair and recombination (246,247). From the early discovery of the first 5’ nuclease of E. coli Pol I (248), it was clear that these nucleases present a new paradigm of substrate recognition. The peculiarity of their substrate recognition that relies on structure rather than the classical sequence-specificity has promoted extensive biochemical and structural research over three decades to decode their sophisticated substrate recognition mechanisms.
3.2.1 FEN1: a multi-functional small protein

In the context of DNA replication, FEN1 stands out as the primary nuclease at the center of Okazaki fragment (OF) maturation (249). The semi-discontinuous nature of DNA replication results in ~50 million OF being generated per each cell cycle. The processing of these OFs creates flap substrates that if left unprocessed are toxic to the genome stability and integrity. FEN1, as an archetype of 5’ structure-specific nucleases, employs a highly sophisticated mechanism to recognize these flap substrates. This relatively small protein is a monomeric metal ion-dependent nuclease comprised of three domains; the N-terminal (N) (1-104 a.a.), intermediate (I) (122-253 a.a.) and C-terminal (336-344 a.a.) domains with a further extended C-terminal region. Its catalytic core is composed of the N and I domains and hosts its 5’ endonuclease and 5’-3’ exonuclease activities. The C-terminal region is not essential for its catalysis but is reserved for its protein-protein interactions with several protein partners such as PCNA (250).

In addition to its vital role in OF maturation, FEN1 is involved in other cellular processes including long-patch base excision repair, stalled replication fork rescue, suppression of trinucleotide repeats expansion and telomere maintenance (244,251-253). Although FEN1 is not essential in yeast, its knockdown mutation in both S. cerevisiae and S. pombe cause severe mutator phenotypes resulting in genomic instability (254-257). More importantly, the implication of its mammalian homologs in disease states has been well documented. For example, the deletion of both fen1 alleles causes lethality in mice embryos (258). With its established role in DNA replication, it is only logical that it is expressed in all proliferating human tissues especially the fast-dividing cancerous cells.
(259,260). In fact, its expression level has been associated with the proliferation state of normal tissues and the degree of malignancy of their cancerous counterparts (261). However, mutations that cause FEN1 reduced expression or altered activity lead to cancer predisposition in both humans and mice (262-264). Therefore, a balance of both FEN1 expression and activity seems to be vital for cancer prevention. It is not surprising then that FEN1 inhibitors have been researched as potential anti-cancer drugs (265,266). Moreover, its suggested role in suppressing trinucleotide repeats expansion (267) could be essential for prevention of several neuromuscular and neurodegenerative diseases, including Huntington's disease and myotonic dystrophy (268). Thus, deciphering FEN1 substrate recognition and catalytic precision would present a platform for tackling the above challenges.

## 3.2.2 FEN1 cognate substrate

FEN1, as a structure-specific nuclease, has to rely on the structural features of the substrate itself. Early *in vitro* studies suggested that it recognizes substrates with single 5’ flaps (SF substrates) (269,270), consistent with the suggested cognate substrate of the bacteriophage homolog (T5FEN) (271,272). Later research confirmed that eukaryotic FEN1 preferentially recognizes double flap substrates with variable lengths of 5’ ssDNA/ssRNA flaps and a strict 1 nt 3’ flap (Figure 3.1A). This preferential recognition towards DF over SF substrates was confirmed through studies showing FEN1 differential binding affinity (159,273), catalytic efficiency and precision (159,241) of these substrates. However, these
*in vitro* studies used static DF substrates with an intentional unpaired 3’ flap. In contrast, during OF maturation *in vivo*, flap substrates are generated when Pol δ strand displaces the previous OF’s RNA/DNA hybrid primer synthesized by Pol α-primase complex. Since both polymerases use the same template strand for their synthesis, both ends of the nick junction can complement the template strand. Therefore, the *in vivo* FEN1 substrate has an equilibrated structure where the 5’ flap and 3’ flap can equilibrate between a SF and DF substrate (Figure 3.1A). The 1 nt 3’ flap in the equilibrated DF can reanneal back after cleavage to generate a ligateable product in contrast to the *in vitro* static DF substrate (254).

Biochemical experiments showed that FEN1 maintains exquisite specificity with extreme efficiency that enhances the hydrolysis rate of target phosphodiester bonds by ~10^{17}, and *in vitro* reaction rates resemble those of enzyme-substrate encounters (241). FEN1 cleavage reaction generates two products; a small ssDNA that constitutes the 5’ flap in the substrate and a nicked dsDNA. Kinetic characterization of FEN1 nuclease activity under single turnover and multiple turnover conditions showed that the nicked dsDNA product, but not the ssDNA 5’ flap product, is a competitive inhibitor of FEN1 nuclease activity (241). Therefore, under multiple turnover conditions, nicked product release is the rate-limiting step.
Figure 3.1. FEN1 substrate recognition. (A) FEN1 cleavage reaction. Schematic showing the equilibration of a flap substrate junction from a single to a double flap and its subsequent cleavage by FEN1 to generate a nick that can be sealed by DNA Ligase 1. (B) Ordering of FEN1 upon DNA binding. Structure of FEN1 in the apo form (PDB 1ULI; Left) (274) and in complex with bent DNA (PDB 3Q8L; Right) (156), highlighting the various structural features of FEN1 and the regions that undergo disorder-to-order transitioning upon DNA binding. (C) Active versus DNA conformational capturing models for forming the FEN1 complex with the bent DNA conformer.

3.2.3 Substrate recognition

The intricacy of FEN1 substrate recognition and its highly selective catalysis motivated
paramount biochemical and structural research investigating its specificity and selectivity. The most obvious physical feature of the flap substrate is its 5’ flap. Therefore, it is not surprising that the earliest proposed model of substrate recognition focused on this structural feature. The first model, called “tracking model” postulated that FEN1 first recognizes the 5’ flap then tracks along until it reaches the flap base where it catalyzes its incision reaction 1 nt inside the ssDNA/dsDNA junction towards the downstream strand. This model was supported by biochemical research investigating the effect of 5’ flap modifications on FEN1 nuclease efficiency. These studies showed that blocking the 5’ end of the flap by biotin/streptavidin interactions or inserting chemical adducts into the 5’ flap have inhibitory effects on FEN1 nuclease activity (275). Similarly, modifying the 5’ flap to include secondary structures or dsDNA (244), or altering the overall structure of the substrate to bubbles or branched structures (275) impaired FEN1 cleavage. Therefore, these studies concluded that the free 5’ end of the flap is integral to FEN1 activity.

However, the later structural and biochemical work changed our understanding as to how FEN1 recognizes its cognate substrate. This work on both FEN1 and another superfamily member, EXO1, shifted the focus from the 5’ flap to the dsDNA as the initial point of substrate recognition. The observation that FEN1 can bind substrates blocked at the 5’ end regardless of the inhibited nuclease activity (276) opposed the tracking mechanism where the 5’ end is the first point of recognition and gave credence to two proposed competing models, “threading” and “clamping”.

The threading model was first suggested by Joyce and coworkers while working with eubacterial FEN1 (277) and was later adopted for mammalian FEN1. This model describes
FEN1 first binding to the duplex region around the ssDNA/dsDNA junction then threading in the 5’ flap to position the scissile phosphate in the active site. This initial binding to the flap base is further stabilized by threading the 5’ flap through a structured channel in the nuclease, called the helical arch. The initial crystal structures of human FEN1 bound to product or substrate complexes and the more recent structure of FEN1-flap substrate complex further supported this threading model (156,157). A brief description of the main structural features revealed by these structures and their implications on FEN1 substrate recognition is discussed below.

The clamping model, on the other hand, was proposed when the crystal structure of human EXO1 was solved and clamping was suggested to be a unifying mechanism for both nucleases (278). This third proposed mechanism also suggests that dsDNA is the first point of recognition, but instead of the 5’ flap being threaded or pushed through the structured helical arch, the helical arch clamps around the 5’ flap. The authors of this study postulated that threading the 5’ flap through a restricted helical arch poses an energy cost that might not be accessible to the nuclease without a coupled energy source (278). Biochemical data in support of the clamping model mainly stems from the observation that flap substrates containing a short duplex region at the tip of the 5’ flap, named fold-back flap substrates, can be cleaved by FEN1 through its endonuclease activity (241). Such substrates challenge the proposition of threading the 5’ flap through a narrow structured helical gateway. On the other hand, clamping model is opposed by the biochemical studies showing that FEN1 incision activity is inhibited on substrates with blocked 5’ flaps at the tip of the flap through biotin/streptavidin interaction (279,280). For these substrates, if clamping model were to hold, then the substrates would present enough ssDNA for FEN1 to clamp around the 5’
flap close to the base and catalyze its incision.

### 3.2.4 FEN1 crystal structure

Structural work of FENs in their apo form from different organisms revealed a common overall architecture of their nuclease domain adopting an $\alpha/\beta$ structure named PIN or SAM fold (274,281-286). In these structures, a mixed-twisted $\beta$-sheet is flanked by two $\alpha$-helical regions and assembles in a saddle-like shape. A protrusion of two $\alpha$-helices from the saddle-like structure is also observed. The structures further supported the phosphodiester bond hydrolysis mechanism by the conserved seven acidic residues (287,288) through coordinating two metal ions (289).

However, a clearer picture of substrate recognition emerged as the crystal structures of FEN1-substrate and FEN1-product complexes were solved (Figure 3.1B-Right) (156). These structures showed the protein folding in an overall architecture reminiscent of a left-handed boxing glove interacting with both sides of the ssDNA/dsDNA nick junction through the palm and fingers. Through this interaction, which constitutes the majority of the protein-DNA interface, FEN1 induces a severe kink to the DNA at $\sim$100° angle between the two duplex regions. This interaction is mainly directed towards the template strand and supported by FEN1 helix-two-turn-helix (H2TH) motif coordinating a potassium ion. Furthermore, the active site hosting the two coordinated metal ions is seen assembled at the flap base. The structure further highlights other structural features that contribute to FEN1 sophisticated substrate recognition.
FEN1 recognizes the strict 1 nt 3’ flap through a specialized 3’ flap binding pocket (156). Next to the 3’ flap binding pocket is FEN-specific structure called the acid block which presumably blocks the DNA from moving further (Figure 3.1B-Right) (155). Another structural feature that was remarked is the hydrophobic wedge, which stabilizes the bent conformer. Comparison of these structures with the structure of human FEN1 in the apo form (Figure 3.1B-Left) (274) and FEN1 from *Archaeoglobus fulgidus* bound to the upstream region (282) led to the proposition that a FEN1 disorder-to-order transition occurs upon binding to the 3’ flap (Figure 3.1B). Moreover, the comparison with the human FEN1 in the apo form (274,286) revealed that the helical arch structures in a capped helical gateway-like conformation in the presence of DNA implying that FEN1 binding to the DNA induces the structuring of this region as well (Figure 3.1B) (155). This capped helical gateway, which forms a narrow channel incapable of accommodating dsDNA, is positioned on top of the active site (156). Whereas the helical gateway is a superfamily shared feature, the cap is specific to FEN1 and EXO1 (156,278), thus restricting the passage of structures that do not contain a free 5’ end such as bubbles and four-way junctions. It is worth mentioning that this ordering of the helical gateway cannot exclusively support the threading versus clamping model; however, the authors suggest that if FEN1 encounters substrates with sufficient ssDNA at the flap base, it can thread and order around the flap, and thus would be capable of cleaving structures such as those presented by fold-back flap substrates (155). The most recent crystal structure of human FEN1 with its cognate DF substrate settled the debate in favor of the threading mechanism (157). The authors further presented an elegant phosphate steering route to position the scissile phosphate in the active
site with an inverted threading mechanism. Nevertheless, the data presented in this chapter mostly precedes solving this structure.

3.2.5 FEN1 state of affairs

FEN1-substrate and FEN1-product crystal structures provide a prototypic system for unveiling the extreme catalytic selectivity of structure-specific 5’ nucleases. Like other 5’ nucleases, FEN1 displays maximum catalytic efficiency for its cognate substrate but it is only residually active on substrates that deviates only slightly from the cognate substrates (159,241). Catalysis is proposed to require both conformational changes in the protein itself, which result in active-site assembly (156,274,278,290-292), and movement of the scissile phosphate of the DNA substrate closer to the catalytic metals (156,278,292). Possible steps in the substrate selection and cleaving process have been previously described (156,274,278,290-292). However, key information about the mechanistic details of FEN1 and its substrate conformational changes that lead to its exquisite catalytic selectivity and efficiency remain controversial and largely undetermined. For example, the question of whether 5’ nucleases actively bend the DNA or selectively bind to a DNA that bends spontaneously remains ambiguous (Figure 3.3C) (158,293). The underlying mechanism of how FEN1 highly discriminates against noncognate substrates is also unclear. Moreover, a comprehensive kinetic scheme of FEN1 reaction from substrate binding to product release is still missing. Such mechanistic knowledge pertains not only
to biological understanding but also to strategies for the design of specific inhibitors as potential chemotherapeutic drugs in the fight against cancer (294).

Here, we used single molecule experiments to resolve some of the standing mechanistic questions by deconvoluting DNA bending, protein disorder-to-order transitioning, active-site assembly and incision. We further probed FEN1 high selectivity for its cognate substrate by perturbing key recognition features of the DNA or protein. We thus showed that FEN1 actively bends its substrate DNA with diffusion-limited kinetics in an induced-fit mechanism, undergoes disorder-to-order transitioning while recognizing its 3’ and 5’ flap features, catalyzes its incision reaction with high efficiency and selectivity instantaneously releasing the 5’ flap product while releasing its nicked product from an unbent form in two-steps.

3.3 Results

3.3.1 FEN1 actively bends the DNA

A major question in structure-specific recognition is whether the DNA distortion observed in protein–DNA complexes occurs spontaneously and is therefore captured by the protein in a process termed conformational sampling or if the protein actively sculptures the DNA into the distorted conformation (Figure 3.1C). To test which of these mechanisms best
describes FEN1 substrate recognition, we started by establishing the DNA conformational state of an ideal non-equilibrated (Non-EQ) DF substrate containing 6 nt ssDNA 5’ flap and an intentional unpaired 1 nt ssDNA 3’ flap (Non-EQ DF-6,1) on its own or in the presence of FEN1. The DNA conformational state in both cases was inferred from the FRET efficiency between an Alexa Fluor 647 acceptor placed at 12 nt into the upstream dsDNA and a Cy3 donor at the 5’ flap end (Non-EQ DF-6,1Flap) (Figure 3.2A) or 15 nt into the downstream dsDNA (Non-EQ DF-6,1Internal) (Figure 3.2A). With the flap-labeling scheme, a decrease in FRET is expected upon FEN1 binding and bending its substrate, in contrast with the internal-labeling scheme where an increase in FRET is predicted. The experiments were performed using custom-built setups operating in either the total internal reflection fluorescence (TIRF) mode at a standard temporal resolution of 100 ms (234) as the primary method or the confocal mode for higher temporal resolution.

The single-molecule time traces of the substrate alone in TIRF mode showed a single FRET state with no transition from this state (Figure 3.2B). To test for short-lived alternative conformers, we used confocal-based smFRET to increase the temporal resolution to 5 ms on surface-immobilized DNA (Figure 3.2E) and to sub-ms on freely diffusing DNA in solution (Figure 3.2F). Importantly, we found that the substrate exhibited a single FRET state implying that it existed as a single conformer. In fact, adding FEN1 to DNA Non-EQ DF-6,1 in both labeling schemes showed transitions to the bent states in a single step to form a stable FEN1–DNA_bent complex that rarely dissociated during our 60 s standard acquisition time (Figure 3.2B,C). We calculated the DNA bending dissociation constant \( K_{d\text{-bending}} \) from the FRET efficiency histogram-binding isotherm to be 3.9±0.4
nM and 4.6±0.6 nM for Non-EQ DF-6,1_{Flap} and Non-EQ DF-6,1_{Internal}, respectively. This dissociation constant agreed with the nM range of $K_m$ from bulk cleavage assays (241).

**Figure 3.2. FEN1 actively bends its DF substrate.** (A) Schematic of smFRET assay with two labeling schemes used to study FEN1 bending. The left panel shows the flap-labeling scheme with Cy3 donor at the
end of 5′ flap and Alexa Fluor 647 acceptor in the upstream primer of the DF substrate; the FRET ratio decreases upon FEN1 binding and bending. The right panel shows the internal-labeling scheme where both donor and acceptor are in the template strand; the FRET ratio increases upon FEN1 binding and bending. Monitoring FEN1 bending of non-equilibrated DF-6,1 using the flap-labeling scheme (Non-EQ DF-6,1Flap) (B) and internal labeling-scheme (Non-EQ DF-6,1Internal) (C). For each labeling-scheme, the smFRET time traces of the substrate alone (upper panel) and in presence of FEN1 (middle panel), as well as the binding isotherms of percentage of bent substrate (%) versus FEN1 concentration (nM) (lower panel) are shown; the binding isotherms are fitted with hyperbolic function to yield the reported $K_{d-bending}$ with standard error of the fit reported. (D) Bending of equilibrated DF-6,1 (EQ DF-6,1Internal) by FEN1. smFRET time traces of EQ DF-6,1Internal alone (upper panel-left) and in the presence of FEN1 (middle panel), FRET histogram of EQ DF-6,1Internal alone (upper panel-right), and analysis of its DNA bending association rate constant ($k_{on-bending}$) and dissociation rate constant ($k_{off-unbending}$) (lower panel) are shown. $k_{bending}$ and $k_{unbending}$ were calculated by fitting an exponential function to the histogram from the population of dwell times of bent ($\tau_{bending}$) and unbent ($\tau_{unbending}$) conformers, respectively; error bars correspond to the standard deviation of the fit. $k_{on-bending}$ and $k_{off-unbending}$ are calculated from the slope of $k_{bending}$ from a linear regression fit and the mean of $k_{unbending}$, respectively; the error bars correspond to the standard deviation of the fit. $K_{d-bending} = k_{off-unbending}/k_{on-bending}$. (E) Confocal-smFRET time traces of surface-immobilized Non-EQ DF6,1Flap (left panel) and Non-EQ DF6,1Internal (right panel) with 5 ms temporal resolution. (F) Burst confocal-smFRET histograms from freely diffusing DNA in solution of Non-EQ DF6,1Internal (0.5 nM) (left panel) and Non-EQ DF6,1Flap (0.5 nM) (right panel) acquired at sub-ms temporal resolution.

To mimic the in vivo substrate, we used equilibrated (EQ) DF-6,1 (Figure 3.1A). FEN1 actively bent EQ DF-6,1Internal to a similar extent and similar $K_{d-bending}$ as Non-EQ DF-6,1Internal (Figure 3.2D). Nonetheless, the single molecule time traces showed multiple transitions between bent and unbent states (Figure 3.2D). The reduced stability of the bent conformer in the equilibrated substrate suggests that a bound 3′ flap could dissociate from the 3′ flap-binding pocket. The dissociated 3′ flap in the equilibrated substrate would pair with the template strand before FEN1 could rebind it while in the non-equilibrated substrate it would remain available for rebinding FEN1. Dwell time analysis of the bent ($\tau_{bending}$)
and unbent ($T_{\text{unbending}}$) states at increasing FEN1 concentrations indicated that the apparent first-order rate constant for DNA bending ($k_{\text{bending}} = 1/T_{\text{bending}}$) increased linearly while that for DNA unbending ($k_{\text{unbending}} = 1/T_{\text{unbending}}$) remained constant (Figure 3.2D). This trend is expected for a 1:1 binding equilibrium where $k_{\text{bending}}$ and $k_{\text{unbending}}$ correspond to association and dissociation of FEN1, respectively. Notably, the second-order association rate constant ($k_{\text{on-bending}}$) calculated from the slope of the linear fit of the concentration dependence of $k_{\text{bending}}$ was diffusion-limited ($1.4 \pm 0.03 \times 10^8$ M$^{-1}$ s$^{-1}$), and the average value of $k_{\text{unbending}}$ ($k_{\text{off-unbending}}$) was $0.45 \pm 0.05$ s$^{-1}$ (Figure 3.2D). Overall, we concluded, with our temporal resolution, that FEN1 actively bends its double flap substrate with diffusion-limited kinetics. We further confirmed this active bending with another human 5’ nuclease, EXO1 (228).

