Using single molecule fluorescence to study substrate recognition by a structure-specific 5’ nuclease

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

© December 2018

Fahad Rashid

All rights reserved
The dissertation of Fahad Rashid is approved by the examination committee.

Committee Chairperson: Prof. Samir M. Hamdan

Committee Members: Prof. Satoshi Habuchi (KAUST), Prof. Enzo Di Fabrizio (KAUST), Prof. Joseph Laporo (Harvard University)
ABSTRACT

Using single molecule fluorescence to study substrate recognition by a structure-specific 5' nuclease

Fahad Rashid

Nucleases are integral to all DNA processing pathways. The exact nature of substrate recognition and enzymatic specificity in structure-specific nucleases that are involved in DNA replication, repair and recombination has been under intensive debate. The nucleases that rely on the contours of their substrates, such as 5' nucleases, hold a distinctive place in this debate. How this seemingly blind recognition takes place with immense discrimination is a thought-provoking question. Pertinent to this question is the observation that even minor variations in the substrate provoke extreme catalytic variance. Increasing structural evidence from 5' nucleases and other structure-specific nuclease families suggest a common theme of substrate recognition involving distortion of the substrate to orient it for catalysis and protein ordering to assemble active sites.

Using three single-molecule (sm)FRET approaches of temporal resolution from milliseconds to sub-milliseconds, along with various supporting techniques, I decoded a highly sophisticated mechanism that show how DNA bending and protein ordering control the catalytic selectivity in the prototypic system human Flap Endonuclease 1 (FEN1). Our results are consistent with a mutual induced-fit
mechanism, with the protein bending the DNA and the DNA inducing a protein-conformational change, as opposed to functional or conformational selection mechanism. Furthermore, we show that FEN1 incision on the cognate substrate occurs with high efficiency and without missed opportunity. However, when FEN1 encounters substrates that vary in their physical attributes to the cognate substrate, cleavage happens after multiple trials.

During the course of my work on FEN1, I found a novel photophysical phenomena of protein-induced fluorescence quenching (PIFQ) of cyanine dyes, which is the opposite phenomenon of the well-known protein-induced fluorescence enhancement (PIFE). Our observation and characterization of PIFQ led us to further investigate the general mechanism of fluorescence modulation and how the initial fluorescence state of the DNA-dye complex plays a fundamental role in setting up the stage for the subsequent modulation by protein binding. Within this paradigm, we propose that enhancement and quenching of fluorescence upon protein binding are simply two different faces of the same process. Our observations and correlations eliminate the current inconvenient arbitrary nature of fluorescence modulation experimental design.
ACKNOWLEDGEMENTS

“It takes a village to raise a child” is a proverb that comes to my mind when I think of this journey throughout my PhD. My first and foremost appreciation and gratitude goes to my supervisor Prof. Samir M. Hamdan for his guidance and expert advice, and for giving me the opportunity to undertake and be part of this fascinating learning and research process. Secondly, I am deeply thankful to my fellow PhD. student Manal S. Zaher for being part of all the work in this thesis and for all her support during this journey. My heartfelt appreciation goes to the many collaborators I have worked with and learned from, especially Dr Susan Tsutakawa (LBNL), Prof. John Tainer (MD Anderson) and Prof. Ivaylo Ivanov (GSU). I would also like to extend my gratefulness to my collaborators here in KAUST, especially Prof. Satoshi Habuchi, Paul Harris, Prof. Lukasz Jeremko, Prof. Mariusz Jeremko, Prof. Enzo De Fabrizio and Dr. Monica Marini.

This work would not have been possible without the help and support of my colleagues in Samir M. Hamdan Lab. You have made my experience in lab a memorable and enjoyable one far beyond science did. Special mentions: our founding members Mohi, Luay and Mostafa, our research scientists Masateru Takahashi, Mohammed Sobhy, and Muhammed Tehseen and last but not least, my friend and collaborator Vlad. I am deeply grateful to my friends Hanan and Junaid for their endless support and encouragement.

On a more personal note, I am eternally indebted and grateful to my parents and my siblings (Amir and Iram) for their constant support and belief in me. You
have been a source of inspiration and encouragement. Special “appreciation” to my two-year-old niece, Sarah, whose constant support during the last two years made this thesis possible.

Lastly, I would like to acknowledge KAUST for providing generous support for conducting this thesis and the late King Abdullah for his wisdom and vision in establishing this Institute.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EXAMINATION COMMITTEE PAGE</strong></td>
<td>2</td>
</tr>
<tr>
<td><strong>ABSTRACT</strong></td>
<td>3</td>
</tr>
<tr>
<td><strong>ACKNOWLEDGEMENTS</strong></td>
<td>5</td>
</tr>
<tr>
<td><strong>TABLE OF CONTENTS</strong></td>
<td>7</td>
</tr>
<tr>
<td><strong>LIST OF ABBREVIATIONS</strong></td>
<td>10</td>
</tr>
<tr>
<td><strong>LIST OF ILLUSTRATIONS</strong></td>
<td>13</td>
</tr>
<tr>
<td><strong>MAIN TEXT</strong></td>
<td></td>
</tr>
<tr>
<td>Chapter 1</td>
<td></td>
</tr>
<tr>
<td>1. Using single molecule fluorescence to study substrate recognition</td>
<td>15</td>
</tr>
<tr>
<td>by structure-specific 5’ nuclease</td>
<td></td>
</tr>
<tr>
<td>1.1 Nucleases</td>
<td>15</td>
</tr>
<tr>
<td>1.1.1 Nucleases: the phosphoryl transfer reaction</td>
<td>16</td>
</tr>
<tr>
<td>1.1.2 Mechanism</td>
<td>16</td>
</tr>
<tr>
<td>1.1.3 Metal ions and nucleases</td>
<td>18</td>
</tr>
<tr>
<td>1.1.4 Structure-specific nucleases</td>
<td>20</td>
</tr>
<tr>
<td>1.2 The superfamily of 5’ nucleases</td>
<td>23</td>
</tr>
<tr>
<td>1.2.1 Flap Endonuclease 1</td>
<td>24</td>
</tr>
<tr>
<td>1.2.2 Exonuclease 1</td>
<td>39</td>
</tr>
<tr>
<td>1.2.3 XPG</td>
<td>42</td>
</tr>
<tr>
<td>1.2.4 GEN1</td>
<td>45</td>
</tr>
<tr>
<td>1.3 Unifying features of 5’ nucleases</td>
<td>49</td>
</tr>
<tr>
<td>1.3.1 DNA bending</td>
<td>49</td>
</tr>
<tr>
<td>1.3.2 Disorder-to-order transition</td>
<td>51</td>
</tr>
<tr>
<td>1.3.3 Selectivity of primary substrates and substrate promiscuity</td>
<td>54</td>
</tr>
<tr>
<td>1.3.4 Protein-protein interactions and post translational modifications</td>
<td>54</td>
</tr>
<tr>
<td>1.4 Use of fluorescence to study nucleases</td>
<td>56</td>
</tr>
<tr>
<td>1.4.1 Principle of fluorescence</td>
<td>56</td>
</tr>
<tr>
<td>1.4.2 FRET theory</td>
<td>61</td>
</tr>
<tr>
<td>1.4.3 Protein-induced fluorescence enhancement</td>
<td>66</td>
</tr>
<tr>
<td>1.4.4 Why single molecule spectroscopy</td>
<td>68</td>
</tr>
<tr>
<td>1.5 Motivation of study</td>
<td>74</td>
</tr>
<tr>
<td>Chapter 2</td>
<td></td>
</tr>
<tr>
<td>2. Single-molecule FRET unveils induced-fit mechanism for substrate</td>
<td>76</td>
</tr>
<tr>
<td>selectivity in flap endonuclease 1*</td>
<td></td>
</tr>
<tr>
<td>2.1 Abstract</td>
<td>76</td>
</tr>
<tr>
<td>2.2 Introduction</td>
<td>77</td>
</tr>
<tr>
<td>2.3 Results</td>
<td>79</td>
</tr>
<tr>
<td>2.3.1 FEN1 actively bends the DNA</td>
<td>79</td>
</tr>
<tr>
<td>2.3.2 FEN1 never misses cleavage of its correct substrate</td>
<td>91</td>
</tr>
<tr>
<td>2.3.3 FEN1 actively creates a 3’flap in equilibrated DF substrate</td>
<td>98</td>
</tr>
</tbody>
</table>
2.3.4 Substrate verification by 3’flap-induced protein ordering ................. 104
2.3.5 FEN1 avoids off-target DNA cleavage in the DNA lockdown step 109
2.3.6 Role of active site residues in positioning the 5’flap and junction . 114
2.4 Discussion .................................................................................................. 117
2.5 Methods.................................................................................................... 121
  2.5.1 Protein purification ............................................................................. 122
  2.5.2 DNA substrate preparation ................................................................. 122
  2.5.3 TIRF-based smFRET .......................................................................... 123
  2.5.4 Confocal-based smFRET ................................................................. 125
  2.5.5 Single-molecule cleavage assays ..................................................... 128
  2.5.6 Time-resolved bulk FRET measurements ...................................... 130
  2.5.7 Surface plasmon resonance (SPR) binding ..................................... 131
  2.5.8 Steady-state bulk cleavage assays .................................................. 132
  2.5.9 Molecular dynamics simulations .................................................... 133

Chapter 3
3. Missed cleavage opportunities by FEN1 lead to Okazaki fragment maturation via the long-flap pathway* .................................................. 137
  3.1 Abstract ................................................................................................. 137
  3.2 Introduction ......................................................................................... 138
  3.3 Results .................................................................................................. 145
    3.3.1 Effect of 5’ flap length on the DNA bending activity of FEN1 ....... 145
    3.3.2 A kinetic mechanism for short 5’ flap recognition and cleavage by FEN1 ............................................................... 150
    3.3.3 Effect of 5’ flap length on the FEN1 kinetic mechanism .......... 165
    3.3.4 Coordination between RPA and FEN1 on short- and long-flap substrates ................................................................. 175
  3.4 Discussion ............................................................................................ 182
  3.5 Materials and Methods ....................................................................... 185
    3.5.1 Proteins expression and purification ............................................. 185
    3.5.2 DNA substrates ............................................................................. 186
    3.5.3 TIRF-based smFRET ................................................................. 187
    3.5.4 PIFE bulk and single molecule experiments ............................... 191
    3.5.5 Ensemble single turnover and steady state cleavage assays ...... 192
    3.5.6 Steady state bulk cleavage in the presence of RPA ................. 193

Chapter 4
4. Initial state of the DNA-dye complex sets the stage for on demand fluorescence modulation upon protein binding* .................................. 196
  4.1 Abstract ............................................................................................... 196
  4.2 Introduction .......................................................................................... 197
  4.3 Results .................................................................................................. 204
    4.3.1 A novel DNA-mediated bidirectional fluorescence modulation observed through FEN1/DF system ........................................ 204
    4.3.2 The initial lifetime of Cy3-DNA sets the stage for fluorescence enhancement or quenching in the presence of RPA ................. 213
    4.3.3 Insights into the structural properties of the DNA-dye complexes. 219
4.3.4 PIFQ and PIFE can be used at the single molecule level to follow catalysis, DNA conformational changes and protein binding ............. 227

4.4 Discussion ........................................................................................................... 234

4.5 Materials and Methods ......................................................................................... 241
  4.5.1 Proteins expression and purification ................................................................. 241
  4.5.2 DNA oligos and substrates ............................................................................... 241
  4.5.3 Steady state fluorescence .................................................................................. 243
  4.5.4 Absorbance measurements .............................................................................. 245
  4.5.5 Quantum Yield measurements ......................................................................... 246
  4.5.6 Theoretical non-radiative lifetime loss calculations ........................................... 246
  4.5.7 Time-resolved fluorescence lifetime measurements ........................................... 248
  4.5.8 Determination of the hyperbolic viscosity dependence of fluorescence lifetime and decoupling of rates ........................................... 250
  4.5.9 Single molecule fluorescence measurements ................................................... 252

Chapter 5

5. Discussion .................................................................................................................. 257
  5.1 Conformational dynamics and substrate recognition ............................................. 257
  5.2 FEN1 actively sculpts its cognate substrate ............................................................ 259
  5.3 5' Flap threading as a part of an overall FEN1 vetting process ............................... 262
  5.4 Coupling of substrate and enzyme conformational changes ................................... 263
  5.5 Avoiding off-target cleavage by probabilistic active site assembly ...................... 264
  5.6 Backup pathway for FEN1 mediated Okazaki fragment maturation ..................... 267
  5.7 FEN1 binding and cleavage of long flap substrates ............................................... 269
  5.8 FEN1 missing cleavage prompts RPA-mediated second nuclease pathway ........... 270
  5.9 Use of single molecule fluorescence to answer biological questions ................. 273
  5.10 Understanding fluorescence modulation of cyanine dyes ................................... 274
  5.11 Characterization of protein-induced fluorescence modulation ............................ 276
  5.12 PIFE and PIFQ both depend on initial fluorescence state of Cy3-DNA complex ... 277
  5.13 Characterization of initial fluorescence state of Cy3-DNA complex .................... 279
  5.14 Adaptation of fluorescence modulation at single molecule level ......................... 282
  5.15 Concluding remarks ........................................................................................... 283

BIBLIOGRAPHY ........................................................................................................... 285

Statement of Author’s Contributions ............................................................................. 309