### 3.3.2 FEN1 never misses cleavage of its correct substrate

To examine the active-site assembly with respect to DNA bending, we replaced Ca$^{2+}$ with Mg$^{2+}$ to simultaneously monitor DNA bending and 5’ flap cleavage using the flap-labeling scheme (Figure 3.3A). Time traces indicated that FEN1 always bent Non-EQ DF-6,1$^\text{Flap}$ (Figure 3.3B) and EQ DF-6,1$^\text{Flap}$ (Figure 3.3C) before cleaving the Cy3-containing 5’ flap; remarkably almost every DNA bending event led to a successful cleavage reaction (Figure 3.3B,C). We confirmed DNA bending before cleavage by the clear anti-correlated change in the donor and acceptor intensities (Figure 3.3B,C). Direct comparison of donor fluorescence in the presence of FEN1 and either Mg$^{2+}$ or Ca$^{2+}$ ions indicated that there is a
strong correlation between the loss of donor particles and the presence of Mg$^{2+}$ that coincided with the introduction of FEN1 into the flow cell (Figure 3.3D). This confirms that the loss of donor particles is due to 5’ flap cleavage and not due to donor photobleaching. We further characterized each particle behavior and discarded those that showed aberrant intensities, donor or acceptor photobleaching, or no cleavage (Figure 3.3D). Analysis of FRET values before cleavage from individual time traces showed that FEN1 cleaved Non-EQ DF-6,1$^{\text{Flap}}$ from a fully bent state (Figure 3.3E). The distributions of the time spent in the bent state prior to cleavage fitted to gamma distributions for both Non-EQ DF-6,1$^{\text{Flap}}$ and EQ DF-6,1$^{\text{Flap}}$ averaged around 160±7 ms and 155±30 ms, respectively (Figure 3.3B,C). FEN1 cleavage generates two products; 5’flap ssDNA and nicked dsDNA (Figure 3.1A). Previous studies demonstrated that excess nicked dsDNA but not 5’ flap ssDNA influences FEN1 activity, which suggests that only nicked dsDNA is a competitive inhibitor of FEN1 release (241,295). Consistent with these findings, we also observed that the time spent in the bent state before cleavage is not influenced by the presence of excess 5’flap ssDNA (Figure 3.3F). Therefore, single-turnover rates ($k_{\text{STO}}$) could be determined directly from the time spent in bent state prior to cleavage ($k_{\text{STO}} = 1/\tau_{\text{bending-flap}}$). However, since 5’ flap release would still contribute to the dwell time before cleavage, our single turnover should be treated as an apparent value. Biologically relevant, the cleavage behavior from the first DNA bending and the $k_{\text{STO}}$ were similar whether there was a deliberate mismatch 3’ flap (Non-EQ DF-6,1; Figure 3.3B) or an equilibrating 3’ flap in EQ DF-6,1 (Figure 3.3C), consistent with bulk cleavage reactions (295). The diffusion-limited rates of DNA bending and cleavage before protein dissociation provide
direct evidence that the reaction of FEN1 on a cognate substrate is limited by encounters between the enzyme and the substrate.

Moreover, the distributions of $T_{bending-flap}$ for both Non-EQ DF-6,1 and EQ DF-6,1 show a rise and decay behavior (Figure 3.3B,C), suggesting that the underlying catalytic mechanism after the diffusion-limited DNA bending step involves two or more steps, as a single-step process would have resulted in a single exponential decay. We reasoned that these steps likely include 3’ flap-induced disorder-to-order transitioning and cleavage chemistry. To test this hypothesis, we employed glycerol as low-molecular-weight viscogen to slow any local protein conformational change that mediates catalysis and/or product release. Increasing glycerol concentration decreased $k_{STO}$ linearly with a slope of 1.5±0.2 (Figure 3.3G) but remained unaffected by polyethylene glycol-8000, a high-molecular-weight viscogen that is too large to interfere with local protein conformational changes (data not shown) (228). $k_{STO}$ is not influenced by 5’ flap ssDNA product release, suggesting that 3’ flap-induced protein ordering is a terminal step to verify the substrate before incision.
Figure 3.3. Cleavage of cognate substrate by FEN1. (A) Schematic representing smFRET cleavage assay using the flap-scheme. (B) Simultaneous monitoring of FEN1 DNA bending and 5' flap cleavage of Non-EQ DF-6,1_{Flap}. Left: A representative smFRET time trace (upper panel) with a zoomed-in view showing the
cleavage of Non-EQ DF-6,1\textsubscript{Flap} in which FEN1 almost never misses the opportunity to bend the DNA and cleave it (lower panel). Right: distribution of dwell times of the bent state prior to cleavage of Non-EQ DF-6,1\textsubscript{Flap} ($\tau$\textsubscript{bending-flap}) for N=227 cleavage events fitted with a gamma distribution. The mean and the standard error of the mean are reported. $k_{STO} = 1/\text{Avg } \tau$\textsubscript{bending-flap}. Cleavage was performed at 50 ms temporal resolution. (C) smFRET cleavage of EQ DF-6,1\textsubscript{Flap} as described for B shows comparable $k_{STO}$ to that of Non-EQ DF-6,1\textsubscript{Flap}. (D) Left: bar chart comparing the Cy3 donor signal lost in cognate substrate due to photobleaching (+Ca\textsuperscript{2+}) or to both photobleaching and incision (+Mg\textsuperscript{2+}) in presence of FEN1. The donor loss increases significantly in the presence of Mg\textsuperscript{2+} as compared to what would be expected for its loss in the presence of Ca\textsuperscript{2+}. Right: detailed classification of all single-molecule time traces in the field of view for cognate substrate in the presence of Mg\textsuperscript{2+}. The uncertainty corresponds to the standard deviation between multiple movies in the presence of either CaCl\textsubscript{2} or MgCl\textsubscript{2}. (E) FRET efficiency of the bent state before cleavage of Non-EQ DF-6,1\textsubscript{Flap} fitted with a Gaussian distribution from multiple cleavage events. (F) Distribution of $\tau$\textsubscript{bending-flap} of N=227 FEN1 cleavage events of Non-EQ DF6,1\textsubscript{Flap} in the presence of excess 10 nt ssDNA as a competitor. The distribution was fitted as in B and the mean and standard error of the mean are reported. This distribution shows no effect of ssDNA on $k_{STO}$ when compared to B. (G) Effect of low molecular weight viscogen (glycerol) on disorder-to-order transitioning in FEN1. Graph showing relative $k_{STO}$ of Non-EQ DF6,1\textsubscript{Flap} cleavage upon addition of glycerol at increasing relative viscosity fitted with a linear regression to calculate the slope of the curve; the error corresponds to the standard error of the fit. $k_{STO}$ was determined as in B.

### 3.3.3 A comprehensive kinetic scheme of FEN1 cleavage reaction

Our next goal was to build a comprehensive kinetic model of FEN1 reaction on DF-6,1 from substrate bending to product release. The diffusion-limited association rate (Figure 3.2D) and the productive catalysis from the first bending event (Figure 3.3B,C) demonstrated that FEN1 cleavage is not limited by substrate binding/bending. Since the previously reported steady state cleavage kinetics ($k_{cat}$) is significantly slower than the single turnover kinetics ($k_{STO}$) (241,295), FEN1 is widely accepted to be product-inhibited. Therefore, 5' flap release is fast following cleavage whereas nicked duplex release limits
the $k_{\text{cat}}$. We have also demonstrated that 5’ flap release is not rate-limiting (Figure 3.3F). Our rapid quench-flow and stopped-flow bulk cleavage experiments, consistent with previous reports (151,296), determined the $k_{\text{STO}}$ at 21±0.9 s$^{-1}$ and $k_{\text{cat}}$ at 1.4±0.1 s$^{-1}$ (Figure 3.4A) with a 15-fold difference. We attributed the ~3-fold slower $k_{\text{STO}}$ in single molecule experiments (Figure 3.3C) to the difference in temperature; 22 °C in single molecule versus 37 °C in bulk.

The cleavage kinetics with donor-labeled 5’ flap DNA were determined by monitoring the bending step just before loss of the donor signal (Figure 3.3A). Thus, these kinetics report on DNA bending, protein ordering for active site assembly, chemistry, and flap release. However, the flap-labeling scheme offers no information about the nicked DNA product release. Therefore, we used a complementary assay based on the internal-labeling scheme to monitor steps subsequent to flap release (Figure 3.4B). This assay detects the time spent by the substrate in bent state at a high FRET of E ~0.52 before the signal decreases to E ~0.25, which we interpret as the nicked product in an unbent state (Figure 3.4B). This interpretation is supported by the following information. The unbent EQ DF-6,1$_{\text{Internal}}$ substrate exhibits a difference of E ~0.05 from the unbent nicked product, as evident from the DNA-only histograms of EQ DF-6,1$_{\text{Internal}}$ (Figure 3.2D) and nicked DNA (Figure 3.4C). Thus, if DNA bending is not followed by 5’ flap cleavage, the FRET should change from 0.3 (unbent substrate) to 0.51 (bent substrate) and back to 0.3 (unbent substrate), whereas in the case of DNA bending followed by 5’ flap cleavage, the FRET would change from 0.3 (unbent substrate) to 0.51 (bent substrate) and back to 0.25 (unbent nicked product). Importantly, the experiment was designed to minimize FEN1 rebinding and bending the nicked product while allowing substrate binding/bending. Specifically, the
cleavage reaction was performed at 250 nM FEN1, which is ~50-fold higher than $K_{d}$-bending of DF-6,1$_{\text{Internal}}$ (3.2 nM; Figure 3.2D) but well below $K_{d}$-bending of the nicked product (lower estimate of 580 nM; Figure 3.4C). Furthermore, since FEN1 always cleaves DF-6,1 after the first encounter (Figure 3.3C), this FEN1 concentration ensures a high fraction of cleavage within a single turnover, and no signal from nicked product rebinding and bending after it is released.

Figure 3.4. Internal-labeling scheme infers on FEN1 product release. (A) Ensemble cleavage kinetics of FEN1 on EQ-DF6,1. Left: Single turnover cleavage was measured on a rapid quench-flow instrument at a FEN1:DNA ratio of 35:1. The amount of 5' flap product formed was analyzed by denaturing PAGE. Average product concentration from two replicates was plotted versus time and fitted to a single exponential equation to determine the cleavage rate ($k_{\text{STO}}$). Right: steady state cleavage was measured with FEN1:DNA at a ratio of 1:800. Average data from two replicates fitted to a linear regression yielded $k_{\text{cat}}$ (slope/[FEN1]). Error bars
correspond to the variation of the two replicates, and the error of the fit is reported. (B) Internal-labeling smFRET cleavage assay. Top: schematic of the assay. EQ DF-6,1\textsubscript{Internal} is labeled as described in Figure 3.2A. In the presence of Mg\textsuperscript{2+}, FEN1 binds and bends EQ DF-6,1\textsubscript{Internal}, increasing FRET from 0.3 to 0.52. Upon cleavage, the 5’ flap is released and a nicked duplex is generated, which has a FRET of 0.25 when unbent. The assay follows the time spent by DNA in bent state (0.52) before the product achieves unbent state. Bottom left: a representative single molecule time trace showing FEN1 bending and cleaving the substrate before FRET drops to 0.25; the inset zooms in on a vbFRET-fitted version of the cleavage event showing a three-state fit (0.3, 0.52 and 0.25) corresponding to the three DNA conformers, unbent EQ DF-6,1\textsubscript{Internal}, bent EQ DF-6,1\textsubscript{Internal} and unbent nicked product, respectively. Bottom right: distribution of the dwell times spent in bent state (τ\textsubscript{bending-internal}) for N=64 cleavage events fitted to a gamma distribution. The average τ\textsubscript{bending-internal} is reported with the standard error of the mean. The cleavage reaction was performed at 50 ms temporal resolution. (C) smFRET bending efficiency of FEN1 on Nick\textsubscript{Internal}. Left panel shows smFRET histograms of Nick\textsubscript{Internal} DNA-only (top) with a single peak centered ~0.25 and upon addition of 200 nM FEN1 (bottom) with two peaks. The two peaks are merged and the centers are shifted. If the unbent peak (highlighted in magenta) center is fixed at 0.25, then the bent peak (highlighted in blue) appears to be centered ~0.41. Right top: a representative single molecule time trace of Nick\textsubscript{Internal} upon addition of 500 nM FEN1. The time trace shows fast transitions between unbent (0.25) and bent (0.52) states. The transitions appear to be much faster than the acquisition temporal resolution of 100 ms. Right bottom: an isotherm of the percentage of bent substrate (%) versus FEN1 concentration (nM) fitted to as described for Figure 3.2B showing a lower estimate of FEN1 $K_{d-bending}$ of Nick\textsubscript{Internal} due to the averaging effect. This averaging effect is caused by the fast transitions seen in the time traces and, consequently, merging and shifting of the peak centers in histograms, which complicates fitting and estimation of the percentage of bent substrate at each concentration, and thus, the explicit determination of $K_{d-bending}$.

The data show that the time spent by the internal-labeled substrate in bent state ($\tau$\textsubscript{bending-internal}) is 270±70 ms (Figure 3.4B), while that of the flap-labeled substrate ($\tau$\textsubscript{bending-flap}) is 155±30 ms (Figure 3.3C). The difference between $\tau$\textsubscript{bending-flap} and $\tau$\textsubscript{bending-internal}, which reports the dwell time of the nicked product in bent state ($\tau$\textsubscript{product-bent}), is only 115±75 ms. This value yields an apparent rate of 8.7±5.7 s\textsuperscript{-1} (1/$\tau$\textsubscript{product-bent}) for nicked product release if we assume that FEN1 dissociation is coupled with unbending of the DNA. This rate is comparable to the $k_{STO}$ of 6.5±1.2 s\textsuperscript{-1} (Figure 3.3C), and is much higher than the $k_{cat}$
measured in bulk under similar reaction conditions (1.4±0.1 s⁻¹; Figure 3.4A), which would imply that another step after nicked DNA release limits steady state turnover. However, as noted above, a product inhibition study predicted that release of nicked DNA is the rate-limiting step (241).

In order to reconcile these contradictory findings, we considered the possibility that FEN1 remains bound to the unbent nicked product for some time before dissociating into solution. In order to test this hypothesis, we performed the same experiment under conditions that favor FEN1 binding to nicked DNA. The prediction was that if after cleavage FEN1 remains bound to the unbent nicked product for some time, we would observe a lag phase at low FRET of E ~0.25, which reports the dwell time of this complex before it dissociates and another FEN1 rapidly binds and bends the released product to increase the signal to E ~0.52 (as illustrated in Figure 3.5A). Lowering KCl concentration to 40 mM from 100 mM increases the affinity of FEN1 for the DNA products of both EQ DF-6,1 and Non-EQ DF-6,1 (K_d = 62 nM and 12 nM, respectively; Figure 3.5E,F). A control experiment with flap-labeled EQ DF-6,1 showed that k_{STO} is not affected by lowering KCl concentration (τ_{bending-flap} is 155±30 ms at 100 mM KCl and 180±40 ms at 40mM KCl; Figure 3.3C and Figure 3.5C, respectively). Since FEN1 bends (Figure 3.2C,D) and cleaves EQ and Non-EQ substrates (Figure 3.3B,C) with similar K_{d-bending} and k_{STO}, respectively (228,295), Non-EQ DF-6,1/Internal was used as the substrate in this experiment to ensure rapid and high affinity rebinding of FEN1 to the nicked product. The results show an increase in τ_{bending-internal} from 270±70 ms at 100 mM KCl (Figure 3.4B) to 570±115 ms at 40 mM KCl (Figure 3.5B), indicating that the dwell time of bent product, τ_{product-bent}, has been extended from 115±75 ms to 390±120 ms with lower KCl. Notably,
we observed an extended phase with unbent DNA at E ~0.25 that lasted for 2100±420 ms ($\tau_{\text{product-unbent}}$), before FEN1 dissociation followed by fast rebending re-established the bent state at E ~0.52 (Figure 3.5B). These results could be interpreted such that product release by FEN1 occurs in two steps: $\tau_{\text{product-bent}}$ wherein FEN1 briefly holds the product in bent state for 390±120 ms after 5' flap departure and $\tau_{\text{product-unbent}}$ wherein FEN1 remains bound to the unbent product for 2100±420 ms before dissociating into solution. Therefore, the actual $\tau_{\text{release}}$ could be the sum of the two dwell times, which yields a $k_{\text{release}}$ of 0.40±0.07 s$^{-1}$ for the nicked product. The resulting $k_{\text{cat}}$ of 0.37±0.06 s$^{-1}$ ($1/(\tau_{\text{bending-flap}} + \tau_{\text{release}})$) is in line with rates determined by bulk experiments at 1:800 of FEN1:DNA (Figure 3.5D and Figure 3.4A at 40 and 100 mM KCl, respectively) and previous reports (241,295). Taken together, these results suggest that FEN1 turnover is limited by release of the nicked DNA product, mainly from an unbent state.
**Figure 3.5. Building a kinetic scheme of FEN1 cognate substrate recognition and catalysis.** (A) Schematic of the multi-step FEN1 reaction as revealed by the internal-labeling scheme: (1) Substrate bending: The extended DF substrate exhibits FRET of 0.3, which increases to 0.52 upon FEN1 binding and bending. (2) Incision/5’ flap release: FEN1 incises the 5’ flap 1 nt inside the junction, and the flap is instantaneously released; at this stage the FRET is 0.52. The time spent by the substrate in bent state just before 5’ flap release is accessed by the flap-labeling cleavage assay ($\tau_{\text{bending-flap}}$). (3) Product unbending: The nicked duplex product remains bound and bent by FEN1 at a FRET of 0.52 (similar to that of bent substrate) before unbending to FRET ~0.25. The time spent by the product in bent state ($\tau_{\text{product-bent}}$) is the difference between the bent state dwell times in the internal-labeling ($\tau_{\text{bending-internal}}$) and flap-labeling ($\tau_{\text{bending-flap}}$) assays. (4) FEN1 dissociation: FEN1 remains bound to the unbent product (E ~0.25) for some time ($\tau_{\text{product-unbent}}$) before dissociating into solution. Thus, product release occurs in two steps and $\tau_{\text{release}}$ is the sum of $\tau_{\text{product-bent}}$ and $\tau_{\text{product-unbent}}$. (5) Product bending: FEN1 can rebind/rebend the product again; hence, FRET fluctuates between 0.25 (unbent product) and 0.52 (bent product) at the end of the reaction. The rebinding step is detected by lowering KCl from 100 mM to 40 mM to increase FEN1 affinity for nicked DNA. (B) smFRET cleavage of Non-EQ DF-6,1_internal. Top: representative single molecule time trace showing cleavage of Non-EQ DF-6,1_internal and exhibiting the substrate and product dynamics described in (A). The FRET state and the substrate/product conformer in each step is illustrated. $\tau_{\text{bending-internal}}$ is highlighted in red and $\tau_{\text{product-unbent}}$ is highlighted in blue on the time trace. The distributions of $\tau_{\text{bending-internal}}$ (bottom left) and $\tau_{\text{product-unbent}}$ (bottom right) for N=99 cleavage events were fitted to gamma distributions, and the means with standard errors are reported. The cleavage reaction was performed at 100 ms temporal resolution. (C) smFRET cleavage of EQ DF-6,1_flap by FEN1 at 40 mM KCl. Distribution of the dwell times spent in the bent state ($\tau_{\text{bending-flap}}$) for N=72 cleavage events. The average $\tau_{\text{bending-flap}}$ at 40 mM KCl is comparable to that obtained at 100 mM KCl (Figure 3.3B,C). The cleavage reaction was performed at 50 ms temporal resolution. (D) Bulk cleavage kinetics of EQ-DF6,1 by FEN1 at 40 mM KCl. Single turnover (left) and steady state (right) kinetics were measured and fit as described in Figure 3.4A, yielding $k_{\text{sto}}$ and $k_{\text{cat}}$, respectively. $k_{\text{sto}}$ is slightly faster and $k_{\text{cat}}$ is slightly slower than the rates obtained at 100 mM KCl (Figure 3.4A). (E) smFRET bending kinetics of Nick_internal by FEN1 at 40 mM KCl in the presence of Ca^{2+}. Top: a schematic showing the Nick_internal structure. Left: smFRET histograms of Nick_internal alone and upon addition of 500 nM FEN1. Unbent DNA-alone peak (shown in magenta) is centered ~0.24 and bent peak (in blue) is centered ~0.5. Right Top: a representative single molecule time trace of Nick_internal upon addition of 50 nM FEN1. The trace shows transitions between unbent (0.25) and bent (0.52) states. Right Bottom: a graph of $k_{\text{bending}}$ (s^{-1}) and $k_{\text{unbending}}$ (s^{-1}) versus FEN1 concentrations (nM). The values of $k_{\text{bending}}$, $k_{\text{unbending}}$, $k_{\text{off-unbending}}$, $k_{\text{on-bending}}$ and $K_d$-bending are determined as described for Figure 3.2D. (F) smFRET bending kinetics of Product_internal by FEN1 in the presence of Ca^{2+}. Top: a schematic showing the Product_internal structure with a static non-equilibrated 1 nt 3’ flap mimicking the cleavage product of a Non-EQ DF substrate. Left: bending kinetics of Product_internal at 40 mM KCl, with the middle panel showing a single molecule time trace of Product_internal upon addition of 20 nM FEN1 exhibiting dynamic behavior, and the bottom panel showing the association and dissociation rate constants...
as described in (E). Right: bending kinetics of Product\textsubscript{Internal} at 100 mM KCl. At higher salt, Product\textsubscript{Internal} displays faster transitions with a significantly elevated $k_{\text{off-unbending}}$, and hence $K_d$-bending. However, $k_{\text{on-bending}}$ is not influenced by salt concentration.

However, the smFRET data described above does not provide direct evidence that FEN1 is still bound to the unbent product during the lag phase. For more direct evidence, we employed protein-induced fluorescence enhancement (PIFE), in which the fluorescence of certain fluorophores (mainly cyanine dyes) is enhanced upon protein binding rather than induced conformational changes. We placed Cy3 at position 10 on the downstream duplex DNA (Figure 3.6A) and observed 35% enhancement of Cy3 fluorescence with Non-EQ DF-6,1\textsubscript{PIFE} in the presence of FEN1 using time-resolved fluorescence measurements in the presence of Ca\textsuperscript{2+} (Figure 3.6A). smPIFE experiments were performed under standard smFRET cleavage conditions (100 mM KCl, 250 nM FEN1). In this assay, the time spent in the enhanced-fluorescence state ($\tau_{\text{PIFE}}$) is interpreted as time spent by FEN1 binding to the substrate, cleaving the flap, and any subsequent binding to the nicked product. By comparing the kinetics from the smFRET and smPIFE cleavage reaction, we hypothesize that if FEN1 releases the product from a bent state right after cleavage, $\tau_{\text{PIFE}}$ should be relatively short while a longer-lived enhanced-fluorescence state would imply that FEN1 remains bound to the product after unbending occurs. The smPIFE assay was performed with DF-6,1, which is cleaved by FEN1 almost always in the first encounter (Figure 3.3B), and at 100 mM KCl and 250 nM FEN1 when no product rebending is observed, as shown in Figure 3.4B and discussed above. These conditions support our interpretation that the observed enhanced-fluorescence state stems from a productive binding event that leads to
cleavage. Fitting the $\tau_{\text{PIFE}}$ for N=77 cleavage events with a gamma distribution yields a lengthy average $\tau_{\text{PIFE}}$ of 2210±500 ms (Figure 3.6B) which translates to a rate of 0.45±0.10 s$^{-1}$. This rate is in line with the $k_{\text{cat}}$ we observed in bulk as well as the suggested $k_{\text{cat}}$ in smFRET experiments. Taken together, smFRET and smPIFE assays potentially explain FEN1 product release mechanism with two steps, a fast unbending step and a relatively slow release of product after it achieves an unbent state.

To summarize, FEN1 binds and actively bends DF-6,1 with diffusion-limited kinetics, cleaves the 5’ flap rapidly after protein ordering and active site assembly, followed by slow product release that occurs in two steps, highlighting the relatively high affinity and stability of the FEN1-nicked product complex.
Figure 3.6. smPIFE cleavage confirms FEN1 product release from an unbent state. (A) Top: schematic showing the position of Cy3 fluorophore in Non-EQ DF-6,1_{PIFE}. Bottom: bulk time-resolved fluorescence lifetime measurements of Non-EQ DF-6,1_{PIFE} in the absence (black curve) and presence (red curve) of 1 μM FEN1. Inset shows the quantification of fluorescence lifetime in the absence and presence of FEN1. The lifetimes are determined using a 2-exponential decay fit, and show 35% fluorescence enhancement upon FEN1 binding. (B) Representative time trace showing a smPIFE cleavage experiment with Non-EQ DF-6,1_{PIFE}. The substrate/product conformer in each state is illustrated. The time spent in the enhanced-fluorescence state $\tau_{PIFE}$ is highlighted in green. The distribution of $\tau_{PIFE}$ for $N=77$ cleavage events was fitted to a gamma distribution, with the mean and standard error of the mean reported. The cleavage reaction was performed at 100 ms temporal resolution.

3.3.4 FEN1 verifies its substrate to avoid off-target DNA cleavage in the DNA lockdown step

To further interrogate FEN1 sophisticated recognition mechanism and to investigate its high selectivity towards its cognate substrate, we perturbed key structural features and studied their effect on FEN1 bending and catalytic efficiency. The main key structures of FEN1 cognate substrate are the strict 1 nt 3’ flap and the ssDNA flap. We reasoned that since FEN1 predominantly binds and bends the DNA at the duplex region as evident from the major interactions of protein-substrate in the crystal structure, disturbing the 3’ or 5’ flap structures would influence FEN1 substrate verification.

From bulk measurements, it remains unclear how 3’ flap-induced protein ordering operates in the case of the *in vivo* equilibrated DF substrate. The equilibrated junction may exist as a single 5’ flap that requires active molding by FEN1 into a double 5’- and 3’-flap or as a DF with a readily available 3’ flap for FEN1 capturing. To address this, we started
by investigating the requirement of having a preformed 3’ flap for inducing DNA bending. Removal of the 3’ flap from Non-EQ DF generating a SF substrate decreased FEN1 cleavage activity by 34 fold (241). Time traces on surface-immobilized SF-6,0Flap accessed at 5 ms using confocal-based smFRET of the DNA alone showed a single FRET state E~0.8 signifying that at this temporal resolution, SF-6,0Flap existed in a single conformer (Figure 3.7A). Upon FEN1 addition, the FRET state rapidly transitioned between E~0.8 and E~0.48 demonstrating that FEN1 actively bent the SF substrate (Figure 3.7A). However, the $\tau_{\text{bending}}$ was markedly reduced to ~43 ms ($k_{\text{off-unbending}}=23.3\pm3.8$ s$^{-1}$) in contrast to that of the stable bent conformer in Non-EQ DF-6,1Flap and EQ DF-6,1Internal ($k_{\text{off-unbending}}=0.45\pm0.05$ s$^{-1}$) (Figure 3.2B,D). However, the $k_{\text{on-bending}}$ at $(1.4\pm0.2)\times10^8$ M$^{-1}$s$^{-1}$ (Figure 3.7A) remained limited by diffusion and similar to that of EQ DF-6,1Internal (Figure 3.2D). $K_d$-bending was 50-fold higher than that of Non-EQ DF-6,1 and EQ DF-6,1 (Figure 3.2B,D). These results show that a 3’ flap is not required for DNA bending but it is critical for DNA binding stability.

We next investigated the communication between the 3’ flap-induced protein ordering and the distant gateway with respect to 5’ flap recognition. It has been postulated that the 5’ flap may thread through the cap-helical gateway and that this threading is needed for catalysis (158,276,280) or that the 5’ flap may be clamped away from the active site for catalysis (278) as discussed in the introduction to this chapter. To test the possibility of a threading mechanism and its coordination with the 3’ flap-induced protein ordering, we used a modification that prevented 5’ flap threading. Blocking the threading by immobilizing a DF substrate through a biotin attached at the end of a 30 nt ssDNA 5’ flap (termed DF-30,1blocked-Internal) showed unstable bending with rapid transitioning (Figure
3.7B), whereas trapping the threaded conformer by prior incubation with FEN1 before blocking the 5’ flap with biotin-NeutrAvidin reestablishes stable bending (data not shown). The impaired DNA bending due to blocked threading (Figure 3.7D,E) displayed comparable $K_{d\text{-}bending}$ and $k_{off\text{-}unbending}$ of SF-6,0 (Figure 3.7A). Notably, blocking the threading led to a distorted DNA structure that did not reach the same final FRET state as when the 5’ flap was not blocked (Figure 3.7C). Importantly, these results indicate that the initial DNA bending by FEN1 did not require threading, but that full bending required 5’ flaps, if present, to be able to thread. The increased $K_{d\text{-}bending}$ of the unthreaded substrate upon removal of its 3’ flap (SF-30,0blocked-Internal) (Figure 3.7D) indicates that 3’ flap binding did not require 5’ flap threading. However, the ability of the 5’ flap to thread is required for the 3’ flap-induced protein ordering to form the stably and correctly bent DNA conformer.

The combined requirement for 3’ flap and base pairing at the junction suggests that signaling occurs via ordering from the 3’ flap-binding pocket to the distant gateway where the 5’ flap is recognized and cleaved. The superfamily semi-conserved R47 in the hydrophobic wedge is poised to mediate this coordination: it stacks against the first base pair on the 3’ flap side of the junction while its side chain C-caps the $\alpha2$ in the gateway and stacks with K128 on $\alpha5$ in the cap (156) (Figure 3.7F). Mutating R47 to A (FEN1-R47A) impaired FEN1’s cleavage on DF substrates to a similar extent as wild-type (WT)-FEN1’s cleavage on SF-6,0 (156). To test this allosteric signaling idea, we maintained the 3’ flap binding using Non-EQ DF-6,1Internal while altering R47 using FEN1-R47A. The defects in $K_{d\text{-}bending}$ and $k_{off\text{-}unbending}$ (Figure 3.7D,E) resembled those in WT-FEN1 on SF-6,0 (Figure 3.7A). We believe that FEN1-R47A engaged the 3’ flap because its $K_{d\text{-}bending}$
and $k_{\text{off-unbending}}$ increased on SF-6.0$_{\text{Internal}}$ as compared to that on Non-EQ DF-6.1$_{\text{Internal}}$ (Figure 3.7D,E).

Collectively these results demonstrate that FEN1 bends both cognate and noncognate substrates and that $K_{d-bending}$ is higher for noncognate substrates. They further showed that FEN1 stabilizes the cognate substrate through remarkable selectivity for its key features of a fully paired nick junction, a 3’ flap and a 5’ flap while promoting the dissociation of noncognate substrates. These results are consistent with a model in which FEN1 actively bends DNA to interact with the ss/ds-DNA junctions and subsequently verifies these interactions by the 3’ flap-induced protein ordering and full engagement of the 5’ flap. It is worthy to note that other noncognate substrates such as DF-7.1$_{\text{mismatch(1nt)-Flap}}$ and DF containing 2 nt 3’ flap (DF-6.2$_{\text{Flap}}$) exhibited $k_{\text{off-unbending}}$ (Figure 3.7E) that was ~13-15 fold slower than that of SF-6.0 (Figure 3.7A) and ~3-4 fold longer than $k_{\text{STO}}$ of the cognate substrate (Figure 3.3B,C). However, FEN1 still bent these substrates multiple times without cleaving them (228). Therefore, it is likely that FEN1 possesses intrinsic mechanisms that block the probable formation of catalytically competent active sites in noncognate substrates to inhibit off-target incision.
Figure 3.7. FEN1 multi-step vetting process of its substrate. (A) Confocal-based smFRET time traces of surface-immobilized SF6,0Flap alone (left panel) and in the presence of FEN1 (middle panel) acquired at 5 ms temporal resolution, showing rapid transitions from high to low FRET upon DNA bending. $k_{on}$, $k_{bending}$ and $k_{off-unbending}$ of SF6,0Flap by FEN1 (right panel) calculated as in Figure 3.2D. (B) The effect of
blocking 5’ flap threading on DNA bending by FEN1. Top: schematic showing the strategy used to block 5’ flap threading into the cap-helical gateway by introducing NeutrAvidin/biotin linkage at the 5’end of the 5’ flap of DF-30,1 (termed DF-30,1blocked-Internal) prior to the addition of FEN1. Bottom: surface-immobilized confocal-smFRET time traces of DF-30,1blocked-Internal alone (left) and in the presence of FEN1 (right) at 5 ms temporal resolution. The substrate was immobilized by surface-coated NeutrAvidin via the biotin group on the 5’ flap. (C) A bar chart comparing final bent FRET states of DF-30,1trapped-Internal, DF-30,1blocked-Internal, SF-30,0trapped-Internal and SF-30,0blocked-Internal using burst confocal-smFRET histograms from freely diffusing substrates acquired at sub-ms temporal resolution. FEN1 concentrations were 5000 nM for SF-30,0blocked-Internal, 1000 nM for DF-30,1blocked-Internal, 1000 nM for SF-30,0trapped-Internal and 200 nM for DF-30,1trapped-Internal. To trap a threaded 5’ flap, FEN1 was first pre-incubated with the substrate before NeutrAvidin was added to bind the biotin on the 5’ flap (as shown in the schematic in the upper panel). The FRET value in each condition represents the average of N = 3 and the uncertainty corresponds to the standard error of their fits. (D) Bar chart comparing $K_{d}$-bending for FEN1-WT or FEN1-R47A on various non-equilibrating flap substrates using the internal labeling scheme. Used noncognate substrates include SF-6,0, DF containing 1 nt mismatch at the nick junction (DF-7,1mismatch(1nt)), DF containing NeutrAvidin/biotin at tip of the 5’ flap to block 5’ flap threading (DF-30,1blocked) and its SF version (SF-30,0blocked), and DF containing 2 nt 3’ flap (DF-6,2). (E) Bar chart comparing $k_{off}$-unbending for FEN1-WT or FEN1-R47A on various non-equilibrating flap substrates using the internal labeling scheme. The lower estimate of $k_{off}$-unbending for FEN1-WT on DF-6,1 corresponds to the 60 s acquisition time where transitions were rarely detected. $K_{d}$-bending and $k_{off}$-unbending are calculated as in Figure 3.2B and Figure 3.2D, respectively. $k_{off}$-unbending was determined from multiple FEN1 concentrations except for FEN1-R47A on SF-6,0 and FEN1 on DF-7,1mismatch(1nt), which were determined from two and one concentration, respectively. (F) R47 acts as a sensor that couples structuring of the 3’ flap-binding pocket and the cap-helical gateway. R47 in the hydrophobic wedge mediates multiple interactions, where it stacks against the first base pair on the 3’ flap side of the junction while its side chain C-cap the α2 in the gateway (highlighted in green) and stacks with K128 on α5 in the cap (highlighted in purple) (PDB 3Q8L).

We reasoned that there are two possible mechanisms for controlling the incision reaction, which are experimentally feasible. The 3’ flap-induced protein ordering could act once per DNA bending event, locking the DNA into either a catalytically competent or incompetent conformation. In this mechanism, the $k_{STO}$ after DNA bending should be similar between cognate and noncognate substrates regardless of whether or not the
lifetime of the bent conformer is limiting. Alternatively, the protein could lock the DNA into a bent conformer and go through multiple cycles of disorder-to-order transitioning to search for a catalytically competent conformation of protein and DNA. In this mechanism, the $k_{STO}$ would be slower for noncognate substrates, particularly under conditions when the lifetime of the bent conformer exceeded that required for cleavage. We found that the $k_{STO}$ of FEN1 in all tested noncognate substrates was similar and comparable to that in the cognate substrate (data not shown) (228). Therefore, we concluded that the 3’ flap-induced protein ordering locked the DNA into either a catalytically competent conformation to be immediately incised or into an incompetent conformation that led to immediate DNA release from the bent conformation. Directly observing FEN1 conformation will lead to further understanding of how it prevents cleavage in noncognate substrates.
Chapter 4

4. Missed cleavage opportunities by FEN1 lead to Okazaki fragment maturation via the long-flap pathway

4.1 Abstract

To ensure genomic stability, low fidelity RNA/DNA hybrid primers synthesized by Pol α-primase complex are removed and OFS are ligated by a highly coordinated mechanism called OF maturation. In this process, the primer is displaced as a 5’ flap for its cleavage by structure-specific 5’ nucleases, mainly FEN1. At least two models (short- and long-flap pathways) have been suggested to describe how the primer is removed. Here, we focused on the mechanism of FEN1 substrate recognition and catalytic efficiency on longer flaps and how would that trigger the long-flap pathway. Using single molecule FRET and stopped-flow ensemble cleavage assays we showed that, unlike in the short flaps case, where FEN1 cleaves its substrate from first DNA bending encounter, on longer flaps FEN1 occasionally bends the DNA without cleaving it. Moreover, in the presence RPA, FEN1 can access its short- and long-flap substrate, but its DNA bending and cleavage efficiencies are severely inhibited in long-flap substrates. We propose that missed cleavage in long-flap substrates allows RPA binding to the 5’ flap and necessitates the long-flap pathway.
4.2 Introduction

Genomic instability has been implicated in a plethora of diseases. DNA is under constant attack by endogenous as well as exogenous factors. The endogenous factors are in part produced during DNA replication process (297). Genomic stability necessitates the coordinated action of multiple DNA metabolic pathways including DNA replication, DNA repair, transcription and cell cycle progression (298,299). Hence, these DNA processes are tightly coordinated, coupled and regulated to maintain genome stability (300). One of the areas where this regulation is of utmost significance is the OF maturation (249).

Due to the antiparallel nature of DNA and the strict 5’-3’ directionality of replicative DNA polymerases, DNA replication is semi-discontinuous, where the lagging strand is synthesized in short OFs (~200 bp for eukaryotic systems). Most DNA polymerases cannot synthesize DNA de novo requiring an RNA primer to initiate synthesis. This initiator RNA primer (iRNA) is synthesized by specialized DNA-dependent RNA polymerase called DNA primase. Thus, the discontinuous synthesis of OFs dictates the priming of synthesis every ~200 bp leaving behind the iRNA (8-12 nucleotide long) with each fragment. For eukaryotic DNA replication, the primer is handed-off intramolecularly to Pol α, a DNA
polymerase that lacks the exonuclease proofreading capacity. Pol α extends the iRNA further with ~20-30 deoxyribonucleotides with low fidelity. This hybrid iRNA/DNA primer is further extended by higher fidelity Pol δ. Thus, for genomic stability, the iRNA has to be removed, Pol α misincorporated nucleotides have to be corrected for, and the nicks have to be sealed. This feat is achieved through a process called maturation of OFs (164).

Approximately 50 million OFs are generated per each cell cycle, all of which have to be processed to ensure genome integrity. The maturation of OFs starts when Pol δ reaches the end of the nascent OF and is faced with the 5’-end of the previous OF. Pol δ strand displaces the 5’-end of the previous OF generating a 5’-flap protruding from a nick junction, that consists of ssDNA and/or ssRNA of various lengths (Figure 4.1A). The complementarity between the 5’ flap and the template strand would potentially equilibrate the nick junction to form one nucleotide 3’ flap (159,228,301). This DF structure is recognized by structure-specific 5’ endonucleases and cleaved one nucleotide into the junction generating a nick that is sealed by DNA Ligase 1 (Figure 4.1A). The main endonuclease involved in this process is FEN1, although other pathways have been suggested depending on the flap length (for reviews, refer to (119,164,249,254)). The primary short-flap pathway involves FEN1 recognizing short flaps and cleaving them to generate the nick while the secondary, and less frequent, pathway involves the generation of a longer flap (Figure 4.1A). This longer flap can be accessed by RPA and once RPA is bound, FEN1 cleavage is inhibited and Dna2 is recruited (162,302). Dna2 displaces RPA (170) and progressively cleaves the flap to make it shorter. At this point, Dna2 dissociates due to its lower affinity for the short flap (172) or FEN1 disengages it (173,174), and the
DNA becomes a substrate for FEN1 again. Albeit the infrequency of the secondary pathway, Dna2 is essential in *S. cerevisiae* (165,166), while FEN1 is not. This rather counterintuitive result is clarified by the presence of backup endonucleases such as RNaseH2 and EXO1 that can compensate for FEN1 on short flaps (175,303). On the other hand, longer flaps, although infrequent, can lead to deleterious consequences and cannot be tolerated which necessitates the action of Dna2.