Accomplishments .......................................................................................................... 312
## LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</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>FEN1</td>
<td>Flap Endonuclease 1</td>
</tr>
<tr>
<td>EXO1</td>
<td>Exonuclease 1</td>
</tr>
<tr>
<td>XPG</td>
<td>Xeroderma pigmentosum group G protein</td>
</tr>
<tr>
<td>GEN1</td>
<td>Flap endonuclease GEN homolog 1</td>
</tr>
<tr>
<td>RPA</td>
<td>Replication protein A</td>
</tr>
<tr>
<td>Dna2</td>
<td>DNA Replication Helicase/Nuclease 2</td>
</tr>
<tr>
<td>Lig1</td>
<td>DNA ligase 1</td>
</tr>
<tr>
<td>PCNA</td>
<td>Proliferating cell nuclear antigen</td>
</tr>
<tr>
<td>H2TH</td>
<td>Helix-two turn-helix motif</td>
</tr>
<tr>
<td>HLH</td>
<td>Helix-loop-helix motif</td>
</tr>
<tr>
<td>PCA</td>
<td>Proto-catechuic acid</td>
</tr>
<tr>
<td>PCD</td>
<td>Protocatechuate-3,4-dioxygenase</td>
</tr>
<tr>
<td>PEG</td>
<td>Polyethylene glycol</td>
</tr>
<tr>
<td>SF-substrate</td>
<td>Single flap substrate</td>
</tr>
<tr>
<td>Term</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------------------------------------------</td>
</tr>
<tr>
<td>DF-substrate</td>
<td>Double flap substrate</td>
</tr>
<tr>
<td>EQ-DF-substrate</td>
<td>Equilibrating Double flap substrate</td>
</tr>
<tr>
<td>NEQ-DF-substrate</td>
<td>Non-Equilibrating Double flap substrate</td>
</tr>
<tr>
<td>HJ</td>
<td>Holliday junction</td>
</tr>
<tr>
<td>HR</td>
<td>Homologous recombination</td>
</tr>
<tr>
<td>E. coli</td>
<td><em>Escherichia coli</em></td>
</tr>
<tr>
<td>dsDNA</td>
<td>Double-stranded DNA</td>
</tr>
<tr>
<td>KDa</td>
<td>Kilo Dalton</td>
</tr>
<tr>
<td>SPR</td>
<td>Surface plasmon resonance</td>
</tr>
<tr>
<td>FRET</td>
<td>Förster Resonance Energy Transfer</td>
</tr>
<tr>
<td>smFRET</td>
<td>Single molecule Förster Resonance Energy Transfer</td>
</tr>
<tr>
<td>PIFE</td>
<td>Protein-induced fluorescence enhancement</td>
</tr>
<tr>
<td>PIFQ</td>
<td>Protein-induced fluorescence quenching</td>
</tr>
<tr>
<td>PAGE</td>
<td>Polyacrylamide Gel Electrophoresis</td>
</tr>
<tr>
<td>TIRF</td>
<td>Total-internal-reflection-fluorescence</td>
</tr>
<tr>
<td>EMCCD</td>
<td>Electron Multiplying Charge Coupled Device</td>
</tr>
<tr>
<td>ns</td>
<td>Nanoseconds</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>HMM</td>
<td>Hidden Markov Modeling</td>
</tr>
<tr>
<td>BSA</td>
<td>Bovine serum albumin</td>
</tr>
<tr>
<td>PBS</td>
<td>Phosphate buffered saline</td>
</tr>
</tbody>
</table>
LIST OF ILLUSTRATIONS

Figure 1.1. Mechanism of two metal ion dependent nuclease mediated DNA hydrolysis. ................................................................. 17
Figure 1.2. Structure-specific nucleases involved in various DNA metabolic pathways. ................................................................. 22
Figure 1.3. FEN1 cleavage reaction. ................................................................. 26
Figure 1.4. Structure of FEN1 .................................................................... 32
Figure 1.5. Two nucleotide unpairing ................................................................ 36
Figure 1.6 FEN1 protein-interaction network ................................................. 56
Figure 1.7. Energy landscape of a cyanine dye .................................................. 58
Figure 1.8. FRET as a molecular ruler ............................................................... 64
Figure 1.9. Schematic showing protein- induced fluorescence enhancement ... 67
Figure 1.10. Ensemble versus single molecule detection .................................... 71
Figure 2.1 Active junction bending by structure-specific 5’nucleases .................. 83
Figure 2.1.1 Bending of equilibrated and non-equilibrated DF-6,1 by FEN1 ..... 85
Figure 2.1.2 Flap substrates exist as a stable extended conformer ................. 86
Figure 2.1.3 MD simulations of the conformational states and DNA bending energy of nick and various flap structures ..................................... 87
Figure 2.1.4 Ensemble assays support single molecule results ............................ 89
Figure 2.1.5 Active bending of nicked DNA by EXO1 ....................................... 90
Figure 2.2. Cleavage of cognate substrate by FEN1 ........................................ 94
Figure 2.2.1 Single-molecule cleavage of cognate substrates and probing conformational changes to FEN1 by viscogens ..................... 96
Figure 2.2.2 Controls for assigning donor loss to the cleavage and immediate departure of the 5’flap ......................................................... 98
Figure 2.3. Active sculpting of the 3’end of a nick Junction creates a 3’flap and drive protein ordering .................................................... 101
Figure 2.3.1 All histograms corresponding to the data shown in Figure 2.3B ... 104
Figure 2.4. Verification of the bent DNA conformer by the 3’flap-induced protein ordering .............................................................. 107
Figure 2.4.1 Bending kinetics of various noncognate substrates ...................... 108
Figure 2.5. Cleavage of noncognate substrates by FEN1 ............................... 111
Figure 2.5.1 FEN1 cleavage of various noncognate substrates ....................... 114
Figure 2.6. Role of active-site residues in active positioning of the 5’flap and the junction ................................................................ 116
Figure 2.7. Sequences of DNA constructs used in this study ............................ 121
Figure 3.1. FEN1 actions in Okazaki fragment maturation ............................... 143
Figure 3.2. Effect of 5’ flap length on FEN1 DNA bending activity ................. 149
Figure 3.3. Single molecule and bulk cleavage kinetics of FEN1 on short DF- substrates .................................................................. 151
Figure 3.3.1. FEN1 bending of the nick DNA supports internal-labeling cleavage ................................................................ 154
Figure 3.4. Kinetic scheme for short 5’ flap recognition and cleavage by FEN1. ................................................................................. 159
Figure 3.4.1. Lower salt concentration stabilizes FEN1 bending of its products. .................................................................................................................. 162
Figure 3.5. Kinetic scheme for the FEN1 reaction. ........................................ 164
Figure 3.6. Effect of 5’ flap length on FEN1 cleavage activity. ...................... 168
Figure 3.6.1. Further characterization of missed cleavage events. .................. 171
Figure 3.6.2. Kinetic parameters of FEN1 substrate bending and catalytic efficiency on DNA versus RNA EQ DF substrates. ............................... 174
Figure 3.7. Coordination between FEN1 and RPA actions on short- and long-flap substrates........................................................................................................ 178
Figure 3.7.1. RPA does not affect FEN1 product release. ............................... 181
Figure 3.8. Substrates used in this study.......................................................... 187
Figure 4.1. Energy landscape of a cyanine dye............................................. 200
Figure 4.1.1. Structures of the different fluorophores used in this study.......... 201
Figure 4.2. A novel DNA-mediated bidirectional fluorescence modulation observed through FEN1/DF system ......................................................... 210
Figure 4.2.1. PIFQ likely stems from modulation of photoisomerization rates .. 212
Figure 4.3. The initial lifetime of Cy3-DNA sets up the stage for the fluorescence enhancement or quenching in the presence of RPA. ....................... 216
Figure 4.3.1. Steady state measurements correlate with time-resolved results. .................................................................................................................. 218
Figure 4.4 Insights into the structural properties of the DNA-dye complexes. .. 223
Figure 4.4.1 Decoupling the rates generating the initial lifetime. ................... 226
Figure 4.5. Single-labeled fluorescence modulation (PIFQ and PIFE) can be used at the single molecule level. ............................................................. 231
Figure 4.5.1. Examples of single molecule traces supporting Figure 4.5 .......... 233
Figure 4.6. List of all the DNA oligos and substrates used in this study. ......... 243
Chapter 1
1. Using single molecule fluorescence to study substrate recognition by structure-specific 5’ nuclease

1.1 Nuclease

DNA encodes the heritable instructions for the growth, evolution, development and proliferation of all life forms. Thus, the need for DNA replication and its maintenance through repair presents in itself a basis of life (1). The accurate replication of DNA is an elaborate process that is highly conserved in all domains of life. The role of the 5’-3’ endo/exonuclease in excising the primer and the 3’-5’ exonuclease activity of the DNA polymerase in removing misincorporated nucleotides renders the nuclease activities to be equally important to the DNA polymerization activity during DNA replication. Furthermore, the integrity of DNA is under constant threat from endogenous and exogenous factors that introduce wide range of DNA lesions (2). Nuclease execute the steps that eradicate the damaged parts during DNA repair. Nuclease, therefore are integral in all aspects of genome maintenance during replication, recombination and repair.
1.1.1 Nucleases: the phosphoryl transfer reaction

The choice of DNA as a genetic material is not arbitrary. Extreme stability is among the myriad of properties that make DNA ideal as a genetic material. The half-life ($t_{1/2}$) of a DNA phosphodiester bond at room temperature in neutral conditions is around 30 million years (3), thus making it extremely resistant to natural breakdown. However, DNA needs to be amenable to wide variety of processes that involve hydrolysis of its phosphodiester bond under physiological conditions. Nature has solved this paradox by evolving a set of the enzymes called nucleases, which catalyze the hydrolytic degradation of a phosphodiester backbone of DNA, or RNA molecules, by up to $10^{17}$-fold. Nucleases are usually proteins but can include RNA catalysts (Ribozymes) as well (4, 5).