Recent studies have provided richly detailed information on the FEN1 mechanism of action. FEN1 initially binds and bends its flap substrate to an angle of 100° at the nick junction (156,158,293). The FEN1 substrate recognition mechanism protects the template strand against inadvertent incision through a cap-helical gateway structure that oversees the active site. The cap-helical gateway selects for threading of 5’ flaps with free ends. Blocking 5’ flap threading with biotin/streptavidin at the end of the 5’ flap demonstrates that threading is a prerequisite for catalysis (275,280). Single molecule experiments with such substrates show that FEN1 achieves a weak bent state that cannot support cleavage (228). Furthermore, the conformer of the blocked 5’ flap (158) and the degree of DNA bending (Chapter 3) vary from that in the threaded complex. These single molecule findings are consistent with time-resolved crystallography data on human EXO1, which illustrate step-wise threading of the 5’ flap through the cap-helical gateway that is coupled with progressive strengthening of the interactions with the bent DNA (304). In case of both FEN1 and EXO1, the flap strand is inverted such that the phosphates are turned away from the active site metal ions, which protects the strand from inadvertent incision during threading (157,304). Threading can also be viewed in the context of regulation and the choice between the short- and long-flap pathways; for instance, RPA-coated flaps may be
blocked from threading, thus limiting catalysis. Collectively, these studies highlight the importance of 5’ flap threading in FEN1 catalytic cycle, and the possible consequences of 5’ flap length and structure on the reaction mechanism.

Garg et al. (150) and Stodola & Burgers (151) have suggested that the main product of FEN1 cleavage is a monoribonucleotide. This is further explained by a tight coupling between Pol δ and FEN1 in what they suggested as an active hand-off mechanism (nick translation). This mechanism entails Pol δ’s limited strand displacement of the 5’-end of the previous OF where FEN1’s substrate is generated readily and a monoribonucleotide is cleaved. This occurs in multiple cycles until the iRNA primer is completely removed. Besides Pol δ’s limited strand displacement, its idling between the polymerase and exonuclease activities maintain a short flap (Chapter 1). Nevertheless, long flaps do occur as evident from in vitro and in vivo studies (149,172,177-179).
A

Short-Flap Pathway

- FEN1 binds and recognizes DF-substrate
- FEN1 cleaves 5' flap
- FEN1 dissociates
- DNA Ligase 1 seals the nick
- The Result: Contiguous dsDNA

Long-Flap Pathway

- Flap escapes FEN1's cleavage
- RPA binds the long 5' flap inhibiting FEN1's cleavage
- Dna2 is recruited and actively displaces RPA
- Dna2 cleaves long 5' flap stepwise and RPA is completely displaced
- Dna2 dissociates leaving short DF, a perfect substrate for FEN1

B

Flap-Labeling Scheme

<table>
<thead>
<tr>
<th></th>
<th>Unbent DNA</th>
<th>Bent DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High FRET</td>
<td>Low FRET</td>
</tr>
</tbody>
</table>

Internal-Labeling Scheme

<table>
<thead>
<tr>
<th></th>
<th>Unbent DNA</th>
<th>Bent DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low FRET</td>
<td>High FRET</td>
</tr>
</tbody>
</table>

Acceptor

Donor

Biotin/NeutrAvidin

Ca²⁺
Deletion mutations of Pif1 and Pol δ’s third subunit (Pol32) in *S. cerevisiae*, as well as post-translational modifications (especially acetylation) of human FEN1, Dna2 and Pol δ have shed some light on how and why these long flaps are generated. Pif1, a 5’-3’ helicase, has been shown to augment Pol δ’s strand displacement capacity (in *S. pombe*) by unwinding the previous OF end (180). Similar results have been reported in the presence of Pol32 in *S. cerevisiae* as compared to the 2-subunit Pol δ (148). Therefore, both Pol32 and Pif1 are expected to increase the chance of generating longer flaps. In fact, the deletion mutation of either protein rescued the lethality caused by ΔDna2 mutation in *S. cerevisiae*, albeit ΔPol32 had a lesser effect (148). On the other hand, acetylation of FEN1, Dna2 and Pol δ by the histone acetyltransferase p300 gave insight into the regulation of the long-flap pathway. While the complete acetylation of FEN1 reduces its cleavage activity by ~90% (181), the same modification has different effect on Dna2 and Pol δ. It stimulates Dna2 nuclease and helicase activities by many folds (182), and enhances Pol δ’s strand
displacement activity (164,183). Taken together, the acetylation of these lagging strand proteins might favor the long-flap pathway.

The long-flap pathway is particularly favorable when viewed in the context of displacing and cleaving of a longer flap where the full hybrid iRNA/DNA primer incorporated with low fidelity by Pol α-primase complex, rather than just the iRNA, is removed and then replaced with higher fidelity by Pol δ. Yet, if this is a favorable pathway, why does the short-flap pathway still prevail? The long-flap pathway with its multistep processing would cause a delay in the maturation of OFs. Furthermore, if for each OF, 30-40 bp have to be removed and replaced, a significant portion of the genome would have to be made twice. Thus, it would make sense that the cell would opt for this pathway only in selected regions. The p300 acetylase has been hypothesized to acetylate replication/repair proteins at the sites of active genes (164). Moreover, components from the long-flap pathway, including Dna2 and Pif1, have been suggested to stimulate FEN1 in S. Cerevisiae (242).

Given that longer flaps do exist in vivo and in vitro, we set out to understand how FEN1 deals with these flaps. FEN1 is a ubiquitous protein and if present in the vicinity of these longer flaps would it compete with RPA to bind to these longer flaps? If it does indeed compete with RPA for the binding, once bound to the long flap, is it capable of recognizing its substrate, ordering itself, and assembling the active site in a similar fashion to the short-flap substrate? FEN1’s bending of its substrate at the nick junction to a 100° angle has been one of the much-studied steps in FEN1’s substrate recognition (156,158,293). Using DF-substrates containing short 5’flap, we showed recently that DF-substrate exists in an
extended conformer that is being actively bent by FEN1 in diffusion-limited kinetics (Chapter 3). This bending mutually induces FEN1’s protein ordering for active site assembly in a mechanism that is mediated by binding of the one-nucleotide 3’flap to the 3’flap-binding pocket (Chapter 3) (156,274,282). This DNA-protein mutual-induced fit mechanism always leads to the assembly of a catalytically-competent active site from the first DNA bending step in case of cognate DF-substrate, while in the case of non-cognate flap substrates, it suppresses the probability of forming catalytically-competent active site and promotes their pre-incision dissociation (Chapter 3 and (228)). Here, we have employed single molecule FRET (smFRET) and stopped-flow ensemble cleavage assays with equilibrated DF-substrates of different 5’flap lengths to study human FEN1’s bending as well as catalytic efficiency on long flap substrates. We found that upon encountering a longer flap, FEN1’s bending efficiency is not significantly affected, its single turnover kinetics is slightly decreased, but the chance of cleaving from the first DNA bending step significantly decreases by flap length increase. These non-tolerated missed opportunities would give a delay before which FEN1 can rebind and have a second chance at cleaving the longer flaps. This delay would give RPA the perfect chance to bind and inhibit the downstream cleavage by FEN1 unless Dna2 gets involved. Furthermore, we demonstrate that FEN1 can access RPA-bound long flaps, which might suggest a handoff mechanism to Dna2.
4.3 Results

In this study, we set out to investigate whether 5' flap length affects FEN1 substrate recognition and catalytic efficiency, and whether any variation in FEN1 activity might influence the choice between short- versus long-flap pathways for OF maturation. We addressed these questions by measuring transient events in the reaction using smFRET complemented with bulk cleavage experiments and determining the FEN1 kinetic mechanism on short and long 5' flaps, both in the absence and presence of RPA.

Most DF substrates used in this study, unless otherwise noted, are capable of flap equilibration, i.e., the template base at the nick junction is complemented by both upstream and downstream strands to generate either a SF or DF substrate. EQ DF substrates best reflect the dynamicity of flap substrates encountered by FEN1 in vivo; although, it has been shown in vitro that FEN1 cleaves non-equilibrated (Non-EQ) DF substrates with similar single turnover rate ($k_{STO}$) as EQ DF (Chapter 3) (295). Similar to the work presented in Chapter 3, we used two DNA labeling schemes for the smFRET assays, namely "flap-labeling" and "internal-labeling". In the flap-labeling scheme (Figure 4.1B), the DF substrate has a Cy3 donor at the 5’-end of the flap and an Alexa Fluor 647 acceptor at nucleotide position 12 upstream of the nick junction. In the internal-labeling scheme, the dyes are located on either side of the nick junction, with Cy3 at position 15 downstream and Alexa Fluor 647 at position 12 upstream from the nick (Figure 4.1B).
4.3.1 Effect of 5’ flap length on the DNA bending activity of FEN1

First, we assessed the ability of FEN1 to bind and bend DNA substrates with varying flap lengths. FEN1 was titrated with internally labeled EQ DF substrates with 6, 29, 50 or 60 nt long 5’ flaps under non-catalytic conditions in the presence of Ca$^{2+}$ ions. Histograms were constructed for each DF substrate as described in Chapter 2. The FRET peaks were centered around 0.30 for all unbent DF substrates and around 0.52 for the FEN1-bent substrates (Figure 4.2A-D). The percentage of bent substrate at each FEN1 concentration was determined by the integrated area of the Gaussian-fitted bent peak, and plotted versus FEN1 concentration to generate the isotherms shown in Figure 4.2A-D. The data yielded the following DNA bending dissociation constants ($K_{d\text{-bending}}$): EQ DF-6,1\text{Internal} = 4.8±0.6 nM; EQ DF-29,1\text{Internal} = 3.3±0.4 nM; EQ DF-50,1\text{Internal} = 4.1±0.5 nM; and EQ DF-60,1\text{Internal} = 17.6±3.1 nM (Figure 4.2A-D). The time traces showed that in all cases, the DNA alone has a single FRET conformer that is actively bent upon addition of FEN1, consistent with our previous results (Chapter 3). While EQ DF-6,1\text{Internal} and EQ DF-29,1\text{Internal} substrates transitioned between bent and unbent states in the presence of FEN1, EQ DF-50,1\text{Internal} and EQ DF-60,1\text{Internal} displayed higher stability in the FEN1-bound bent state, including a significant portion of bent particles that did not undergo any transition (Figure 4.2C and D). This higher stability observed with EQ DF-50,1\text{Internal} and EQ DF-60,1\text{Internal} implies that FEN1 has a lower rate of binding/bending given that the $K_{d\text{-bending}}$ does not change in case of EQ DF-50,1\text{Internal} and increases slightly in case of EQ DF-60,1\text{Internal}. EQ DF-6,1\text{Internal} and EQ DF-29,1\text{Internal} displayed diffusion-limited association rates; (1.57±0.47) x 10$^8$ and (1.33±0.01) x 10$^8$ M$^{-1}$s$^{-1}$, respectively, with similar dissociation rates.
Figure 4.2. Effect of 5' flap length on FEN1 DNA bending activity. (A) FEN1 bending efficiency of internally-labeled equilibrated DF-6,1\textsubscript{Internal} (EQ DF-6,1\textsubscript{Internal}). Top Left panel shows smFRET histograms of EQ DF-6,1\textsubscript{Internal} alone and upon addition of 5 nM FEN1. The histograms were fitted by one or two Gaussian distributions for DNA-only and DNA + FEN1, respectively. In the DNA + FEN1 histogram, the fitted unbent peak (shown in magenta) has same FRET center as DNA-only, and the fitted bent peak is shown in blue. Top
Right panel shows representative single molecule time traces of EQ DF-6,1_int as alone and upon addition of 5 nM FEN1. Bottom Left panel is an isotherm of percentage of bent substrate (%) versus FEN1 concentration (nM). The percentage of bent substrates was estimated by the Gaussian-fitted bent peak for three replicates at each FEN1 concentration. The isotherm was fitted to a one-site binding model with Bmax ≤ 100 and yielded the DNA bending dissociation constant ($K_{d\text{-bending}}$). Error bars reflect variation of the % bent substrate from the three replicates and the reported error is the error of the fit. Bottom Right panel shows a graph of $k_{\text{bending}}$ (s$^{-1}$) and $k_{\text{unbending}}$ (s$^{-1}$) versus FEN1 concentrations (nM). At each FEN1 concentration, time traces were idealized and fit by vbFRET to calculate the dwell times $\tau_{\text{bending}}$ and $\tau_{\text{unbending}}$ spent in the bent and unbent states, respectively. The histograms from the population of the dwell times were fit to exponential functions yielding $k_{\text{bending}}$ (1/$\tau_{\text{bending}}$) and $k_{\text{unbending}}$ (1/$\tau_{\text{unbending}}$). The association rate constant ($k_{\text{on\text{-bending}}}$) was calculated from the slope of the linear regression fit of $k_{\text{bending}}$ versus FEN1 concentration. The dissociation rate constant ($k_{\text{off\text{-unbending}}}$) was calculated as the mean of $k_{\text{unbending}}$. The error bars correspond to the standard error of the exponential fit of $k_{\text{bending}}$ and $k_{\text{unbending}}$ and the errors reported for the association/dissociation constants correspond to the error of the fit and the standard error of the mean, respectively. $K_{d\text{-bending}} = k_{\text{off\text{-unbending}}} / k_{\text{on\text{-bending}}}$. (B) FEN1 bending efficiency of EQ DF-29,1_int as described in A. (C) FEN1 bending efficiency of EQ DF-50,1_int as described in A. (D) FEN1 bending efficiency of EQ DF-60,1_int as described in A.

### 4.3.2 Effect of Mg$^{2+}$ on DF substrates

The requirement of 5’ flap threading for catalysis is supported by the inhibition of FEN1 cleavage of substrates with flaps containing chemical adducts or blocked flaps with biotin/streptavidin interaction (275). With this pre-requisite for catalysis, it is logical that the ssDNA conformation of the 5’ flap might play a role in the overall catalysis efficiency. To gain insight into the possible ssDNA conformational changes within the 5’ flap and their dependence on 5’ flap length, we investigated the effect of Mg$^{2+}$ on DF substrates with two different flap lengths (DF-6,1 and DF-12,1). To address this question, we used the flap-labeling scheme. Increasing Mg$^{2+}$ concentration has been shown to affect the electrostatic interactions of the DNA through its interactions with the phosphate backbone, and
consequently rigidifying the DNA structural conformation, thereby slowing down the transition rates between different conformers (305-308). With that in mind, we hypothesized that increasing the Mg$^{2+}$ concentration would slow down the 5' flap ssDNA conformational changes if present. Moreover, we sought to investigate the effect of Mg$^{2+}$ and flap length on the dsDNA of the nick junction; thus, we used the internal-labeling scheme.

**Figure 4.3. Effect of Mg$^{2+}$ on the conformational changes of DF substrates.** (A) Effect of various Mg$^{2+}$ concentration on DF-6,1. TIRF-smFRET histograms of Non-EQ DF6,1$_{\text{Internal}}$ (left panel) and Non-EQ DF6,1$_{\text{Flap}}$ (right panel) with increasing concentrations of MgCl$_2$ (0 mM, 1 mM, 10 mM and 50 mM). The insensitivity of both labeling schemes to increasing divalent metal ion concentrations demonstrates that both labeling schemes report directly on the geometry of the duplex DNA. (B) Effect of various Mg$^{2+}$ concentration on DF-12,1. TIRF-smFRET histograms of Non-EQ DF12,1$_{\text{Internal}}$ (left panel) and Non-EQ DF12,1$_{\text{Flap}}$ (right panel) with increasing concentrations of MgCl$_2$ (0 mM, 1 mM, 10 mM and 50 mM). The
sensitivity of the flap-labeling scheme to varying divalent metal ion concentrations but not the internal-labeling scheme demonstrates that the geometry of the duplex DNA is not influenced by the length of the 5’ flap and that the flap-labeling scheme is inappropriate to describe the geometry of the duplex DNA only when the 5’ flap length exceeds 6 nt. All TIRF-smFRET measurements were acquired at 160 ms temporal resolution. FWHM represents the full width at half maximum of the Gaussian peak.

We tested the FRET efficiency of Non-EQ DF-6,1 and Non-EQ DF-12,1 in the two labeling schemes with increasing Mg$^{2+}$ concentration from 0 to 50 mM. For Non-EQ DF-6,1, Mg$^{2+}$ displayed no effect on the center of the FRET histogram peak nor its broadness in both labeling schemes (Figure 4.3A). In contrast, on Non-EQ DF-12,1, Mg$^{2+}$ demonstrated differential effects with two different labeling schemes (Figure 4.3B). While the internal-labeling scheme showed consistent FRET histogram peak centers and broadness with the varying Mg$^{2+}$ concentrations, the flap-labeling scheme exhibited an increase in the FRET histogram peak center from E=0.51 to E=0.73 as the Mg$^{2+}$ concentration increased from 0 to 50 mM. Moreover, the flap-labeling scheme showed that the FRET histograms broadness of Non-EQ DF-12,1 decreased from FWHM=0.25 to FWHM=0.19 with the increasing Mg$^{2+}$ concentration. Furthermore, the FRET efficiency of Non-EQ DF-6,1 and Non-EQ DF-12,1 in the case of the internal-labeling scheme are equivalent.

The results with the flap-labeling scheme suggest that flaps with lengths below 6 nts most likely exist in one conformation. However, as the flap length increases beyond 12 nts, one can extrapolate that the 5’ flap would undergo various conformational changes with rates that were slowed down with increasing Mg$^{2+}$ concentration. The results with the
internal-labeling scheme, on the other hand, suggest that increasing Mg$^{2+}$ or flap length did not affect the duplex portion of the nick junction.

### 4.3.3 Effect of 5’ flap length on FEN1 kinetic mechanism

To test FEN1 activity on varying flap lengths (6, 29, 50 and 60 nt), we performed cleavage experiments as established in Chapter 3, mainly with internal-labeled substrates since in the presence of 10 mM Mg$^{2+}$, the long flaps have higher flexibility as shown in Figure 4.3B. This flap movement of long flap substrates complicated the identification of real bending and cleavage events in flap-labeling scheme (data not shown). smFRET cleavage experiments were performed with 250 nM FEN1, which is ~15-fold higher than the weakest $K_{d-bending}$ measured for the DF substrates (17.6 nM for DF-60,1$_{\text{Internal}}$; Figure 4.2B-D), but lower than the $K_{d-bending}$ of nicked DNA (as described in Chapter 3). We were able to identify cleavage events due to the slight difference in FRET efficiency between unbent DF substrates (E~0.3) and the unbent nicked product (E~0.25), and thus obtain the average $\tau_{\text{bending-internal}}$ for each DF substrate as shown in Figure 4.4A-D. Interestingly, there were no significant differences in $\tau_{\text{bending-internal}}$ values across the flap lengths (315±65 ms for DF-29,1, 350±70 ms for DF-50,1 and 360±65 ms for DF-60,1), and even compared to DF-6,1 (270±70 ms). Given that $k_{\text{release}}$ for the common nicked product is not expected to vary with 5’ flap length, similar $\tau_{\text{bending-internal}}$ values mean that $k_{\text{STO}}$ is similar across the flap lengths as well.
Alternatively, we conducted complementary rapid quench-flow bulk cleavage experiments with the same EQ DF substrates to validate the finding that $k_{STO}$ and $k_{cat}$ do not vary significantly with increasing 5' flap length. $k_{STO}$ was determined at 35:1 ratio of FEN1:DNA, while $k_{cat}$ was determined at 1:800 ratio of FEN1:DNA. The $k_{STO}$ was slightly higher for 30 nt ($32.5\pm1.4 \text{ s}^{-1}$) compared to 6 nt ($21\pm0.9 \text{ s}^{-1}$) (Chapter 3) and slightly lower for 50 nt ($12.6\pm0.7 \text{ s}^{-1}$) and 60 nt flaps ($9.9\pm0.5 \text{ s}^{-1}$) (Figure 4.4E). The $k_{cat}$ also followed the same trend as $k_{STO}$, showing a small change with increasing flap length relative to DF-6,1 (2-fold maximum; Figure 4.4E), which is not surprising since the rate-limiting product release step should be comparable for all the substrates. While this finding is consistent with our smFRET results, it contrasts with a recent report suggesting that the catalytic efficiency of FEN1 is significantly compromised with 5' flaps longer than 45 nt (~10-fold decrease in $k_{STO}$) (295). The authors compared $k_{STO}$ to $k_{cat}$ and suggested that beyond 45 nt flaps, the rate limiting step in the reaction switches from product release to 5' flap threading. We note, however, that while we utilized poly(T) flaps in this study, Tarantino et al., (295) had mixed sequence flaps, and an analysis of the sequences by IDT OligoAnalyzer 3.1 indicates a propensity for secondary structure, especially in their DF-60,1M substrate (hairpin of $T_m = 51.9^\circ\text{C}$). The $k_{STO}$ of FEN1 is significantly reduced on DF substrates containing 5' flap hairpins (241,301); hence, it is possible that secondary structures formed due to the mixed sequence rather than flap length account for the reported impairment in FEN1 activity on long flap substrates.
Figure 4.4. Effect of 5' flap length on FEN1 cleavage activity. (A) smFRET cleavage of EQ DF-6,1\textsubscript{internal} as described in Chapter 3. Top: distribution of the dwell times spent in bent state ($\tau_{\text{bending-internal}}$) for $N=64$ cleavage events fitted to a gamma distribution. Average $\tau_{\text{bending-internal}}$ is reported with the standard error of the mean. Bottom: a representative single molecule time trace showing FEN1 bending and cleaving the substrate.
before FRET drops to 0.25; the inset zooms in on a vbFRET-fitted version of the cleavage event showing a three-state fit (0.3, 0.52 and 0.25) corresponding to the three DNA conformers, unbent DF-6,1\textsubscript{Internal}, bent DF-6,1\textsubscript{Internal} and unbent nicked product. The cleavage reaction was performed at 50 ms temporal resolution. (B) smFRET cleavage of EQ DF-29,1\textsubscript{Internal} as described in A. (C) smFRET cleavage of EQ DF-50,1\textsubscript{Internal} as described in A. (D) smFRET cleavage of EQ DF-60,1\textsubscript{Internal} as described in A. (E) Ensemble cleavage kinetics of FEN1 on longer flap EQ DF substrates. Left: Single turnover cleavage of DF-30,1, DF-50,1 and DF-60,1. Right: steady state cleavage of DF-30,1, DF-50,1 and DF-60,1. The rates were measured and reported as described in Chapter 2,3.

Interestingly, a closer look at the time traces revealed some particles undergoing cleavage after missed opportunities in which FEN1 bends the DF substrate but does not cleave the flap (Figure 4.5A). These missed events resulted in a return to the unbent substrate with FRET efficiency ~0.3 rather than the unbent nicked product (E ~0.25), followed by one or more additional binding and bending events until cleavage occurred. The assignment of a bending event as either a missed cleavage or cleavage event was further confirmed by constructing FRET histograms from the traces that showed multiple bending events in DF-50,1. As shown in Figure 4.5B, the FRET values of the section of traces before any bending occurs average 0.29 (highlighted in green), those of sections with periods separating any two bending events also average around 0.29 (highlighted in blue), but the FRET values of the section after the last bending event average around 0.23 (highlighted in red). This pattern confers confidence in the assignment of the first bending event(s) as missed cleavage event(s) and the last one as a cleavage event. Note that FEN1 concentration is below $K_d$-bending of nicked DNA and the experiments are performed at 100 mM KCl; therefore, FRET increases subsequent to the first one do not represent FEN1 binding to nicked product. On some of the particles, FEN1 went through multiple missed
tries at cleavage, but the majority were cleaved after one miss. Interestingly, the percentage of particles with at least one missed event increased significantly from 3.1% in the case of DF-6,1 to 21% in the case of DF-60,1 (Figure 4.5C). We also found that the fraction of particles with more than one missed event increased with flap length. Analysis of the dwell time spent in the bent state during missed events ($\tau_{\text{bending\text{-}missed}}$) showed that $\tau_{\text{bending\text{-}missed}}$ did not differ significantly across the flap lengths (note: in case of DF-6,1 the sample size of 3.1% was too small to draw any conclusion). Moreover, the $\tau_{\text{bending\text{-}missed}}$ (~270 ms) was in the same range as $\tau_{\text{bending\text{-}internal}}$ (Figure 4.5C and Figure 4.4A-D, respectively). This result indicates that during the missed opportunities FEN1 can access and bend all the DF substrates with similar efficiency but still cannot assemble a catalytically competent active site. We also note that the 5' flap might be getting threaded into the cap-helical gateway in these missed-cleavage bending events since a DF substrate in which the 5' flap is blocked from threading dissociates ~13-fold faster at 47.2 s$^{-1}$ (Chapter 3) (228) than the ~3.7 s$^{-1}$ rate ($1/\tau_{\text{bending\text{-}missed}}$) measured here (Figure 4.5C). It is also possible that these events result from trapping of partially threaded complexes in a fashion similar to that observed with EXO1 (304).

We also noticed that in contrast to DF-6,1, cleavage of DF-50,1 and DF-60,1 occurs in an asynchronous manner after FEN1 enters the flow cell, even under single turnover conditions. This behavior was clear when we plotted the distribution of the initial time point in each cleavage event for DF-6,1, DF-50,1 and DF-60,1. The distributions are broader for the 50 and 60 nt flaps as compared to the 6 nt flap, suggesting that a decrease in the DNA binding and bending rate with increasing flap length also influences cleavage (Figure 4.5D). We speculate that this reduction reflects challenges faced by FEN1 in
binding the nick junction in the context of long 5' flaps rather than flap threading, since significant DNA bending occurs even without threading of the 5’ flap into the cap-helical gateway (Chapter 3) (228).

To summarize our results thus far, both single molecule and bulk cleavage analyses demonstrate that 5’ flap length has a minor inhibitory effect on the FEN1 reaction mechanism. The missed cleavage behavior on long flap substrates indicates that the most striking difference occurs at a step after DNA binding and bending but before, chemistry and product release. However, the steady state rate-limiting step in the reaction remains nicked product release. Accordingly, the overall effect of increasing flap length on FEN1 catalytic efficiency is relatively small.
Figure 4.5. FEN1 misses cleavage, as the flap grows longer. (A) FEN1 almost always cleaves a short-flap substrate (EQ DF-6,1Internal) in the first bending event (Figure 4.4). With longer flaps, FEN1 exhibits missed opportunities at cleavage (defined as an unproductive bending event wherein FRET drops to unbent substrate state (0.3) rather than unbent product state (0.25)). Representative single molecule time traces show a majority of these events are single missed opportunities (left) and a minority are multiple missed opportunities (right). Insets zoom in on vFRET-fitted versions of the missed events as well as the cleavage events with three-state fits. The fits show FRET returning to the substrate unbent state (0.3) in missed events and dropping to the product unbent state (0.25) in cleavage events. (B) Assignment of a bending event as missed cleavage or actual cleavage. Time traces that exhibit multiple bending events in EQ-DF-50,1Internal cleavage reaction (N=15) were considered for further quantification of the unbent FRET state. As shown on the representative time trace, three regions of the traces were taken into consideration, the region before any
bending event occurred (green), the region between any two bending events (blue), and the region after the
last bending event (red). The distributions of the FRET states occupied in each region is plotted and shown
with the corresponding color. (C) Quantification of missed cleavage events. Top: A bar chart shows that the
percentage of particles exhibiting cleavage with missed opportunities increases with flap length from 3.1%
(DF-6,1) to 21.0% (DF-60,1). Bottom: A bar chart shows the average time spent in bent state during missed
events ($t_{bending\text{-}missed}$) by the different DF substrates. The reported N is the number of missed events, and not
the number of particles, accounting for particles with multiple missed events. (D) Cleavage of DF-50,1 and
DF-60,1 occurs in an asynchronous manner after FEN1 enters the flow cell in contrast to DF-6,1. Distributions of the first time point of the bending step in each cleavage event were plotted for the different
substrates (DF-6,1-cyan, DF-50,1-green and DF-60,1-purple). The data were fit to normal distributions and
the means with standard deviation and N are reported. The mean and standard deviation values indicate
broader distributions for DF-50,1 and DF-60,1 as compared to DF-6,1.

To date, most in vitro characterizations of FEN1 substrate recognition and catalytic
efficiency have involved substrates with DNA flaps. Since RNA flaps arise in vivo and
could be substrates for FEN1 cleavage, we investigated the effect of replacing a DNA flap
with RNA in EQ DF-6,1 and EQ DF-29,1 substrates. Single molecule bending experiments
in the presence of Ca$^{2+}$ showed that FEN1 is only modestly (~3-fold) defective in accessing
both RNA-flap substrates (Table 4.1). However, the stability of the bent complexes varies
with flap length, such that FEN1 shows a modest, ~3-fold reduction in stability on EQ DF-
6,1-RNA and ~10-fold reduction on EQ DF-29,1-RNA when compared with the
corresponding DNA substrates (Table 4.1). Nonetheless, single molecule cleavage assays
with internal-labeled substrates showed slightly faster cleavage kinetics with RNA-flap
substrates as compared with corresponding DNA substrates. These findings are in line with
a previous report on SF RNA substrates (309). Importantly, the missed cleavage behavior
of FEN1 remains the same with an RNA flap as with DNA. Overall, we conclude that
although FEN1 exhibits reduced stability on RNA-flap substrates, its catalytic efficiency is not limited significantly by the higher dissociation rate.

Table 4.1. Kinetic parameters of FEN1 substrate bending and catalytic efficiency on DNA versus RNA DF substrates.

<table>
<thead>
<tr>
<th></th>
<th>EQ DF-6,1 DNA</th>
<th>EQ DF-6,1 RNA</th>
<th>EQ DF-29,1 DNA</th>
<th>EQ DF-29,1 RNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_{d-binding}$ (nM)</td>
<td>4.8 ± 0.6</td>
<td>32.1 ± 8.0</td>
<td>3.3 ± 0.4</td>
<td>103.0 ± 15.4</td>
</tr>
<tr>
<td>$k_{on}$ ($10^4$ M$^{-1}$s$^{-1}$)</td>
<td>1.57 ± 0.47</td>
<td>0.58 ± 0.02</td>
<td>1.33 ± 0.01</td>
<td>0.39 ± 0.02</td>
</tr>
<tr>
<td>$k_{off}$ (s$^{-1}$)</td>
<td>0.45 ± 0.02</td>
<td>1.38 ± 0.02</td>
<td>0.37 ± 0.02</td>
<td>4.35 ± 0.04</td>
</tr>
<tr>
<td>$T_{bending-internal}$ (ms)</td>
<td>270 ± 70</td>
<td>160 ± 30</td>
<td>315 ± 65</td>
<td>220 ± 50</td>
</tr>
<tr>
<td>Missed cleavage events (%)</td>
<td>3.1</td>
<td>4.7</td>
<td>10.2</td>
<td>14.1</td>
</tr>
</tbody>
</table>

4.3.4 Coordination between RPA and FEN1 on short- and long-flap substrates

We have shown that the probability of FEN1 missing cleavage within its first encounter with a DNA substrate increases with 5' flap length. While this probability seems low, the outcomes of any missed cleavage event can be significant. If left unprocessed, long flaps can form secondary structures that may impede DNA replication and repair, undergo recombination at ectopic sites, and result in duplication of sequences, among other deleterious effects (164); hence, they are not generally tolerated by the cell. As noted
earlier, deletion of Dna2, which is required to process long flaps along with RPA (Figure 4.1), is lethal in *S. cerevisiae* whereas deletion of the FEN1 orthologue (Rad27) is tolerated, albeit with a severe mutator phenotype (165,166). We posit that the availability of an alternate pathway for processing long flaps is very important, given our discovery that FEN1 occasionally misses cleaving such DNA substrates. To test this idea, we asked whether FEN1 activity has any impact on how long flaps are diverted to the secondary long-flap pathway.

One way to address this question was to monitor the actions of RPA and FEN1 with respect to each other on short- as well as long-flap substrates. A gel-based assay showed that FEN1 cleavage activity on DF substrates with varying flap lengths (DF-2,1, DF-6,1 and DF-30,1) remains unchanged on DF-2,1 and DF-6,1 but is inhibited on DF-30,1 with increasing RPA (Figure 4.6A). This result can be explained by the dual single stranded DNA-binding modes of RPA: a weak, transient 8-nucleotide mode (*K*<sub>d-binding</sub> ~50 nM) and a stronger, more stable 30-nucleotide mode (*K*<sub>d-binding</sub> ~0.5 nM) (310), indicating RPA forms a stable complex with the 30 nt flap. The next series of experiments measured DNA binding/bending and cleavage by FEN1 using the smFRET assays described above. First, as a control, we tested the effect of RPA binding to two internal-labeled short- and long-flap DF substrates (DF-6,1<sub>Internal</sub> and DF-29,1<sub>Internal</sub>) in the presence of Ca<sup>2+</sup>. For both substrates, FRET efficiency histograms with DNA alone or DNA with increasing concentrations of RPA showed a single peak centered at ~0.33 and ~0.31, respectively (Figure 4.6B); thus, RPA does not appear to distort these DF substrates, which simplified analysis of FEN1 activity in the presence of RPA. Subsequent experiments were performed with 100 nM RPA for DF-6,1 and 1 nM RPA for DF-29,1, which is two fold higher than
**K\textsubscript{d-binding}** of both binding modes. As shown in Figure 4.6C, FEN1 binds and bends DF-6,1 similarly in the absence or presence of 100 nM RPA. The histograms in presence of both proteins showed two separable peaks centered around the same FRET efficiencies of E~0.3 and E~0.52 (Figure 4.6C) as with FEN1 alone (Figure 4.2A). The time traces also showed similar transitions between bent and unbent states as seen with FEN1 alone. Moreover, the bending efficiency of FEN1 in presence of RPA was comparable to that of FEN1 alone, as evident from similar **K\textsubscript{d-binding}** constants (4.8±0.6 nM and 8.6±1.2 nM in the absence and presence of RPA, respectively; Figure 4.6E). In sharp contrast, with DF-29,1, the presence of 1 nM RPA caused the FRET histogram peaks to merge with the centers shifted (Figure 4.6D) when compared with FEN1 alone (Figure 4.2B); note: the time traces showed that the bent and unbent state transitions were faster than our 100-ms temporal resolution, resulting in an averaging effect and merged peaks in the FRET histograms. With DF-29,1, unlike DF-6,1, RPA had an inhibitory effect on FEN1 binding and bending efficiency. The averaging effect complicated data fitting and determination of **K\textsubscript{d-binding}** for FEN1 in the presence of RPA, nevertheless, a lower estimate of 318.0±30.7 nM (Figure 4.6E) was ~100 fold higher than that in the absence of RPA (3.3±0.4 nM; Figure 4.2B). This result indicates that RPA significantly lowers FEN1 affinity for a long-flap substrate, but FEN1 can still access the DNA and bend it (Figure 4.6D). In other words, the inhibitory effect of RPA does not stem from complete blockage of FEN1 from the substrate, but is more likely due to the inability of FEN1 to form a stable bent conformer as a result of blocked 5' flap threading by bound RPA. By analogy with 5' flaps bound and blocked from threading by biotin/streptavidin, the lifetime of the bent conformer with an RPA block is significantly less than that required for FEN1 to catalyze cleavage (Chapter 3) (228).
Figure 4.6. Coordination between FEN1 and RPA actions on short- and long-flap substrates. (A) Steady state FEN1 cleavage activity in the presence of RPA. Gel showing short- and long-flap substrates cleaved by FEN1 in the presence of increasing RPA (0-5 nM). While RPA has no effect in the case of short flaps (DF-2,1 and DF-6,1), it inhibits FEN1 activity on DF-30,1 in a concentration dependent manner. (B) RPA shows no effect on the structure, and thus the FRET, of EQ DF-6,1_internal and EQ DF-29,1_internal. Left: EQ DF-6,1_internal DNA-only histogram (top) with FRET centered ~0.33, and upon addition of 200 nM RPA (bottom) with FRET centered ~0.31. Right: corresponding histograms for EQ DF-29,1 internal substrate. (C) smFRET bending of EQ DF-6,1_internal by FEN1 in the presence of 100 nM RPA. Top: histogram showing distribution of FRET states upon addition of 20 nM FEN1 (unbent peak shown in magenta, bent peak shown in blue). The peaks are well separated and centered around the same FRET values as in the absence of RPA (Figure 4.2A). Bottom: a representative single molecule time trace showing similar transitioning rates between bent and unbent states as seen without RPA. (D) smFRET bending of EQ DF-29,1_internal by FEN1 in the presence of 1 nM RPA. Top: histogram showing distribution of FRET states upon addition of 2000 nM FEN1. The peaks (unbent in magenta, bent in blue) are merged and the centers are shifted from those seen in the absence of RPA (Figure 4.2B). Bottom: a representative single molecule time trace showing fast transitions between bent and unbent states that cannot be resolved within the temporal resolution of acquisition (100 ms). With
such fast transitions, the FRET state captured within each frame (100 ms period) is an average and not a true FRET state. This averaging explains why the traces do not show distinct FRET states, and why the full bending (0.5) state is not reached. This effect appears as merging of the peaks in the histograms. Therefore, at any particular concentration, the percentage of the bent peak is underestimated, and consequently the $K_{d-bending}$ as well (E). A bar chart illustrating RPA effect on FEN1 $K_{d-bending}$. RPA has no effect on the $K_{d-bending}$ of FEN1 for DF-6,1, but increases $K_{d-bending}$ by >100 fold for DF-29,1 (note that this value is a lower estimate, due to the averaging effect noted above, given that the bent state does not saturate even at 2000 nM FEN1; panel D).

Finally, the effect of RPA on FEN1 cleavage activity was measured under single turnover conditions using flap-labeled DF-6,1$_{Flap}$ and DF-29,1$_{Flap}$ substrates. The DNAs were pre-incubated with 100 nM or 1 nM RPA, respectively, prior to co-injection of the same concentration of RPA and 250 nM FEN1. FEN1 cleaved DF-6,1 with comparable efficiency in the presence ($\tau_{bending-flap} = 190\pm40$ ms; Figure 4.7A) or absence of RPA ($\tau_{bending-flap} = 155\pm30$ ms; Chapter 3). We do not anticipate any effect of RPA on $k_{cat}$ with DF-6,1 since $K_{d-bending}$ of the product was not affected by RPA (Figure 4.7B). Together, these results show that RPA neither stimulates nor inhibits FEN1 binding, bending, cleavage or multiple turnover kinetics on short-flap substrates. Moreover, as predicted by RPA-induced inhibition of a stable DF-29,1 bent conformer bound to FEN1 (Figure 4.6D), cleavage of this substrate was severely inhibited by RPA (data not shown, since cleavage events were scarce). It should be noted that a previous study suggests that S. cerevisiae RPA stimulates FEN1 activity on short flaps and inhibits it on long flaps (242). According to the results of our study, human RPA does not affect FEN1 activity on short flaps and inhibits it on long flaps. S. cerevisiae and human RPA have been found to have differential effects on another endonuclease, EXO1, as well (311). While the molecular basis of the
variation is unknown at this time, it may reflect a subtle difference between the structure and function of RPA in these two organisms.

Based on the findings of this study, we propose that while FEN1 is capable of processing a long flap by itself, it misses cleavage every so often, allowing the abundant RPA at the replication fork to compete effectively for binding the flap. The resulting inhibition of cleavage requires Dna2 to displace RPA and shorten the flap as part of the long-flap processing pathway.

Figure 4.7. RPA does not affect FEN1 single or multiple turnover kinetics on short flaps. (A) FEN1 cleavage efficiency on EQ DF-6,1\textsubscript{flap} was assayed in the presence of 100 nM RPA and at 50 ms temporal resolution. Left: single molecule time trace showing cleavage wherein a brief bending event is followed by loss of signal due to flap release. Right: the distribution of $\tau_{bending-flap}$ for N=112 cleavage events in the presence of RPA was fitted with a gamma distribution and the mean with standard error is reported. (B) FEN1 bending efficiency of Product\textsubscript{internal} in the absence (left) and presence (right) of 20 nM RPA. Isotherms
were obtained and fit as described for EQ DF-6,1_{Internal} in Figure 4.2A. RPA does not display any significant effect on FEN1 bending efficiency with Product_{Internal}, and hence is unlikely to affect its product release kinetics.
Chapter 5

A possible role for DNA-mediated and PCNA-coordinated product handoff between FEN1 and DNA Ligase 1*

5.1 Abstract

Okazaki fragment maturation proceeds by the coordinated action of multiple enzymes including FEN1, DNA Ligase 1 and Pol δ. PCNA has been suggested to act as the conductor of this elaborate molecular orchestra. Two competing models have been proposed to describe this coordination; toolbelt and sequential switching models. However, the exact details of this molecular choreography remain elusive. One intriguing aspect is whether the substrate transactions among the different enzymes involved are solely orchestrated by PCNA or the enzymes themselves also participate in the handoff of intermediates. Here, using smFRET, we show preliminary data that suggest that in the case of human FEN1 and LIG1, there is a possible allosteric mechanism of substrate handoff in a PCNA-independent manner, which might be subject to further regulation when PCNA is present. Additional characterization of PCNA involvement showed enhancement of FEN1 product stability, which supports its role in sequestering the toxic intermediates before passing them onto the next appropriate enzyme. We have optimized the purification of human RFC, the PCNA clamp loader, and showed that it is able to load PCNA onto DNA.