1.1.2 Mechanism

The basic mechanism of all nucleases involves the cleavage of either of the two phosphodiester P—O bonds (Figure 1.1) (5), which proceeds via general acid-base catalysis assisted by metal ions. In the first step of this reaction, the nucleophile, usually a water molecule, is activated via deprotonation by the general base. In the second step, the leaving moiety is protonated by the general acid, thus accelerating the product formation. Either associative ($S_{N2}$) or dissociative ($S_{N1}$) pathways can be used for phosphodiester cleavage. However due to the extreme stability of the DNA phosphodiester bond and its poly-anionic nature, only the $S_{N2}$ associative pathway is favored in the nuclease-mediated hydrolysis (6). The $S_{N2}$
mechanism in the nuclease-mediated hydrolysis reaction proceeds in three steps (Figure 1.1). In the first step, deprotonation of water by one of the two active site metal ions (depicted in Figure 1.1 as $M_1^{2+}$) activates it to act as a nucleophile. In the next step, the penta-covalent intermediate is formed, which is highly unstable due to its extreme negative charge. Lastly, breakage of the scissile phosphate takes place, with the second metal ion acting to stabilize the negative charge on the leaving group, thus facilitating its departure (5). Cleavage of the “a” P—O bond (depicted in Figure 1.1) generates a 5’ phosphate and 3’ hydroxyl groups. This cleavage is favored evolutionary, since it leaves groups that other enzymes can act upon. The selective cleavage of bond “a” of the two P—O bonds requires the active site of the enzyme to orient the nucleophilic group to be on the 5’ side of the phosphodiester bond (Figure 1.1) (4, 5).

![Figure 1.1. Mechanism of two metal ion dependent nuclease mediated DNA hydrolysis.](image)

While a base-deprotonated water molecule serves as the most common nucleophile, it is not universally used by all nucleases. Amino acids Serine, tyrosine and histidine residues have all been observed to act as a nucleophiles (7,
8). Similarly, the 3’ hydroxyl group of DNA/RNA can also act as a nucleophile in multiple cleavage pathways including mRNA splicing, DNA strand transfer, or hairpin formation. RNases on the other hand use 2’ hydroxyl ions or even free ribonucleotides as nucleophiles (9, 10).

1.1.3 Metal ions and nucleases

Metal ions serve two primary purposes in accelerating the hydrolysis of a phosphodiester bond (11, 12). Firstly, they act as a general base to activate the attacking nucleophile and secondly, they stabilize the highly charged penta-covalent intermediate and further neutralize the charge on the leaving oxyanion moiety, thus accelerating the forward reaction of bond cleavage (4, 5, 11). Two metal ion-mediated enzymatic catalysis was first proposed based on the observation of two metal ions in the active site of alkaline phosphatase as well as the 3’-5’ exonuclease of E. coli DNA polymerase I structure (13-15). Since then, it has been shown that catalysis via two metal ions is utilized by most nucleases and all polymerases studied to date. Furthermore, this mechanism is also utilized by catalytic RNA, where metal ions can substitute for protein side chains as general acids and bases. In the pre-transition state, two metal ions are around 4Å apart and are invariably coordinated by an aspartate residue and non-bridging oxygen molecules of the scissile phosphate (4, 5). Further coordination of the metal ions is provided by other aspartate residues of the protein. Subsequently, the metal ions move closer (around 3.5Å apart) in order to stabilize the transition state complex.
This results in the nucleophile moving within 2.5 Å distance from the scissile phosphate, which is sufficient for inducing hydrolysis (4, 5, 16, 17). The high sensitivity of the two Mg\(^{2+}\) ions for the ligand geometry and the electrostatic environment makes them ideal to mediate scissile phosphate cleavage with exquisite specificity. Although Mn\(^{2+}\) ions can often replace Mg\(^{2+}\) ions in the active site, this comes on the price of relaxing the specificity of the reaction. Ca\(^{2+}\) ions on the other hand can act as a binding- but not a catalytic-mimic of Mg\(^{2+}\) ions since they cannot move closer than 3.8 Å towards each other (15,16).

Less commonly used mechanisms by nucleases include using either one metal ion or no metal ions at all. In case of one metal ion, the water-activating metal ion is absent. Therefore, either residues, such as Histidine, deprotonate the water molecule, as shown in \(\beta\beta\alpha\)-Me nucleases, or one of the protein residues itself can serve as a nucleophile, as shown in Tyrosine HuH nucleases. The stringency on the identity of metal ions used in this mechanism is lenient and a wide variety of metal ions can serve this purpose including Ca\(^{2+}\), Mg\(^{2+}\), Mn\(^{2+}\), and Zn\(^{2+}\) (16, 17). The last category of nucleases is metal independent. These include DNases and RNases. In case of RNases, 2'-OH is used as a nucleophile while one of the protein resides, such as Histidine, can serve as a nucleophile in the cases of DNases. The Penta-covalent intermediate is stabilized by one of the negatively charged residues of the nuclease (4, 5, 10).
1.1.4 Structure-specific nucleases

Structure-specific nucleases (SSNs) are a class of nucleases that recognize the tertiary structure of DNA, rather than its sequence. These nucleases play a key role in all aspects of DNA metabolism (18-21). Their catalytic repertoire can include incision of large number of aberrant DNA structures ranging from ssDNA- and/or ssRNA-flaps, overhangs, nicks and four way Holliday junctions (Figure 1.2). These aberrant structures are natural intermediates during replication, repair and recombination, but are extremely toxic (20, 21). They break the continuity of the heritable duplex DNA and they impose a significant impediment to replication and transcription. It is astonishing as to how these nucleases recognize their substrates with exquisite specificity based solely on the DNA structure itself rather than its sequence or even the presence of modified nucleotides (20) (Figure 1.2). The first enzyme known to possess a structure-specific nuclease activity was the bacterial DNA polymerase I, that cleaves the 5’ flap of the ssDNA protruding from a ss/dsDNA nick junction using its 5’-3’ exonuclease/endonuclease activity (22). Subsequently, a similar 5’ flap endonuclease, DNAse IV now known as flap endonuclease (FEN1), was discovered and characterized (23). This was followed by the discovery of multitude of other structure-specific nucleases, including those with 3’ flap cleavage activity like XPF-ERCC1 (24). Due to the fact that these nucleases rely on the indirect readout of a DNA structure, these nucleases can potentially be promiscuous in their nuclease activity. Thus, the precise need to control their substrate selection and incision site is central to the efficiency and fidelity of
these nucleases. The importance of this control can be gauged from the fact that alteration of the activity of any of these nucleases is associated with human diseases (25-27). However, although these nucleases have determinative primary substrate preferences, some do catalyze the incision of other substrates as well (Figure 1.2). The relevance of this more relaxed substrate specificity or residual activities on other substrates remains largely unknown.
Figure 1.2. Structure-specific nucleases involved in various DNA metabolic pathways. For each pathway, the nuclease(s) involved are mentioned and a schematic of the DNA intermediary structure is represented with the site of cleavage.
1.2 The superfamily of 5' nucleases