* This Chapter contains preliminary data.
5.2 Introduction

The requirement for efficiency and precision in OF processing suggests that the actions of the polymerase strand displacement, cleavage of the generated flap and sealing of the resulting nick be tightly coupled and coordinated. The enzymes responsible for these three actions, namely the lagging strand polymerase, Pol δ, FEN1 and DNA Ligase 1, all contain a PIP box and have been shown to directly interact with PCNA (187,188,312-316). Hence, it is not surprising that PCNA has been proposed to coordinate the maturation of OF process (136,137). PCNA is a ring-shaped homotrimer and a member of the structurally conserved sliding β clamps (137,317). PCNA tethers the replicative polymerases to the DNA and increases their processivity. However, the list of PCNA interacting partners extends to a diverse set of proteins involved in many DNA metabolic pathways. Not only does PCNA interact with these partners, but also recruits them to the replication fork or their chromosomal loci, enhances their catalysis, and above all choreographs their coordinated multistep processes (134,318).

In addition to DNA replication, these PCNA-interacting partners are involved in translesion synthesis, mismatch, base excision and nucleotide excision repair pathways, chromatin remodeling, cell-cycle control, among others (134,136,137,319-321). Indeed, the tight coordination and partner switching has proven vital to many PCNA-mediated processes in DNA replication and repair pathways (137,318). However, most of these protein partners interact with PCNA via the same binding site. Thus, their coordinated recruitment by PCNA might stem from switching among the partners through affinity-
driven competition to the same binding site (137). Furthermore, posttranslational modifications of PCNA, especially phosphorylation, SUMOylation and both mono- and poly-ubiquitination, have shown to regulate PCNA partner switching by biasing certain partners recruitment (322-324). Nevertheless, eukaryotic PCNA is a homotrimer and can possibly interact with more than one partner protein at the same time.

The coordination of OF maturation is the best studied example of such sequential multistep recruitment of enzymatic activities mediated by PCNA. The general pathway of OF maturation and its enzymatic activities have been well established, albeit with several missing mechanistic details. The temporal and spatial recruitment and regulation of its enzymes at particular time points is of utmost significance. For instance, the regulation and tight coupling of Pol δ and FEN1 activities during nick translation ensures processing of the flaps while still at short lengths without triggering the long-flap pathway (150,151,153,325). The requirement of DNA Ligase 1 nick sealing of DNA-DNA ends necessitates that the RNA primer is removed in its entirety before DNA ligation occurs (185). Keeping in mind that these three proteins contain PIP boxes and the homotrimer nature of PCNA, two models have been proposed to elucidate their PCNA-mediated coordination.

The first model suggests that these partner proteins sequentially bind to PCNA (Figure 5.1A). Their highly dynamic binding and release from the same or different PCNA monomer drives their coordination (137,318). In this context, one can envision that a single functional binding site on PCNA is enough for the coordinated actions of Pol δ, FEN1 and DNA Ligase 1. Support for the sequential binding model mainly comes from the recent
biochemical study of yeast OF maturation proteins where the researchers elegantly designed an approach to purify PCNA heterotrimers with only one or two functional binding sites (318). The study suggested that one functional binding site is sufficient for the coordinated action of Pol δ and FEN1 in nick translation, and therefore, concluded that their simultaneous binding is not required. However, the methodology does not directly show that this dynamic sequential switching occurs. Moreover, the study offers no insight into the coordination of DNA ligation during OF maturation. The low-resolution EM structure of an archael PCNA-Lig-DNA complex also supports this sequential switching model (326).

In contrast, the second model suggests that simultaneous binding of more than one partner protein to different PCNA monomers is necessary for a successful coordination (Figure 5.1B). Here, PCNA can be thought of as a “toolbelt” stably bound to, and moving with, more than one partner that can be readily available to perform their activities when needed (327). It follows then that more than one functional binding site is required for a successful coordinated OF maturation. The “toolbelt” model is supported by earlier biochemical studies with homodimeric bacterial (327,328) and heterotrimeric archael (329) sliding clamps in addition to the recent high-resolution kinetic studies of yeast proteins (151). In the latter study, the researchers performed millisecond kinetic studies to better understand PCNA coordination of Pol δ and FEN1 activities as well as the stimulatory effect of PCNA on their separate activities. Yet again, DNA Ligase 1 was not included in this study either.
Archeal PCNA, with its heterotrimeric nature, can coordinate different enzymes through each monomer. Thus, it has evolved to solve this coordination conundrum by opting for the “toolbelt” model (329-331). However, the eukaryotic case is more sophisticated, where the homotrimeric PCNA can use either of the models or a combination of various models to coordinate the various DNA processing pathways. Solving the puzzle of PCNA-mediated processes is not only vital to understand the recruitment of these proteins, but is also mechanistically interesting since PCNA serves as a mobile platform on which these proteins are plastically tethered and perform their designated functions.
Figure 5.1. The two proposed models for PCNA coordination of Okazaki fragment maturation. (A) Schematic describing the sequential switching model where Pol δ, FEN1, and LIG1 bind to, and release from, PCNA sequentially. This model suggests that the high dynamic association/dissociation of these proteins with/from PCNA drives their coordinated actions during OF maturation. (B) Schematic depicting the toolbelt model where the three enzymes are simultaneously bound to, and traveling with, PCNA occupying all three binding motifs of PCNA.

Insight into how these enzymes interact with PCNA came to light in the elegant work of Sakurai et al. in their structure of PCNA-FEN1 complex (274) (Figure 5.2A). In this structure, three FEN1 molecules were bound to the three subunits of PCNA trimer through their PIP boxes. The structure showed that FEN1 interaction with PCNA is primarily through FEN1 amino acids 336-356. These amino acids form a twin beta zipper separated by an alpha helix that form the major interactions with PCNA (Figure 5.2A). In fact, *in vivo* studies using mice model system showed that FEN1 mutations in this region, which disrupt its PCNA interaction while maintaining its nuclease activity intact, cause aneuploidy-associated cancer (332). The crystal structure further displayed that FEN1 can be maintained in two different orientations, which the authors ascribed to “active” and “inactive” conformers (274) (Figure 5.2B). These two conformers differ by the orientation of FEN1 with respect to PCNA, due to the different angle of rotation around the FEN1 hinge region (333-336) that is just before FEN1 PIP box. The inactive conformer showed FEN1 nuclease core being maintained at perpendicular orientation with respect to where the DNA axis would be located, thereby excluding FEN1 access to the DNA as it passes through PCNA core channel. In contrast, in the active conformation, FEN1 nuclease core runs parallel to the DNA axis and would be able to access the DNA and perform its catalytic
activity (Figure 5.2B). The authors further suggested that maintaining the inactive conformation could be advantageous when PCNA is required to slide, as the inactive conformer would not pose any steric clashing with PCNA sliding motion (274). With this structure in mind, a reasonable extrapolation would be a scenario in which PCNA acts as a toolbelt and simultaneously binds three different enzymes (Pol δ, FEN1 and DNA Ligase 1). These enzymes can all be tethered to a single PCNA trimer, with one enzyme in the active conformation, while the other two maintained in the inactive conformation until they are needed.

Further interesting insight into PCNA coordination came from the crystal structure of human DNA Ligase 1 (LIG1)-bound to DNA (185) (Figure 5.2C). The ligase structure was solved with the N-terminal domain truncated, thus the structure showed three domains; the DNA-binding domain (DBD) which forms the major interaction with the DNA, the adenylation domain (AdD) and an OB domain (OBD). The AdD and OBD are linked via a long linker and contain the six conserved motifs thus constituting the catalytic core of the ligase. The crystal structure revealed that the ligase completely encircles the nick DNA. SAXS structures with archael apo protein showed DNA Ligase 1 to have an extended conformer and suggested that its circularization requires its DNA binding (333). It is worthy to note that although the major interaction between DNA Ligase 1 and PCNA is through its PIP box, there are secondary interaction points in the DBD. The similarity in the ring shape and size between PCNA and DNA Ligase 1, as well as the initial biochemical characterization, suggests that PCNA and DNA Ligase 1 might interact in a 1-1 stoichoiometry (334). Moreover, it has been suggested that DNA Ligase 1 might first interact with PCNA through its N-terminal domain and the PCNA ring encircling DNA.
might then stimulate DNA Ligase 1 transformation from an extended conformer into the ring shape via interactions with the rest of DNA Ligase 1 most likely through its DBD (335). Therefore, one can envision that the ligase can be bound to PCNA in an extended inactive conformer on non-preferred DNA substrates, and consequently not hindering PCNA sliding or interaction with other proteins. However, when DNA Ligase 1 encounters its preferred nick DNA, it could undergo a transition to encircle the DNA. This ring-shaped DNA Ligase 1 can be imagined to occlude PCNA binding sites from other partners. Overall, these results present a complex platform that could be attributed to more than one model.
Figure 5.2. Crystal structures of human PCNA-FEN1 complex and human LIG1 bound to DNA. (A) A front view of the crystal structure of human PCNA (blue) bound to three FEN1 molecules (green, red and grey). Each PCNA monomer consists of two domains (A and B) with an interdomain connecting loop (IDCL). FEN1 interaction with PCNA is supported by the canonical FEN1 PIP box and PCNA IDCL and further contacts with FEN1 extended C-terminal. (PDB: 1UL1) (274) (B) A side view of the same crystal structure showing two FEN1 molecules in an active conformation (green and red), while the third is positioned in an inactive conformation (grey) parallel to the DNA and not hindering PCNA sliding on the DNA. (C) Crystal structure of human LIG1 completely encircling nicked DNA. The structure shows the three domains of LIG1; DBD (red), AdD (green) and OBD (yellow). The template strand (black) and nicked strand (grey) are shown with the AMP moiety (blue) (PDB: 1X9N ) (185).

Despite years of research into the enzymatic processes of each of the OF maturation proteins and their PCNA-orCHECARtigated coordination, many gaps in our understanding persist, and the debate between the two proposed models remain unsettled. What would the trigger for such protein switching from one enzyme to another is still elusive. Moreover, to date, no physical interaction among these three proteins has been reported to support a direct handoff mechanism in the absence of PCNA. Differential binding affinity to the DNA substrates might provide some answers to this question, although no systematic study of binding affinities of these enzymes to various intermediates has been conducted. Furthermore, there might be some unknown DNA-mediated indirect allosteric handoff mechanism currently uncharacterized. Here, we attempt to dissect this PCNA-coordinated mechanism at the single molecule level, particularly using smFRET assays. We present our preliminary data on human FEN1- and LIG1-coordinated PCNA partner switching.
5.3 Results

During our work characterizing FEN1 substrate recognition in the short flap- (Chapter 3) and long flap-pathways (Chapter 4), we developed single molecule FRET assay using internal-labeling scheme that reports on FEN1 intrinsic property of bending its DNA substrate. We utilized this labeling approach to probe the questions related to PCNA-mediated coordination of partner switching between FEN1 and LIG1. The labeling design therefore reports on this coordination from FEN1 substrate recognition perspective. Furthermore, the work presented here mainly targeted PCNA coordination at the level of FEN1 cleavage product of EQ or Non-EQ DF substrates. We believe this work will shed some new light and complement the existing work of PCNA coordination during nick translation between FEN1 and Pol δ.

5.3.1 PCNA stabilizes FEN1 bending of its product

While characterizing FEN1 product release mechanism (Chapter 3), we showed that FEN1 binding/bending of its product is stabilized by the presence of synthetic unpaired 1 nt 3’ flap. In other words, FEN1 binds/bends the product of non-equilibrated flap substrate (termed product) with higher affinity and stability as compared to the product of equilibrated flap substrates (termed nick). FEN1 bending of the product is characterized by a diffusion-limited association rate of \((0.42\pm0.08) \times 10^8 \text{ M}^{-1}\text{s}^{-1}\) with a fast dissociation rate
of $7.25\pm0.07 \text{ s}^{-1}$, yielding an equilibrium dissociation constant of $173\pm33 \text{ nM}$ (Figure 5.3A). In contrast, on a nicked DNA product, FEN1’s bending/unbending was faster than our temporal resolution (100 ms) (Figure 5.3B), which led to merging of the resulting FRET histograms and complicated assigning of the bent FRET state as well as percentage of bent DNA (as discussed in Chapter 3). As a result, the isotherm of bent DNA percentage versus FEN1 concentration represented a lower estimate of the equilibrium dissociation constant of $580\pm130 \text{ nM}$ (Figure 5.3B).

PCNA has been shown to stimulate FEN1 bending (293) and catalytic efficiency of single flapped substrates (151,316). Hence, we sought to characterize PCNA effect on FEN1 bending its product in both forms (product and nick). It is worthy to note that in the presence of PCNA, the FRET state of both DNA constructs remained constant as evident from the time traces and FRET histograms of DNA alone in the presence and absence of PCNA (data not shown); therefore, FRET change is attributed to FEN1 binding/bending its substrate. As both product and nicked DNA are biotinylated at the downstream side, which is used to immobilize the DNA to the surface, the upstream side is unblocked. PCNA can slide on dsDNA with a mean diffusion coefficient of $1.16\pm0.07 \mu\text{m}^2/\text{s}$ (336); it follows then that in our setup with short DNA constructs, PCNA would freely diffuse on/off the DNA. Keeping that in mind, FEN1 titrations to both DNA constructs were performed with the injection of 500 nM PCNA with each FEN1 concentration.

The presence of PCNA led to a higher affinity binding/bending of FEN1 to the product DNA. Plotting the bent product percentage against increasing FEN1 concentration resulted in the isotherm shown in Figure 5.3C. Fitting this isotherm to a hyperbolic function
generated an equilibrium dissociation constant of 7.2±1.0 nM indicating an ~20-fold increase in FEN1 affinity in the presence of PCNA (Figure 5.3A,C). The presence of PCNA also stabilized the bent conformer as FEN1 dissociation rate dropped to 0.58±0.04 s⁻¹ (Figure 5.3C) as compared to 7.25±0.07 s⁻¹ in the absence of PCNA (Figure 5.3A). FEN1 association rate on the other hand was not affected, which was unexpected as PCNA is believed to recruit FEN1 to the DNA substrate and increase its local concentration (337,338). However, in our experimental design, perhaps PCNA loading is not well optimized and does not describe the physiological state. Similarly, on the nicked DNA substrate, PCNA enhanced FEN1 affinity to the DNA as well as stabilized the bent conformer (Figure 5.3D). However, due to the fast bending/unbending in the absence of PCNA (Figure 5.3B), this enhancement of bending cannot be quantified. Overall, PCNA stabilized FEN1 interaction with its product DNA as expected from the formation of a bi-molecular protein complex on the DNA.
Figure 5.3. PCNA stabilizes FEN1 bending of the nick and product. (A) FEN1 bending efficiency of the Product\textsubscript{Internal}. Top: schematic showing FEN1 product resulting from the cleavage of non-equilibrating DF substrate. The DNA is doubly labeled with a Cy3 donor at position 12 on the downstream side and an Alexa Fluor 647 acceptor at position 15 on the upstream side. Bottom Left: representative single molecule time trace showing FEN1 bending/unbending Product\textsubscript{Internal} at a FEN1 concentration of 50 nM. Bottom Right: analysis of FEN1 bending association rate constant ($k_{\text{on-bending}}$) and dissociation rate constant ($k_{\text{off-unbending}}$). FEN1 dwell time analysis of the bent ($\tau_{\text{bending}}$) and unbent ($\tau_{\text{unbending}}$) conformers were determined by idealizing the smFRET time traces to two states. The populations of these dwell times were plotted and fitted to exponential function to calculate $k_{\text{bending}}$ and $k_{\text{unbending}}$ at each concentration. Error at each concentration corresponds to the standard deviation of the exponential fit. These values were plotted against FEN1 concentration to generate the graph shown here. $k_{\text{on-bending}}$ was calculated from the slope of $k_{\text{bending}}$ fit to a linear regression with the error of the fit reported. The reported $k_{\text{off-unbending}}$ is the mean of $k_{\text{unbending}}$ with the standard error of the mean reported. $K_{d-bending} = k_{\text{off-unbending}}/k_{\text{on-bending}}$. (B) FEN1 bending efficiency of the Nick\textsubscript{Internal}. Top: schematic showing FEN1 product resulting from the cleavage of equilibrated DF substrate with internal labeling scheme as described in A. Bottom Left: representative single molecule time trace showing FEN1 bending/unbending Nick\textsubscript{Internal} with fast transitions beyond our temporal resolution (100 ms). Bottom Right: FEN1 bending isotherm generated by plotting the percentage of bent DNA as a function of FEN1 concentration. At each FEN1 concentration, the percentage of bent DNA was estimated by fitting the FRET histograms to one or two Gaussian distributions. The fast transitions in smFRET traces resulted in an
averaging effect, which led to merging of the FRET histograms and complicated the estimation of the bent population. The isotherm was fit to a hyperbolic function to yield the reported $K_{d\text{-bending}}$ with the standard error of the fit shown. Due to the averaging effect, the reported $K_{d\text{-bending}}$ is a lower estimate. (C) FEN1 bending efficiency of the Product$_{\text{internal}}$ in the presence of 500 nM PCNA. Left: FEN1 bending isotherm as described in B. Right: analysis of FEN1 association/dissociation rates as described in A. (D) FEN1 bending efficiency of the Nick$_{\text{internal}}$ in the presence of 500 nM PCNA. Left: FEN1 bending isotherm as described in B. Right: analysis of FEN1 association/dissociation rates as described in A.

5.3.2 LIG1 competes with FEN1 to the nicked product

After investigating PCNA effect on FEN1 binding/bending of its product, we questioned whether LIG1 on its own could compete with FEN1 to the product in a non-PCNA coordinated manner. First, we started by assessing LIG1 ability to bind different DNA constructs; FEN1 preferred substrate with a short 5’ flap and 1 nt 3’ flap (DF-6,1), nick and product DNA. Surface plasmon resonance (SPR) experiments with these different DNA constructs and increasing concentrations of LIG1 (100, 500 and 1000 nM) showed that LIG1 bound the product DNA, nick DNA and the DF substrate. The dissociation phase upon buffer injection shows that LIG1 formed stable complexes with these DNA constructs. However, the maximum response units achieved indicate preferential association to the FEN1 product, then nick DNA and then DF substrate (Figure 5.4A).

In smFRET FEN1 bending assays, we proceeded using product DNA to assay the competition of FEN1 and LIG1 to this DNA construct. As was the case with PCNA, the addition of LIG1 to product DNA did not influence the FRET state (data not shown), and consequently, we attributed the change of FRET to FEN1 bending. Similar to previous
experiments, we maintained an injection of 500 nM LIG1 with each FEN1 concentration to allow for true competition to the DNA. In the presence of full length (FL) LIG1, FEN1 bending of the product DNA was characterized by a diffusion-limited association rate of \((0.19 \pm 0.02) \times 10^{7} \text{ M}^{-1} \text{s}^{-1}\) (Figure 5.4B) an \(~20\)-fold decrease as compared to FEN1 alone case (Figure 5.3A). This decrease indicates that LIG1 can actively compete FEN1 to the product DNA and inhibits its access. On the other hand, the presence of FL-LIG1 seemed to stabilize the bent conformer with a dissociation rate \((0.91 \pm 0.03 \text{ s}^{-1})\) (Figure 5.4B) that is \(~8\)-fold lower than that of FEN1 alone (Figure 5.3A). These association and dissociation rates yielded an equilibrium dissociation constant of \(491 \pm 61 \text{ nM}\), which represented an overall \(~3\)-fold lower affinity of FEN1 to the product DNA in the presence of FL-LIG1 (Figure 5.4B). While the witnessed decrease in FEN1 association rate due to LIG1 competition to the DNA seems logical, the increased stability in the dissociation rate seems puzzling. To date, no direct physical interaction between FEN1 and LIG1 has been reported. Moreover, a control experiment assessing FEN1 bending of DF-6,1 in the presence and absence of LIG1, where SPR established that LIG1 can bind to this substrate, showed practically no effect of LIG1 on FEN1 bending of this substrate (data not shown), indicating that direct physical interaction between FEN1 and LIG1 is highly unlikely.