A subfamily of structure-specific nucleases that is most relevant for DNA metabolic pathways is the 5' nucleases superfamily. This superfamily is composed of four members involved in DNA replication, repair and recombination. These family members are flap endonuclease 1 (FEN1), exonuclease 1 (EXO1), xeroderma pigmentosum complementation group G (XPG) and gap endonuclease 1 (GEN1). These nucleases recognize the 5' end of a ss/dsDNA junction and catalyze the incision reaction of a phosphodiester bond 1 nt inside the 5' end of the junction (28, 29). Despite the significant protein structural conservation among the family members, these nucleases cleave a diverse set of substrates with utmost specificity. FEN1 catalyzes the incision of 5' flap substrates, EXO1 performs exonuclease incision on double strand breaks and nicks, XPG cleaves bubble structures and GEN1 resolves Holliday junctions (Figure 1.2). This paradox that highlights the mechanistic details unifying the structural features and yet diversifying the substrates in these 5' nucleases has motivated remarkable research into the structure and function of each of these proteins. This section commences by highlighting the milestones in shaping our current understanding, prior to starting my PhD thesis work, of the individual four 5' nucleases and concludes with a description of their mechanistic unifying features.
1.2.1 Flap Endonuclease 1

FEN1 is the archetype of the 5’ nuclease superfamily. Early work with bacterial and eubacterial systems identified 5’ exonucleases/endonucleases as integral to DNA replication (30, 31). The discovery of mammalian FEN1 and its subsequent characterization (23, 32, 33) has transformed our understanding of the substrate recognition mechanism in 5’ nuclease and their vast roles in other DNA metabolism pathways. FEN1 is a monomeric metallonuclease 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. The N-terminal region contains the nuclease domain that poses 5’ endonuclease and 5’-3’ exonuclease activities and the C-terminal region marks the site for as many as 20 protein-protein interactions with other partner proteins including the eukaryotic sliding clamp (PCNA) and the Werner syndrome ATP-dependent helicase (WRN) (34-36). FEN1 is vital in many cellular DNA transactions especially Okazaki fragment maturation, long-patch base excision repair (BER), stalled replication fork rescue and telomere maintenance (37-40). Hence, genome stability and integrity necessitates that FEN1 catalyzes the incision of its substrate with high specificity, efficiency and regulation.

1.2.1.1 Biochemical properties

FEN1 was thought to recognize single stranded 5’ flaps as its cognate substrate in line with its bacteriophage homolog, T5 5’ exonuclease (T5FEN). However, its current widely accepted cognate substrate is a double flap (DF) consisting of a 5’
ssDNA or ssRNA flap of variable length and a strict 1 nt 3’ flap (41-43) (Figure 1.3). The DF substrate is considered the preferred FEN1 substrate for the following reasons. First, FEN1 has higher affinity to DF substrates as compared to single flap (SF) substrates as evident with lower apparent $K_D$s (42, 43). Second, FEN1 cleaves DF substrates with better efficiency than single-flap substrates. Third, DF substrates are cleaved with higher precision resulting in 100% ligatable products (42, 44). Importantly, the presence of the 3’ flap is believed to enhance the rate of reaction such that FEN1 cleavage on DF substrate approaches diffusion-limited kinetics (44, 45). Whereas in vitro biochemical experiments have used a static DF substrate where the 1 nt 3’ flap is unpaired, the in vivo substrate is believed to have an equilibrated DF structure (Figure 1.3). During Okazaki fragment maturation, the 5’ flap emerges as a result of the strand displacement activity of the lagging strand polymerase, Pol δ. Pol δ strand displaces the RNA/DNA hybrid primer generated by Pol α-primase complex in the initiation of the previous Okazaki fragment. Since both Pol α-primase complex and Pol δ use the same DNA template for their synthesis, the DF substrate equilibrates between a SF substrate with only 5’ flap and a DF substrate with 1 nt 3’ flap that can reanneal back after the cleavage reaction (Figure 1.3) (46).
Figure 1.3. 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.

Early biochemical characterizations of FEN1 investigated the optimal conditions of its nuclease activity. These studies demonstrated that FEN1’s reaction is independent of the substrate sequence and it can process substrates with 5’ flaps up to 200 nts long (31). For mammalian FEN1, optimum cleavage temperature was found to be 37°C. This cleavage reaction was shown to be metal-ion dependent where it is supported by magnesium and manganese but not by other cations such as calcium and zinc. Moreover, the optimal magnesium range of FEN1 nuclease activity was determined as 1-10 mM (23). This metal ion-independent nuclease activity generates 5’-phosphate monoester and 3’-hydroxyl products (47). Earlier work suggested that salt concentrations beyond 50 mM NaCl is inhibitory to the nuclease activity (23). However, later studies showed that optimal activity occurs at 100 mM NaCl (44).
1.2.1.2 Substrate recognition

Tracking

In the search for FEN1 substrate recognition mechanism, a 5’ flap tracking model emerged. This model was initially supported by the observations that the presence of secondary structures or dsDNA in the 5’ flap was inhibitory to FEN1 cleavage efficiency (40). Similar results were observed when the 5’ flap contained a block, such that generated by the strong binding of streptavidin to a biotin group, placed at the tip of the flap or when the flap contained certain chemical adducts closer to the ss/dsDNA junction (48). The replacement of the 5’ flap structure with bubble or branched structures also yielded analogous inhibitory effect (48). Taken together, these results indicated that the presence of a free 5’ flap end is integral to FEN1 catalytic efficiency. Thus, the tracking model was proposed to describe FEN1’s substrate recognition where the protein enters the free 5’ end and tracks along the 5’ flap until it reaches the base of the ss/dsDNA junction and catalyzes the cleavage reaction.

Threading

Nevertheless, work with eubacterial FEN1 suggested that the endonuclease first binds to the duplex region around the base then threads in the 5’ flap to the active site for cleavage (49). A similar threading mechanism for mammalian FEN1 was then proposed. This threading mechanism is supported by recent biochemical and structural data (discussed in Chapter 5) (50, 51). The biochemical data showed that regardless of the inhibition of FEN1 nuclease activity when blocking the end
of the 5’ flap, the nuclease can still bind these substrates (50). Therefore, FEN1 must still be able to access these blocked substrates by initially binding to the base of the flap. In this context, the initial binding to the base was proposed to be further stabilized when the 5’ flap is threaded through the nuclease. Further support for the threading mechanism comes from its interaction with PCNA. This interaction was shown to stimulate FEN1 nuclease activity (52, 53). Hence, FEN1 can be envisioned to be part of a larger functional protein complexes with PCNA or other binding partners docking at the flap base rather than drifting on its own to find the 5’ flap end (40). Moreover, the threading requirement can be a protective measure against unwanted incisions in the template strand.

Yet, detailed mechanistic findings on the substrate recognition mechanism were not possible until the structural work advanced on the mammalian FEN1 and its homologs, including the structures of the enzyme in its apo form or when bound to its substrate or the product. The apo structures of FEN1 from different organisms were solved and they demonstrated a common overall architecture among FEN proteins (54-60). The nuclease domains in these proteins adopt an α/β structure named PIN or SAM fold. In addition, two α-helical regions flank a mixed-twisted β-sheet arranged in a saddle-like shape. The β-sheet is also flanked with two dsDNA binding sites, with a protrusion of two α-helices from the saddle-like structure (28). The structures were in line with phosphodiester bond hydrolysis mechanism by a two metal-ion coordination (4) through the conserved seven acidic residues (61, 62). Surprisingly, these structures revealed that the N-termini of archael and eukaryotic FENs are structured and folded in the proximity of the active site.
The crystal structure of the FEN1 bacteriophage homolog, T5FEN1, in its apo form defined a structural feature of FEN1 that permits only ssDNA to pass through and access the active site (54). This feature was called the helical arch as it is composed of two $\alpha$ helices assembled in an arch structure above the active site. The subsequent crystal structure of FEN1 from Archaeoglobus fulgidus confirmed the biochemical data that the preferred substrate for FEN1 is a DF with an additional 1 nt 3' flap (55). The structure showed FEN1 forming a hydrophobic wedge around the 3' flap, thus positioning the nuclease at the flap base and in line with a precise incision of 1 nt into the downstream duplex DNA resulting in a perfect nick (Figure 1.3). This FEN1 recognition of the 3' flap was further confirmed by the first crystal structure of human FEN1 (59). Additionally, the apo structure of human FEN1 showed a disordered region around the active site with two metal ions coordinated at a distance <4 Å apart.

**Clamping**

The later crystal structure of another human 5' nuclease, human EXO1, suggested a third possible mechanism called clamping that could be unifying mechanism for both nucleases (63). Similar to threading, with the proposed clamping mechanism, FEN1 would bind first at the flap base, but instead of the 5' flap threading through the helical arch, the helical arch clamps around the 5' flap. Authors in this study favored a clamping model to explain difficulty the threading the ssDNA flap or gapped flap without a coupled energy source through a restricted helical arch (63). In support of this mechanism, biochemical studies have shown that human FEN1 can endonucleolytically cleave gapped flap substrates (44). Gapped flap
substrates are flap substrates where the 5’ flap contains a short duplex region towards the tip of its 5’ end. The cleavage of such substrates cannot be explained with the threading mechanism as the structured helical arch cannot accommodate the passage of the 5’ duplex region. However, the clamping model was subsequently rejected due to observation that FEN1 cannot cleave the streptavidin blocked 5’ flap substrate or RPA bound 5’ flap substrate. The authors in this study proposed a variation of threading model called disorder-thread-order mechanism, in which threading occurs in a disordered state of the cap-helical gateway of FEN1, which structures after threading to assemble the active site and cleave the substrate (64, 65).

1.2.1.3 Human FEN1-DNA crystal structures

Concurrently, two crystal structures of human FEN1 bound to substrate and product DNA were solved (51) (Figure 1.4). Upon incising of the DF substrate, FEN1 generates two products: a 5’ ssDNA flap, which is instantaneously released and a duplex dsDNA product that is a competitive inhibitor of FEN1 binding to its DF substrate (44). In the crystal structures (51), the substrate DNA used, called the EXO substrate, harbors a static 1 nt 3’ flap and a completely complementary downstream primer; in other words, if the 3’ flap were allowed to equilibrate, it would create a 1 nt 5’ flap substrate. The product DNA, on the other hand, is the duplex DNA product upon the cleavage of a static DF harboring 4 nt 5’flap and 1 nt 3’ flap. Together, these two structures detailed a remarkable FEN1 substrate recognition and incision mechanism. Most of the interactions in the protein-
substrate and protein-product complexes are similar, but differences are localized to the interactions with the 5' flap.

*Overall structure*

In the substrate and product complexes, the protein adopted a structure resembling a left-handed boxing glove with its palm and fingers interacting with the upstream and downstream duplex regions on both sides of the ss/dsDNA junction (Figure 1.4). These interactions comprise most of the binding surface and occludes more than 16 nt. Strikingly, the two duplex regions are bent at an angle of ~100° with respect to each other with the active site bearing the dual metal ions assembled at the flap base (51). This bending led to the proposition of a novel initial recognition mechanism where the protein binds the junction and bends the flanking duplex DNA regions at a single phosphate diester. Furthermore, interactions with the template strand constitute about half of the total protein-DNA interface in contrast with the anticipated bias towards the 5' flap strand suggested by the three proposed mechanisms of substrate recognition (28). The interaction with the template strands is supported in part by a coordinated potassium ion in a Hairpin-two Turn-Helix (H2TH) motif, a hydroxyl group in the side chain of a serine residue and the template DNA via a phosphate diester. Overall, the protein-dsDNA binding interface is not contiguous and a template arc is formed due to this discontinuity. This template arc delivers the downstream strand containing the 5' flap to the vicinity of the active site (Figure 1.4).
Figure 1.4. Structure of FEN1. (A) Structure of FEN1 showing DNA bending and various structural features. DNA bending is induced by breaking the path of continuous DNA by FEN1 hydrophobic wedge. The structure further highlights FEN1 3' flap binding pocket cradling the unpaired 3' flap and the cap-helical gateway overseeing the active site and controlling the access to catalysis. Downstream DNA is bound by H2TH and K+ binding motif (PDB 3Q8L). (B) (Left) The structure of FEN1 apo-enzyme shows the disordered cap-helical gateway and 3' flap binding pocket (PDB 1ULI) (60). (Right) The structure of DNA bound-FEN1 shows disorder-to-order transitioning upon DNA binding (PDB 3Q8L) (51).
The upstream side

On the upstream side, the template strand contributes to most of the protein-dsDNA interactions with very localized interactions with the last three nucleotides of the upstream strand next to the junction (51). However, the interactions with the 1 nt 3' flap are significant as the protein bears a specialized 3' flap binding pocket that is composed of ten amino acid residues (Figure 1.4). Most of the interactions within this pocket are directed to the sugar-phosphate backbone and the 3' hydroxyl group but not to the base, which is consistent with the structural rather than the sequence specificity of FEN1. The structure of the 3' flap binding pocket with a formed cleft can only accommodate a 1 nt 3' flap. This selectivity is imposed by a FEN1-unique acidic residues forming what is known as an “acid block”, which putatively blocks the DNA from moving further (28). The substrate and the product complexes also recognize another structural feature known as the hydrophobic wedge. This hydrophobic wedge contributes to the stabilization of the bent conformer through its interaction with a penultimate nucleotide. It is worth noting, that this hydrophobic wedge in the apo structure appears disordered (60). However, the hydrophobic wedge is structured when Archaeoglobus fulgidus FEN1 is bound to only the upstream duplex DNA containing a 3' flap (55). These different states of the hydrophobic wedge suggest that FEN1 undergoes ordering upon binding to the 1 nt 3' flap. The only distinguishing feature for this 1 nt 3' flap is its 3' hydroxyl group which is considered instrumental for substrate recognition by FEN1 (Figure 1.4).
The downstream side

On the downstream side, in contrast to the apo structure (59, 60), the helical arch is structured in a gateway-like conformation. This suggests that this region of the protein becomes ordered only upon encountering DNA (51). The posts of this gateway structure are formed by helix 2 and most of helix 4, while the rest of helix 4, along with helix 5, forms a cap structure that crowns the gateway (Figure 1.4). This helical gateway sits on top of the active site and forms a narrow channel that is 13 Å wide at its narrowest point, suggesting that the helical gateway would not be able to order around dsDNA. A cap structure seals the helical gateway and restricts access to the active site to only 5’ flaps with a free end and not to other structures such as bubbles (51). It is worth mentioning that the helical gateway is conserved in all superfamily members but the cap is exclusively present in FEN1 and EXO1 (51, 63). The protein mainly interacts with the 5’ flap at its base and the rest of the flap is expected to pass through the helical gateway. Therefore, in reconciliation with the biochemical data of FEN1 cleaving gapped flap substrates (44, 48), it seems that FEN1 requires sufficient ssDNA at the flap base for the helical gateway to order. This also explains why FEN1 can tolerate some adducts in the 5’ flap as long they are located few nucleotides away from the flap base (48). Furthermore, based on the protein-product complex with the phosphate backbone orientation through the active site, the researchers predicted that the 5’ flap would thread through the helical gateway. However, the ordering of the helical gateway around the ssDNA cannot distinguish between the threading mechanism and the clamping mechanism. The debate over threading versus clamping is now settled.
after the crystal structure of FEN1 with a 5’ flap substrate was solved (discussed in Chapter 5) (66).