The N-terminal domain of LIG1 contains the PIP box and it is believed to contribute to the majority of the interaction with PCNA (334). In preparation for investigating PCNA-coordinated partner switching between FEN1 and LIG1, we created LIG1 mutation deleting its N-terminal, referred to as \(\Delta\text{N-LIG1}\). In smFRET FEN1 bending assay parallel to that performed in the presence of FL-LIG1, we performed a competition assay in the
presence of ∆N-LIG1. In the presence of ∆N-LIG1, FEN1 association rate to its product ((0.62±0.20) x 10^7 M^-1s^-1) (Figure 5.4C) was comparable to that in the presence of FL-LIG1 case (Figure 5.4B). However, the dissociation rate (5.53±0.02 s^-1) (Figure 5.4C) was comparable to that in FEN1 alone case (Figure 5.3A). Therefore, the resulting equilibrium dissociation constant at 897±298 nM (Figure 5.4C) was ~5-fold higher than in FEN1 alone case (Figure 5.3A). Taken together, these results indicate that LIG1 acts as a competitive inhibitor of FEN1 to the product DNA as evident from the effect on association rate, while the differential effect on the dissociation rate is not fully elucidated and requires further investigation.

Figure 5.4. LIG1 competes FEN1 off the product DNA. (A) SPR experiment assessing FL-LIG1 binding to three DNA constructs; DF6,1 (grey), Product (blue) and Nick (Cyan). The experiment was performed with
3 FL-LIG1 titrations; 100 nM (top), 500 nM (middle) and 1000 nM (bottom). This binding experiment shows FL-LIG1 preferentially binding the Product over the Nick and DF6,1. (B) FEN1 bending efficiency of the Product$_{internal}$ in the presence of 500 nM FL-LIG1. Analysis of FEN1 association/dissociation rates are as described in Figure 5.3A. (C) FEN1 bending efficiency of the Product$_{internal}$ in the presence of 500 nM ΔN-LIG1. Analysis of FEN1 association/dissociation rates are as described in Figure 5.3A.

5.3.3 PCNA-mediated product handoff

Having established the separate effect of both PCNA and LIG1 on FEN1 bending, we aimed to dissect the PCNA-mediated FEN1-LIG1 partner switching from the FEN1 product bending perspective. Similar to previous experiments, PCNA and LIG1 were maintained each at 500 nM concentration in all FEN1 titrations. In the presence of PCNA and FL-LIG1, FEN1 association rate to the product at $(0.38 \pm 0.03) \times 10^7 \text{M}^{-1}\text{s}^{-1}$ (Figure 5.5A) was comparable to that in the presence of FL-LIG1 (Figure 5.4B), indicating that FL-LIG1 inhibition of FEN1 access to the product DNA is neither stimulated nor inhibited in the presence of PCNA. On the other hand, FEN1 dissociation rate from its product DNA in the presence of PCNA and FL-LIG1 $(0.31 \pm 0.03 \text{ s}^{-1})$ (Figure 5.5A) was comparable to the rate achieved in the presence of PCNA (Figure 5.3C). These association and dissociation rates generated an equilibrium dissociation constant of $82 \pm 11 \text{nM}$ (Figure 5.5A), which represents an ~2-fold decrease as compared to the FEN1 alone case (Figure 5.3A) and an ~7-fold increase to the FEN1 in the presence of PCNA case (Figure 5.3B). Therefore, PCNA stabilization of FEN1 on the product DNA seems to counteract LIG1 competition to the same DNA, overall favoring FEN1 binding to the product DNA.
While the stabilization of FEN1 dissociation rate by PCNA (Figure 5.3B) offers an indirect indication of the formation of PCNA-FEN1 bi-molecular complex formation, our experimental design does not give any indication whether PCNA-LIG1 complex is formed or not. Hence, we sought to disrupt this complex formation using LIG1 deletion mutation, ΔN-LIG1, instead of FL-LIG1 in the preceding experiment. This substitution of ΔN-LIG1 failed to induce any significant difference of FEN1 association and dissociation rates (Figure 5.5B) as compared to those in the parallel experiment in the presence of FL-LIG1 (Figure 5.5A). While LIG1 access to the DNA in both forms is indirectly supported by the decrease in FEN1 association rate, our current design still lacked any indication to whether the PCNA-LIG1 complex is being formed or not. We suspected that PCNA might not be properly loaded on the DNA and its stabilization of FEN1 association rates might stem from a physical interaction through FEN1 C-terminal without accessing the DNA and possibly incapable of coordinating the partner switching between FEN1 and LIG1. Thus, we shifted our focus to optimizing PCNA loading on the DNA.

Figure 5.5. FEN1-LIG1 competition to the product in presence of PCNA. (A) FEN1 bending efficiency of the Product\textsubscript{Internal} in the presence of 500 nM FL-LIG1 and 500 nM PCNA. Analysis of FEN1 association/dissociation rates are as described in Figure 5.3A. (B) FEN1 bending efficiency of the
Product\textsubscript{Internal} in the presence of 500 nM ΔN-LIG1 and 500 nM PCNA. Analysis of FEN1 association/dissociation rates are as described in Figure 5.3A.

5.3.4 PCNA loading via RFC in bulk and single molecule

To mimic physiological conditions, we opted for proper loading of PCNA on nicked DNA through its clamp loader, RFC. As mentioned in Chapter 1, RFC is a heteropentameric complex with all of its subunits containing an AAA+ ATPase related domain (138,139). The N-terminal domain of its biggest subunit, RFC1, has been shown \textit{in vitro} to bind non-specifically to the DNA (339); thus, it might lower the efficiency of properly loading PCNA to the 3’terminus of primer/template junctions. Its deletion mutation, however, was shown to enhance PCNA proper loading (340). Therefore, we opted to express and purify RFC complex with deletion mutation of the N-terminal domain of RFC1, termed as ΔN-RFC. The detailed expression and purification of ΔN-RFC is described in Chapter 2. This purification of ΔN-RFC yielded a complex with high purity as shown in Figure 5.6A, but the protein complex seemed to elute in two peaks (P1 and P2) following size exclusion chromatography. To examine which of these two peaks contained the proper RFC complex with the correct subunit stoichiometry, we performed a PCNA loading experiment onto nicked DNA substrate as shown in Figure 5.6B. The nicked substrate was labeled with Cy3 in the template strand, while PCNA was labeled with Cy5-Maleimide non-specifically. The DNA was blocked from both ends with biotin-\textit{NeutrAvidin} interaction on one end and digoxigenin-anti-digoxigenin antibody on the other side (Figure 5.6B). Upon PCNA interaction with Cy3-Nick DNA, a FRET is observed. This FRET efficiency increases as
PCNA is loaded properly in the presence of ATP and ΔN-RFC collected from P2 peak (Figure 5.6C). Hence, we concluded that only P2 peak contained the proper ΔN-RFC complex and proceeded with it for the next experiment.

Figure 5.6. ΔN-RFC purification and PCNA loading in bulk. (A) The N-terminal deletion of RFC1 subunit in complex with the other 4 subunits (termed ΔN-RFC) was cloned, expressed and purified as described in Chapter 2. The protein eluted in two peaks, P1 (blue) and P2 (magenta) on size-exclusion chromatography. These two peaks represent different complex stoichiometry. P2 was determined as the peak with proper stoichiometry since it showed stimulation of PCNA loading as described in B. (B) Schematic illustration bulk FRET assay for PCNA loading via ΔN-RFC on nick DNA. The Cy3 labeled nicked DNA is blocked from both sides with biotin/NeutrAvidin one-side and Dig/anti-Dig interactions on the other side. Loaded Cy5-PCNA (blue) via ΔN-RFC (grey) would result in FRET. (C) Steady state fluorescence emission spectra of the Cy3-Nick DNA in 4 conditions (as color coded in the figure) at 535 nm excitation.
To properly load PCNA in smFRET assay we designed a double-labeled nicked DNA construct similar to the one used in bulk assay but with a Cy3 donor and Alexa Fluor 647 acceptor in the same positions we used for internal-labeling scheme (Figure 5.7A). To dually block the ends of this nicked DNA construct, we first incubated the DNA with anti-dig antibodies for 10 mins at room temperature followed by immobilization on the surface through biotion/NeutrAvidin interaction. For PCNA loading, PCNA and ΔN-RFC each at 20 nM concentration were pre-incubated at 37°C for 10 mins in an ATP-containing buffer, and then injected into the flow cell and incubated for further 10 mins. Finally, the flow cell was washed with imaging buffer. FEN1 was then titrated and DNA bending was characterized by examining the FRET histograms and time traces.

The nicked DNA FRET histogram centered around 0.24 with a single conformer as evident from the time traces verifying that Dig/Anti-Dig interaction did not affect the FRET state (Figure 5.7B,C). Loading PCNA to this nicked DNA did not shift the FRET state either as evident from both the FRET histograms and time traces further confirming that FRET change can be safely attributed to FEN1 recognizing and bending the nicked DNA (Figure 5.7B,C). Upon titrating FEN1 to the loaded PCNA-nick complex, the FRET histogram showed two separable peaks, one centered around 0.24 corresponding to the unbent nicked DNA and a significant population with a peak centered around ~0.5 which corresponds to the fully bent conformer (Figure 5.7B). In contrast, in FEN1 titration experiments in the absence of PCNA at similar concentration, the FRET histogram showed a single peak corresponding to unbent nicked DNA with no significant bent population (data not shown). Taking a closer look at the time traces revealed two distinct time traces
profiles (Figure 5.7C). The first one showed fast bending/unbending behavior, reminiscent of FEN1 binding/bending the nick DNA. More importantly, the second type of time traces showed a very stable FRET state of 0.5 corresponding to the fully bent state. On the other hand, increasing FEN1 concentration by 10 folds resulted in FRET histograms with two merged peaks centered around 0.24 and 0.42 (Figure 5.7B). This peak merging stemmed from fast transitioning between bent and unbent states as evident from the time traces (Figure 5.7C), thus leading to an averaging of the FRET states. This averaging effect is reminiscent of FEN1 binding/bending the nick DNA in the absence of PCNA (Figure 5.3B), albeit with higher population of the bent FRET state at this concentration.

Taken together, these results can be interpreted to stem from a heterogeneity in the DNA molecules. It seems that PCNA did not load on the majority of DNA molecules. In these molecules, FEN1 bending is characterized by fast bending/unbending as seen in a subpopulation of the time traces at 20 nM FEN1. On the remaining DNA molecules, PCNA appears to have been properly loaded as FEN1 bending of these molecules was very stable. In those instances, PCNA loaded via ΔN-RFC stabilized FEN1 bending (Figure 5.7B,C) even further than that in the case where PCNA was recruited from solution (Figure 5.3D). Increasing FEN1 concentration augmented its association to the nick DNA, but with a majority of molecules lacking PCNA, the bending was not stabilized and it resembled the FEN1 alone case (Figure 5.3B).
Figure 5.7. **PCNA loading via ΔN-RFC at single molecule.** (A) Schematic showing the dually labeled nick DNA by the internal-labeling scheme immobilized onto the surface with biotin/NeutrAvidin interactions and blocked from the other end with Dig/anti-Dig interactions. PCNA is loaded via ΔN-RFC as described in Chapter 2. Upon FEN1 addition, the DNA is bent leading to a FRET change from a low state to a higher state. (B) FRET histograms of $\text{Nick}_{\text{Internal}}$ in different conditions: 1) DNA alone, 2) in the presence of loaded PCNA, after washing excess protein, 3) in the presence of 20 nM FEN1 and 4) in the presence of 200 nM. (C) Representative single molecule time traces corresponding to the four conditions described in B. At 20
nM FEN1, two traces are shown to reflect the majority of traces showing fast transitions (left) and a minority population showing single FRET state corresponding to the bent conformer (right).

In conclusion, while the data presented here is still preliminary, the experimental approach to load PCNA is promising. Further optimization is required to obviate the heterogeneity and achieve better PCNA loading. Nevertheless, PCNA-mediated stabilization of FEN1 bending of the nick or product DNA is conclusive as expected from a bi-molecular complex formation. The effect of PCNA on FEN1 association rate as would be expected with PCNA recruiting FEN1 to the DNA would only be answered once PCNA loading is properly optimized. The competition between LIG1 and FEN1 to the product follows with their expected roles on this DNA substrate. However, further experiments are needed to dissect their PCNA-mediated partner switching.
Chapter 6

6. Discussion and Future Directions*

Throughout DNA metabolic pathways, a plethora of toxic structures is generated that if left unprocessed would lead to genomic instability and several disease states. Nucleases are at the center of processing these aberrant structures. A subfamily that is relevant for DNA metabolism is the structure-specific 5’ nucleases superfamily. This superfamily consists of four highly conserved endo- or exo-nucleases involved in DNA replication, repair or recombination. These family members in human are flap endonuclease 1 (FEN1), exonuclease 1 (EXO1), xeroderma pigmentosum complementation group G (XPG) and gap endonuclease 1 (GEN1). They hydrolyze phosphodiester bonds that are one nucleotide into the 5’ end of ssDNA/dsDNA junctions, and hence they are called 5’ nucleases (155,341). One interesting aspect of these nucleases is that they share significant structural similarity but have evolved to process a diverse set of substrates with high degree of specificity on their cognate substrates, while maintaining residual cleavage efficiency of other similar substrates. FEN1 incises 5’ flapped substrates, EXO1 performs exonuclease cleavage on double strand breaks and nicks, XPG cleaves bubble structures and GEN1 resolves Holliday junctions (160,341). This paradox that highlights the mechanistic details unifying the cleavage site on diverse substrate structures in these 5’ nucleases has motivated remarkable research into the structure and function of each of these proteins.
In this dissertation, we focused on dissecting FEN1’s sophisticated substrate recognition and highly selective catalysis in both short- and long-flap pathways as an archetype of 5’ nucleases. Moreover, we attempted to provide an insight into its regulation through PCNA-mediated coordination with other protein partners.

6.1 Uncovering FEN1 mechanistic details

Critical cellular processes such as DNA replication and repair are regulated by molecular properties encoded in their interacting macromolecules whereby distinct dynamic conformations correspond to different functionalities (342,343). The mechanisms for these dynamic changes that occur during macromolecular interactions are the subject of intense interest with two major proposed mechanisms, 'induced fit' and 'conformational selection' (344-348). Studies on multiple enzyme-ligand systems have revealed that the ligand binds to the active conformation of the enzyme, indicating that functional or conformational selection is in play (343,345). However, the high precision required for DNA replication
and repair has consistently raised issues of whether or not these might involve unusual mechanisms of chemistry or physics. In this context, how DNA-repair enzymes specifically recognize and remove damage in DNA is a decades-long debate. Does the damage destabilize the DNA duplex leading to disruption of the DNA structure (extrahelical base flipping or DNA bending) before its subsequent capture by the repair enzymes or do these enzymes actively sculpt the DNA as part of their recognition of the damage?

6.1.1 FEN1 remarkable substrate recognition: an active multi-step vetting process

With our FEN1 single-molecule results, a picture emerges of induced conformational changes to both the substrate and the protein playing key roles in stabilizing a transition state that has been thoroughly vetted using multiple checks and is poised for catalysis with remarkable specificity (Figure 6.1). In this process, FEN1 can differentiate between substrates whose incision is good for the cell (cognate) or toxic (noncognate) even if these substrates have small differences in their binding affinities. Active DNA bending does not create a significant energy barrier as evident by FEN1 diffusion-limited on-rates. We propose that different members of the 5’ nuclease family might share similar DNA-bending-induced disorder-to-order transitioning but differ in the mechanisms that couple this transitioning with active site assembly. In FEN1, the coupling of protein transitions of the 3’ flap-binding pocket and the 5’ flap-binding helical gateway with DNA sculpting uncovers how dynamic protein segments are critical contributors to substrate binding and
catalytic selection. This elaborate allosteric mechanism that underlies the active site assembly provides regulatory framework for cognate substrate selection.

As part of the active DNA sculpting process, we observed single-molecule measurements consistent with a mutual induced-fit mechanism, with the protein bending the DNA and the bent DNA inducing protein conformational change (Figure 6.1). Substrate distortion by $\sim 100^\circ$ and DNA-induced protein conformational changes are features that extend beyond 5’ nucleases (349-351). More generally, the FEN1-type induced-fit mechanism may be central to detecting chemically subtle but biologically critical differences between correct and incorrect substrates for multiple DNA and RNA processes. In this paradigm, nuclease precision in replication comes from an induced-fit mechanism that regulates the compatibility of the distorted DNA conformer with active-site assembly and its off rate to allow cleavage of cognate substrates while avoiding noncognate substrates (Figure 6.1).

The diffusion-limited bending of the cognate substrate by FEN1 and its cleavage from the first encounter represents a practically perfect precision reaction with rates that are limited by diffusion. The stochastic cleavage behavior of noncognate substrates after multiple cycles of DNA bending and dissociation of FEN1 has a fundamental bearing on how enzyme specificity is understood. Cases in which an enzyme encounters noncognate substrates and cleaves them after multiple trials might mislead the interpretations of substrate specificity in classical biochemical techniques that monitor product formation. Furthermore, multiple attempts to cleave noncognate substrates are likely to be insignificant inside the cell. These findings advance our insight into a previously
unidentified mechanism in structure-specific nucleases for extreme specificity towards their cognate substrates inside the cell.

The 3’ flap binding pocket is distant from the active site, raising the question why FEN1, along with some other structure-specific nucleases, utilizes long-range DNA-induced conformational coupling, in contrast with local coupling as observed in EXO1 (156,274,278,349-351). FEN1 cleaves 5’ flaps containing RNA, DNA or mismatches of various lengths. We anticipate that long-range DNA-induced conformational coupling could provide a mechanism that enhances flexibility in nuclease substrates. The action of 3’ flap-induced protein ordering as a key step that locks the FEN1 interaction with the junction could provide the advantage of limiting the sampling time between the disordered protein form and noncognate substrates that could otherwise lead to nonspecific cleavage. This active site control via a long-range induced-fit mechanism suggests why mutations distant from the FEN1 active site have a dramatic effect on genomic stability and disease states (264,352). Thus, evolutionary selection against toxic and mutagenic DNA instability may have developed the unusual DNA-induced conformational coupling seen in FEN1 as a previously unrecognized part of repair and replication nuclease fidelity.
Figure 6.1. Model for control of catalytic selectivity by the DNA mutual-induced fit mechanism in FEN1. DNA sculpting: FEN1 actively bends a variety of structures to verify the key features of its cognate DF substrates of fully paired ss/dsDNA nick junction, threaded 5’ flap into the cap-helical gateway and 3’ flap. Protein ordering: FEN1 actively pulls the 3’ end of the nick junction to create a 3’ flap that drives the protein ordering, which in turn orders the active site and locks the DNA conformation. Decision: the active site and locked DNA conformer are always in catalytically competent form in cognate substrate, while they are primarily in catalytically incompetent form in noncognate substrates (no 5’ flap threading, no 3’ flap, mispaired junctions) or in the case of FEN1 mutants (R47A, K93A and R100A). DNA release or catalysis: the DNA shifts or unpairs to move the scissile phosphate into the active site for cleavage as probed by the flap/junction positioning-residue Y40, while in noncognate substrates FEN1 promotes DNA dissociation.

6.1.2 A comprehensive kinetic scheme of FEN1 reaction

Our work characterizing FEN1 on cognate short flap substrates culminated in a comprehensive kinetic scheme of FEN1 reaction (Figure 6.2). FEN1 binds/bends its
cognate substrate with diffusion-limited rates, then undergoes disorder-to-order conformational changes to assemble the active site and place the scissile phosphate in the active site. It then catalyzes a fast incision reaction of the 5’ flap one nucleotide inside the junction, releasing the flap instantaneously. In contrast, the nicked DNA product was shown to act as a competitive inhibitor of FEN1 in multiple turnover assays. Here, we demonstrated using smFRET and smPIFE that FEN1 releases its nicked product in two steps; a fast unbending step and a slow release step from an unbent conformation. These results were subsequently supported by bulk chase experiments (353). The finding that the nicked product is released in two steps could indicate a product-mediated hand-off mechanism. Since FEN1 holds onto the DNA in unbent form, we speculate that perhaps it contacts part of the duplex while allowing DNA Ligase 1 to access the newly formed nick and complete OF maturation.
Figure 6.2. Comprehensive kinetic scheme of FEN1 reaction. FEN1 binds and bends a short DF substrate at diffusion-limited kinetics and commits to cleavage within the first encounter. After cleavage, the 5’ flap is released instantaneously while nicked product release occurs in two steps. FEN1 can rebind/rebend the nicked product at diffusion-limited kinetics, albeit with a lower on-rate than the substrate. With longer flaps, FEN1 can miss cleavage and dissociate from the bent substrate, requiring more than one attempt at cleavage.