Unpairing

Another structural feature that emerged from the FEN1-product structure and defined a mechanistic criterion of substrate recognition was the observation that the nucleotide just to the 3’ side of the cleaved phosphate (termed -1) was unpaired from the template strand (51). However, this same nucleotide was paired in the protein-substrate product. Structural alignment of the protein-substrate and protein-product complexes suggested that the two nucleotides flanking the scissile phosphate (termed +1 and -1) must unpair to prepare for catalysis (Figure 1.5). Supporting this unpairing mechanism is the helical parameters of the DNA in the protein-substrate complex. While the upstream region conforms to B-DNA, the six bases closest to the active site on the downstream region are more like A-DNA with less stacking of the 5’ flap strand nucleotides (28). In the protein-product structure, three residues Y40, K93 and R100 are implicated in stabilizing the unpaired -1 nucleotide by stacking against its base (Y40) or interacting with its terminal phosphate (K93 and R100) (Figure 1.5). This unpairing requirement to bring the scissile phosphate in position for catalysis was later argued against as the crystal structure of FEN1 with a 5’ flap substrate was solved (discussed in Chapter 5) (66).
Figure 1.5. Two nucleotide unpairing. (Left) FEN1 structure showing key residues postulated to be part of unpairing process (K93, R100 and Y40). The R47 residue is shown, which is placed between the cap-helical gateway and the 3’ flap binding pocket and can potentially act to communicate the ordering between these disordered regions. Colored region of FEN1 shows the cap-helical arch region. (Right) Structure showing FEN1 substrate (PDB 3Q8L) and product (PDB 3Q8K) DNA structure showing two terminal nucleotides (+1 and -1) that are paired in the substrate structure but unpaired in the product structure. The authors hypothesized that unpairing brings the scissile phosphate closer to the metal ions, a requirement of metal ions-mediated catalysis (51).

Active site assembly and catalysis

In the protein-product complex, the 5’ phosphate monoester interacts with the two divalent metal ions in the active site, whereas in the protein-substrate complex, the scissile phosphate is in the proximity of the active site within 5-8 Å but with no direct interactions with the metal ions or the residues of the active site (51). It follows then that FEN1 binding and bending of the duplex DNA in a non-contiguous manner is sufficient to bring the scissile phosphate in the vicinity of the active site.
Overall, the positioning of the scissile phosphate diester in the vicinity of the active site seems to mainly involve the template strand and the upstream primer (3’ flap strand) rather than the downstream primer (5’ flap strand). This could be viewed as a protective measure against imprecise cleavage within the ssDNA of the 5’ flap (28). The protein scarcely interacts with the 5’ flap strand which allows for the advantageous indiscrimination between RNA and DNA-containing 5’ flap strands. However, the helical properties of the bent two-way DNA junctions and the spacing between the binding sites allow FEN1 to distinguish between a 5’- and a 3’- flap, which would inhibit the incision of the latter one. Thus, these helical properties of the bent DNA present an elegant platform by which FEN1 can position the scissile phosphate in the vicinity of the active site while selecting against 3’ flaps.

Additionally, the FEN1-product structure showed the assembly of the active site and shed some light on its metal ion-dependent catalytic mechanism. The assembled active site allows for endonucleolytic hydrolysis of the phosphodiester bond through electrostatic mechanism. In the active site, the two metal ions are directly interacting with the terminal phosphate of the product (51). They are also directly coordinated by four of the seven conserved carboxylate residues in the exact sites that were reported for the apo structure (60). These metal ions are coordinated by D86, E160, D179, and D181 on the inner sphere, and are bridged by E160. On the outer sphere, the remaining three conserved carboxylates, along with phenolic hydroxide of a conserved tyrosine (Y234), constitute the rest of the contacts via water molecules. The researchers proposed that the two metal ions activates a water molecule that attacks the scissile phosphate diester bond and
hydrolyzes it in a mechanism analogous to that proposed for EndoIV (67, 68). This proposition is distinguishable from the proposed two metal ion mechanism of cleavage where one of the metal ions activate the nucleophilic water (63). There is not enough evidence to support either proposition, however the requirement of two metal ions for the hydrolysis of phosphate diester bond is well supported (69).

1.2.1.4 Importance of FEN1

During replication, ~50 million Okazaki fragments are generated per each cell cycle and FEN1 is at the center of processing these fragments. In addition to its instrumental role in Okazaki fragment maturation, FEN1 has been implicated in many cellular processes involving DNA metabolism. In a similar fashion to Pol δ’s strand displacement during Okazaki fragment maturation, Pol β can strand displace at a apurinic/apyrimidinic (AP) site resulting from a damaged base and initiate long-patch BER (40). FEN1 is the nuclease recruited to process the resulting flap. With the cellular DNA being under constant attack from endogenous and exogenous factors, the efficiency of DNA repair pathways is indispensable. Furthermore, FEN1 has been suggested to suppress the trinucleotide repeats expansion (70). Trinucleotide repeats expansion has been associated with several neuromuscular and neurodegenerative diseases, including Huntington's disease and myotonic dystrophy (71). FEN1 is also believed to play a role in telomere maintenance. Moreover, despite its predominant localization to the nucleus, FEN1 was confirmed to localize in the mitochondria (72, 73). Perhaps, it is involved in mitochondrial BER although this involvement has not been well established.
The importance of mammalian FEN1 is further emphasized with the homozygous deletion of the *fen1* gene in mice causing embryonic lethality (74). On the other hand, due to its basic role in DNA replication, it follows that its expression level has been correlated with the proliferative state in human tissues including cancerous cells (26, 75). In other words, in many cancer types, FEN1 is overexpressed, while the level elevated expression has been witnessed and this elevated expression also correlates with the tumor aggressiveness (76). Moreover, a predisposition to cancer in humans and mice is associated with FEN1 mutations (25, 27, 77, 78). It is not surprising therefore that FEN1 has been targeted for anti-cancer drugs (79, 80). For these reasons, tight regulation of FEN1 is a necessity. This regulation is in part controlled by its interactions with protein partners. In addition, its posttranslational modifications could offer another layer of regulation.

### 1.2.2 Exonuclease 1

Human Exonuclease 1 (EXO1) is another member of Rad2/XPG family of structure-specific 5' nucleases that is involved in multiple aspects of genome replication, repair and recombination. Like other members of this family, the catalytic repertoire of EXO1 is broad with dominant trait of processive exonucleolytic cleavage of nicks, gaps, recessed primer/template junction. EXO1 also possesses endonucleolytic activity to cleave 5' flaps (81, 82). Its exonucleolytic activity is important in diverse nucleic acid processing mechanisms including double-stranded DNA break repair, DNA mismatch repair (MMR) as well
as the DNA damage response. The endonucleolytic activity of EXO1 has been proposed to be involved in immunoglobulin class-switching recombination processing, trinucleotide repeat and possibly in Okazaki fragment maturation. An EXO1 knockout in mice shows severe defects including a decrease in the survival rate and an increased vulnerability to cancer (82, 83).