6.1.3 Where we heading

Despite years of biochemical, structural and the more recent single molecule research to understand the mechanistic details of FEN1, there are several key mechanistic details still missing. FEN1’s disorder-to-order conformational change is fascinating; however, the basis of such transition is elusive. Ongoing work in our group, and in collaboration with
Jaremko’s group (KAUST), is focused on deciphering this mechanism employing nuclear magnetic resonance (NMR) spectroscopic approach. The long-term goal is to assign full FEN1 spectra and, in conjunction with large-scale computational work, to design potential FEN1-targeted inhibitors. This will be coupled with extensive biochemical, biophysical and single molecule experimentation to scope out potential anti-cancer drugs as FEN1 implication in cancer is well documented (354,355).

On the other hand, the mechanistic details we learn about FEN1, an archetype of 5’ nucleases, could greatly advance our appreciation of the superfamily unifying features as well as those specific to FEN1 (160). Previous and current work in our group aimed/aims at broadening the lessons we learned from FEN1 to other superfamily members (228). Previous work showed that EXO1 actively bends its substrate in diffusion-limited kinetics similar to FEN1. Other efforts in our group have been centered around deciphering the Holliday junction resolution by GEN1 at single molecule level (356). We have expressed and purified XPG, with some preliminary characterizations on its mechanism of substrate recognition. Future work will be focused on dissecting the role of XPG in nucleotide excision repair pathway with multiprotein fluorescence approach rather than the single protein mechanistic detail-oriented approach.
6.2 FEN1 in the context of Okazaki fragment maturation

Okazaki fragment maturation involves removal of RNA/DNA primers in the form of 5' single-stranded flaps to precisely create nicks that are ligated to complete lagging strand DNA synthesis. Defective or incomplete processing of 5' flaps can interfere with DNA replication and promote sequence expansions, especially of repeat sequences, among other outcomes that have detrimental impacts on genome integrity and stability (157,357,358). The critical importance of accurate and efficient OF maturation is highlighted by the fact that deletion mutations of enzymes primarily responsible for this process are linked to cancer predisposition and neurodegenerative diseases (258,262,299,359).

It has been proposed that when Pol δ generates a short 5' flap at a downstream OF, there is tight coupling and highly efficient hand-off of the DNA to FEN1 for cleavage (150,151). However, there is also evidence that longer 5' flaps are formed. For example, in wild type *S. pombe* cells, 5' flaps visualized by electron microscopy have a mean length of 51 nt, with some exceeding 100 nt (177); deletion of FEN1 in *S. cerevisiae* results in duplications that indicate flap lengths as long as 100 nt as well (178,360). In this case, the evidence indicates that RPA and Dna2 helicase/nuclease are involved in flap removal in addition to FEN1 (153,162,302,361). Processing of long flaps has the benefit of removing both the RNA and the error-containing DNA portion of the primer generated by low fidelity Pol α, but that comes with the cost of significant DNA re-synthesis and potential delays in OF maturation given the larger number of proteins required to perform the task. Not surprisingly, the long-flap pathway is considered a secondary or back up option to the more predominant short-
flap pathway (242,362). The choice between these pathways can have important consequences, and therefore likely involves coordination between the proteins involved and may be subject to regulation as well (164,181,182,363).

6.2.1 FEN1 triggers long flap-pathway by missing cleavage of long flaps

Characterizing FEN1 activity on short and long flaps in order to elucidate the events leading to OF maturation by the short- versus long-flap pathways, we interestingly found that in bulk experiments, FEN1 exhibits only 2-fold lower single turnover and steady state cleavage rates on 60 nt flaps compared with 6 nt flaps, which suggests that it is capable of acting on longer flaps by itself during OF maturation. However, single molecule analysis of FEN1 activity revealed key transient events where its actions differ on short- versus long-flap DNA substrates. First, in smFRET DNA bending experiments we found that substrate recognition, in particular bending efficiency, is not affected by flap length. This finding is supported by the FEN1-flap DNA crystal structure, which shows that most of the interactions of FEN1 are with the duplex portion of the DNA, not the flap (156,157). In smFRET single turnover cleavage experiments, again we found little difference in the rates of multiple steps in the reaction with increasing flap length (Chapter 4). Therefore, we concluded that flap length has no significant impact on bending of the DNA substrate, cleavage chemistry, 5’ flap release and unbending of the nicked DNA product by FEN1 during the reaction. Importantly, we did find that while FEN1 always cleaves a short flap within the first encounter with the substrate, it increasingly misses cleavage as the flap
length increases (Chapter 4). As noted above, FEN1 binds and bends both short and long flaps with similar efficiency. In addition, the average lifetime of the bent FEN1-DNA conformer is the same whether FEN1 misses or completes flap cleavage (Chapter 4). Hence, the missed cleavage opportunities indicate that longer flaps pose a challenge for conversion of the bent FEN1-DNA complex into a catalytically active state. Threading a long 5' flap through the cap-helical gateway to position the scissile phosphate in the active site may pose difficulties that increase the odds of FEN1 dissociating from DNA rather than cleaving the flap.

6.2.2 FEN1 and RPA interplay on DF substrates

To determine the fate of longer flaps that can escape FEN1 cleavage, we examined the interplay between FEN1 and RPA on substrates with varying flap lengths. We found that RPA strongly inhibits cleavage of long-flap substrates, but it has no effect on short-flap substrates at the bending, cleavage or product release steps in the reaction (Chapter 4). Notably, while the presence of RPA weakens FEN1 affinity for a long-flap substrate, the enzyme can still access the DNA and bend it. Cleavage appears to be inhibited mainly because the FEN1-DNA complex cannot achieve a stable, catalytically active conformation with RPA bound to the flap (Chapter 4). Thus, when FEN1 misses cleaving a long flap, RPA has the opportunity to bind it and block subsequent attempts at cleavage even if FEN1 rebinds the DNA. Given the high affinity and stability of RPA interaction with ssDNA, we expect that this protein-DNA complex will remain in pause mode until RPA is actively
displaced by Dna2. In short, the competition between FEN1 and RPA determines the choice between the short- and long-flap pathways. The ability of FEN1 to bind and bend RPA-bound long-flap substrate could facilitate substrate transfer to FEN1 as Dna2 shortens the 5’ flap and displaces RPA.

In conclusion, we found that as the flap gets longer, the probability of escaping cleavage increases, giving RPA the chance to get involved and trigger the long-flap pathway. What might be the structural/dynamics basis for FEN1 dissociation from a long-flap substrate, and how Dna2 acts on a DNA substrate possibly crowded with RPA and FEN1, are intriguing follow-up questions to be addressed in future studies.

6.2.3. Follow-up projects

Current work in our lab is focused at expressing and purifying human Dna2 protein and aims at deciphering the interplay between FEN1, RPA and Dna2 using our established single molecule assays. This work will also focus on mechanistic structural details pertaining to Dna2 recognition of long flap substrates.

Within this paradigm of regulating the choice between short- and long-flap pathways, the key protein players have been shown to undergo posttranslational modifications such as phosphorylation and acetylation. It has been hypothesized that perhaps these posttranslational modifications, especially acetylation by p300 acetyltransferase, might be the trigger for the switch between the two pathways. Acetylation of the key nucleases FEN1
and Dna2 exhibited opposite effect; while the acetylation of FEN1 showed ~90% diminished nuclease activity (181,364), the acetylation of the latter showed several fold increase in its helicase and nuclease activities (182). Moreover, previous work showed that acetylation of yeast Pol δ’s third subunit enhances its strand displacement activity (365). Similar results were also shown for Pif1 helicase acetylation augmenting its helicase activity. A general picture that emerges is that the acetylation of OF maturation proteins might trigger the long-flap pathway presumably in active gene regions of the genome to ensure higher fidelity DNA replication. In this context, our lab is interested in studying the effect of acetylation on FEN1, RPA and Dna2 and the interplay between these acetylated proteins to infer key regulatory mechanisms of their activities.

6.3 PCNA-mediated product handoff between FEN1 and LIG1

In addition to posttranslational modifications, FEN1 has been shown to be regulated by several protein-interacting partners, such as PCNA (332,354,366). In fact, PCNA acts as a platform and coordinator of several DNA metabolic pathways including DNA replication and repair (136,338,367,368). Its involvement in orchestrating the OF maturation pathway is well established (369). It interacts and regulates most of its key enzymatic activities including Pol δ, FEN1 and DNA Ligase 1 (136,316,334,337,338,367,368). However, the mechanism of coordination given its homotrimeric nature in eukaryotes and the vast set of
protein partners has been puzzling. The debate between the two main proposed models; the sequential switching of partners due to their dynamic interactions and release model and the toolbelt model, is far from being settled (148,151,318,319,338,370). There are several lines of evidence supporting either model, but each has its own shortcomings. Moreover, the existing studies focus on PCNA coordination of Pol δ and FEN1 activities and their switching mechanisms, but they do not extend to the role of DNA Ligase 1 (151,318).

Therefore, we sought to shed some light on PCNA coordination of FEN1 and DNA Ligase 1 activities at single molecule level by following FEN1 bending kinetics of its product. Our work characterizing FEN1 product release mechanism suggested that FEN1 releases its nicked product from an unbent state. This allowed us to speculate that it is possible that FEN1 holds on to the product for subsequent handoff to the ligase. The extended state of the DNA perhaps is crucial for the recognition of nicked product by DNA Ligase 1. The FEN1 100˚ bent nicked product is much more distorted than what is expected for the ligase. However, we could not speculate how this would play out in the presence of PCNA.

6.3.1 FEN1 acts as a product chaperone stabilized by PCNA

FEN1 dissociation rates in the presence of PCNA alone demonstrated that PCNA stabilized FEN1 binding/bending of its nicked and non-equilibrated product (Chapter 5). In this context, FEN1 can be viewed as a product chaperone holding onto the product for the
subsequent ligation step and PCNA is helping stabilize this chaperone. This could be a regulatory mechanism ensuring the faithful handoff of DNA between the two enzymes and protecting it from unwanted nuclease activity or excessive strand displacement (134,319,368). PCNA is known to recruit FEN1 to OF maturation sites and increase its local concentration (316). In our setup, we were unable to draw any conclusions regarding this recruitment, as we do not have evidence that PCNA is properly loaded on the DNA. Consequently, FEN1 association rates were not influenced by the presence of PCNA suggesting that in our setup perhaps PCNA is not recruiting FEN1. Future work employing a PIFE effect upon PCNA binding to the DNA or fluorescently labeled PCNA can conclusively confirm PCNA loading on DNA before further probing FEN1 bending.

6.3.2 LIG1 competes with FEN1 to the product

Our SPR experiments with FL-LIG1 on different DNA constructs surprisingly showed that LIG1 slightly prefers binding to non-equilibrated product as compared to nicked product. Although this was surprising as the preferred in vivo substrate of LIG1 is a perfect nick, if it is confirmed, it might unravel an interesting mechanistic product-mediated interplay between FEN1 and LIG1. Perhaps, FEN1 maintains the 3’ flap in an unpaired state that stabilizes its interaction with the DNA, as evident from the comparison of its bending efficiencies of product versus nick substrates, and hands-off this unpaired 3’ flap product to LIG1 in that conformation. The comparison between FEN1 association rates in the absence and presence of LIG1 (in the case of both FL-LIG1 and ΔN-LIG1) showed a
remarkable decrease as LIG1 competes with FEN1 to the product DNA. The stabilization of FEN1 on DNA product witnessed in the presence of FL-LIG1 is indeed puzzling. It is tempting to speculate that there might be a bi-molecular FEN1/FL-LIG1 complex formation, no such direct physical interactions have been previously reported and we confirmed the lack of this direct physical interaction using FEN1 flap substrate (DF-6,1) where we failed to witness any stabilization of the bent conformer. Overall, we believe that LIG1 inhibits FEN1 access to the DNA in a possible product-mediated handoff in the absence of PCNA.

6.3.3 PCNA seemingly favors FEN1

To understand the PCNA-coordinated product handoff from FEN1 to LIG1, we witnessed that in the presence of PCNA and both forms of LIG1 (FL-LIG1 and ΔN-LIG1), LIG1 seemed to compete with FEN1 to the product DNA, while PCNA might be stabilizing FEN1 on the DNA. Thus, PCNA seems to counteract LIG1 competition of FEN1 to the product DNA. This could be viewed as a regulatory mechanism in favor of FEN1 inhibiting unwanted access of the ligase to the product until it is ready for ligation. However, there are several caveats in our setup. First, the reaction is performed in presence of Ca\(^{2+}\) rather than Mg\(^{2+}\) and this could have possible effects on the mode of binding of LIG1 to its substrate. Moreover, we failed to discriminate between the effects of FL-LIG1 and ΔN-LIG1 in this process, suggesting that perhaps the PCNA-LIG1 complex is not being formed on the DNA. This is a likely scenario if PCNA is not properly loaded on the DNA. In this context, perhaps LIG1 is forming a complex with PCNA in solution and thus sequestering LIG1 from accessing the DNA construct. As for FEN1 stabilization by PCNA most likely
stemming from PCNA-FEN1 complex formation, we can envision a scenario where PCNA binds to FEN1 through its C-terminal domain without being properly loaded on the DNA; in other words, it is possible that in these experiments PCNA is being recruited by FEN1 and not the other way around.

6.3.4 Loading PCNA via RFC is a bottleneck: what comes next?

To conclusively probe the PCNA-mediated partner switching between FEN1 and LIG1, PCNA has to be pre-loaded on the DNA before interacting proteins are added. Our approach to load PCNA via RFC seemed promising as we witnessed partial stabilization of FEN1 on nick substrate. This approach needs further optimization with possibly exploring other blocking methods of either end, confirming PCNA loading as mentioned above, or optimizing the loading buffer conditions. Once PCNA loading is optimized, further work employing single molecule-coupled cleavage and ligation assay can address PCNA coordination of these two enzymatic activities in real time.

Other interesting facets of this regulatory mechanism is focused on the ligase itself. LIG1 interaction with PCNA has been shown to extend beyond its N-terminal domain to include further interactions within its DBD. The DBD also mediates the initial interactions with the DNA through the minor grooves (185). These minor grooves are exposed in the FEN1:DNA crystal structure (156). Therefore, if DNA Ligase 1 were to bind a FEN1-DNA complex, it would only sterically clash with FEN1 near its cap. This led the researchers to
suggest that perhaps an order-to-disorder transition of the capped helical gateway occurs, which in turn would destabilize FEN1 binding to the DNA and mediate the handoff to the ligase in a “passing the baton” manner similar to that proposed for base excision repair (156,371). This could be aided and coordinated by PCNA. From this perspective, it is interesting to investigate whether the DBD alone can push FEN1 off its product or if all the domains of ligase are required to mediate the exchange. We have expressed and purified ligase DBD and future work characterizing this aspect would be inspiring.

Finally, but not less importantly, deciphering the mechanistic details of DNA Ligase 1 encircling the DNA would offer one remarkable piece into solving the puzzle of PCNA-coordinated OF maturation. Future work in our group intends to approach LIG1 circularization on nick DNA through single molecule research of different mutations of the protein, in particular the deletion mutation of its flexible linker between the AdD and OBD domains. Other approaches include site-specific fluorescent dual labeling of the ΔN-LIG1 through NHS labeling at the N-terminal domain of DBD and aldehyde tag labeling in the OBD domain. This approach would allow for real-time monitoring of LIG1 encircling its DNA through smFRET.
6.4 Concluding Remarks

This dissertation is aimed at inspiring the readers to appreciate the complexity and fascination of DNA replication as one of the most vital processes in all domains of life. Chapter 1 took the reader on a journey through more than 30 years of research into the initiation, elongation and termination steps of DNA replication with brief descriptions of the foremost biochemical and structural milestones while highlighting major gaps in our knowledge. From there onwards, the focus shifted to Okazaki fragment maturation centered around its key player, FEN1.

Chapter 3 dissected the extreme precision and specificity of this perfect catalyst suggesting an induced-fit mechanism where the enzyme actively induces a severe kink in the DNA to a 100° angle while undergoing key protein conformational changes from a disordered to ordered state. Through protein mutational work and disruption of key features in FEN1 ideal substrate, a multiple enzymatic vetting process of the substrate was revealed. This chapter further highlighted a lock-down mechanism whereby the enzyme cleaves its cognate substrate from first encounter but avoids off-target substrates that eventually can be cleaved after several hit and miss trials. The work also constructed a comprehensive kinetic scheme of FEN1 reaction where the rate-limiting step is its product release, which occurs in two steps.

After establishing key details in substrate recognition and catalytic mechanism in FEN1 short-flap pathway, Chapter 4 delved into deciphering what happens when the perfect
catalyst misses its substrate. The long-flap pathway is thought of as a backup system and only triggered in rare cases. We showed for the first time that FEN1 plays a role in triggering this long-flap pathway through missing the cleavage of longer flaps. Furthermore, the Chapter concluded with protein-protein interplay between FEN1 and RPA on short- and long-flap pathways suggesting a possible DNA-mediated handoff mechanism between Dna2/RPA in the long-flap pathway back to FEN1 in the short-flap pathway.

The dissertation then concluded with preliminary results on PCNA-mediated FEN1 and DNA Ligase 1 partner switching with promising future directions to look forward to.
BIBLIOGRAPHY


Statement of Author’s Contributions

I would like to acknowledge the contribution of the following people to the research presented in this dissertation.

**Professor Samir M. Hamdan’s** Lab (King Abdullah University of Science and Technology, Thuwal, Saudi Arabia)

**Fahad Rashid** provided the expression clones and established the expression and purification protocols of human FEN1, PCNA, RFC, and LIG1. He also established single molecule cleavage assay using flap-labeling scheme and provided the data described in Figures: 3.2-B,C,D; 3.3-D,E,F; 3.7-B,C,D,E; and 4.3.

**Professor Satoshi Habuchi’s** Lab (King Abdullah University of Science and Technology, Thuwal, Saudi Arabia)

**Paul D. Harris** performed the confocal-based FEN1 FRET experiments and contributed with Figures: 3.2-E,F and 3.7-A.

**Professor Manju M. Hingorani’s** Lab (Wesleyan University, Middletown, CT)

**Bo Song** performed FEN1 ensemble single turnover and steady state kinetic assays described in Figures: 3.4-A; 3.5-D and 4.4-E.
ACCOMPLISHMENTS

List of publications:

- Raducanu VS#, Rashid F#, **Zaher MS#**, Hamdan SM. (2019). A novel fluorescent signal transducer embedded in a DNA aptamer paves the way for versatile metal ion detection. (*In final stages for submission*).

- Rashid F#, Raducanu VS#, **Zaher MS#**, Tehseen M, Habuchi S, Hamdan SM. (2019). Initial state of DNA-dye complex sets the stage for protein induced fluorescence modulation. (Accepted in *Nature communications*).


*These authors contributed equally to the work.*

**List of presentations:**

- “Missed cleavage opportunities by FEN1 lead to Okazaki fragment maturation via the long-flap pathway”

  Abstract Selected: Oral Presentation

  Fusion Conference, DNA and Interacting Proteins as Single Molecules - *In Vitro* and *In Vivo*

  21st-24th Feb 2018

  Cancun, Mexico.

- “Missed cleavage opportunities by FEN1 lead to Okazaki fragment maturation via the long-flap pathway”

  Abstract Selected: Poster Presentation

  Fusion Conference, 3rd DNA Replication/Repair Structures and Cancer Conference

  11th-15th Feb 2018

  Cancun, Mexico.
“A kinetic insight into FEN1 escaping its flap and the significance of the secondary pathway of Okazaki Fragment maturation”

Biological and Environmental Sciences and Engineering (BESE) Division Seminar
13th March 2017
KAUST, Thuwal, Saudi Arabia.