The most important function of EXO1 is its role in MMR. EXO1 forms a complex with either Mutsα or Mutsβ depending on the nature of the mismatch that needs to be repaired. Both Mutsα or Mutsβ are heterodimeric sliding complexes made of MSH2/MSH6 or MSH2/MSH3 subunits, respectively. This interaction of EXO1 with MSH2 increases the processivity of the exonuclease activity of EXO1 in an ATP dependent manner. Further interaction with RPA confers EXO1 with highly processive exonuclease activity (81, 83).

The EXO1 catalytic domain is located in the N-terminal region of the protein (1-352). The rest of the protein consists of a long carboxy-terminal (353-846) which is predicted to be highly disordered. This region plays an important role in regulating the nuclease activity of EXO1 by binding to a multitude of partner proteins and hosting a large number of post-translational modification sites as well (63, 83).

The catalytic domain of EXO1 has been determined in the DNA bound form. FEN1 and EXO1 catalytic regions share very low sequence homology but share structural similarities of different domains. However, due to the different primary role of FEN1 and EXO1 as endonuclease and exonuclease, they possess
additional features for these functional fortes. Geometrically, the active site is located similarly to FEN1, where it lies buried with restricted access. The active site coordinates two metal ions using 5 conserved aspartate residues (51, 63).

The DNA is contorted by EXO1 using its hydrophobic wedge, similar to FEN1, by inducing a severe bending of 90° at the nick junction. This bent DNA is stabilized by the surface architecture of the protein using different binding domains that interact with the template strand. These DNA-protein interactions include the contacts between the H2TH and the downstream side of the nick junction as seen in the case of FEN1. The H2TH harbors the K⁺ binding site that coordinates the interactions with the phosphate backbone. The upstream side of the nick junction is bound by hairpin-helix-hairpin domain (63).

The access to the active site is controlled via a highly positive α4-α5 helical arch that controls the access to the active site as is the case in FEN1 (63). In FEN1, this region is disordered in the absence of DNA (60). It is reasonable to assume that in EXO1, the situation will be similar. However due to the absence of an apo form structure of EXO1, the best evidence that supports this proposition is the large B factor of these residues which suggests some degree of disorder (63). Further structural information was provided by time-resolved crystal structures of exonucleolytic cleavage before and after the hydrolysis step of cleavage (84). Two metals were within 3.6 Å distance, sufficient to induce phosphodiester cleavage of the bond. Also, a water molecule is situated to perform an in line attack on the phosphodiester bond. The arch domain was proposed to undergo a series of transitions from open/close equilibria to the clamped state. In the open state, the
DNA is mobile and possesses a higher degree of freedom. The transition occurs when the complex equilibrates to the closed state, in which both the junction and the DNA are held with more restriction and order. Finally, the complex moves into the clamped state, in which the active site is ordered, the metal ions are engaged, and the hydroxyl group takes the in-line position to attack the phosphodiester bond. Furthermore, it was shown that at least two residues (H36 and Y32) act to courier the DNA along the different conformations and guide it to its proper orientation such that the scissile bond is near the metal ions and the hydroxyl ion is in line to attack. Particularly, H36 flips its conformation from up to down when the transition between closed to clamped state occurs (84).

1.2.3 XPG

XPG is the last member to the 5’ nucleases superfamily. XPG is primarily involved in nucleotide excision repair (NER) (85). NER is characterized by its ability to detect and remove DNA damages based on their alteration of the Watson-Crick structure of the double helix. Therefore, unlike other pathways for DNA repair (BER, MMR) that act against a particular type or set of lesions or damages, NER is versatile in its ability to find many kinds of damages and repair them. The repertoire of lesions that can be detected by NER include cyclobutane–pyrimidine dimers (CPDs), 6–4 pyrimidine–pyrimidone photoproducts (6-4PPs), DNA crosslinks, and multitude of other bulky modifications. Any defect in any factors of
NER pathways can cause serious diseases including: xeroderma pigmentosum (XP), Cockayne syndrome (CS) and trichothiodystrophy (TTD) (1, 86, 87).

NER operates via either of the two sub-pathways: Global genome-NER (GG-NER) and Transcription coupled-NER (TC-NER) which differ in the mechanism of damage recognition. GG-NER is active throughout cell cycle whereas TC-NER is specialized to sense the lesions in actively transcribed regions in coupled manner with stalling of the RNA polymerases (87). Overall, NER uses at least 30 different component proteins to complete the execution of repair from the recognition of the damage to the ligation step. NER starts via recognition of bulky DNA lesions by XPC-HR23B. XPA is the next protein recruited, although the exact role of XPA in repair is still under investigation. Transcription factor TFIIH is the next component to be recruited to the damage site by direct interaction with XPC-HR23B. TFIIH, with its two helicases XPB and XPD, opens the DNA around the lesion. This DNA opening step facilitates the formation of the pre-incision bubble complex composed of XPG, XPF-ERCC1, XPA and RPA, which has been shown to be around 25 bp. Sequential cleavage of this bubble DNA by XPF-ERCC1 and XPG leads to expulsion of the damaged DNA 25 bp ssDNA. Further synthesis by either of the replication associated polymerases (Pol δ or Pol ε) or the translesion polymerase (Pol κ) along with PCNA, RFC and DNA ligase 1 completes the synthesis and the ligation steps and thus repairs the damage fully (1, 85, 86).

XPG is recruited to the pre-incision complex by its interaction with the transcription factor TFIIH (88). It has been shown to be constitutively associated
with TFIIH and is believed to play a role in transcription that is different than TC-NER (86). XPG also has a structural role in NER in the pre-incision complex, apart from its catalytic role of the 3’ incision of the bubble. The 5’ incision of the bubble by XPF-ERCC1 happens only upon binding of XPG. It is hypothesized that XPG is required to stabilize the pre-incision complex and to mediate the protein ordering in XPF-ERCC1 that is required to assemble its active site (89).

XPG, like other Rad2/XPG family members, contains an N-terminal catalytic domain and an I- domain, which together form the catalytic core of this nuclease. The N- and I-domains are conserved in all 5’ nucleases and usually are connected by a short linker domain. However, in XPG, the two domains are interspersed by a large (600 amino acids) linker domain called the spacer region. This unique spacer region is not conserved and is predicted to be disordered. However, this spacer region is expected to mediate its interaction with other NER proteins including TFIIH and RPA. The last region of XPG is its C-terminal, which mediates its interactions with other proteins such as PCNA (86, 90).

The crystal structure of the catalytic domain of Rad2 (S. cerevisiae XPG homolog) without the disordered spacer domain was determined in the presence of recessed 5’ end dsDNA (91). The structure shows that most of the interactions of Rad2 are with the dsDNA part of the substrate. Moreover, most of the structural features defining other 5’ nucleases, such as the H2TH DNA binding motif, the hydrophobic wedge and the helical arch, are present in Rad2. The Rad2 active site is made of D30, D77 (from N-region), E792, E794, E813, E815 and E864 (from I-domain). An important structural element present in Rad2 is the presence of the
G37 amino acid which forms a H-bond with the base of the cleaved strand. Other important difference in Rad2 as compared to FEN1 and EXO1 structures is the nature of the helical arch. Whereas α4 and α5 form the helical arch in FEN1 and EXO1, which shields the active site to prevent the inadvertent cleavage in their template strands, Rad2 lacks α5 and the position of its α4 is away from the active site. This makes the active site more accessible and can create a channel for an intact ssDNA bubble to move into the active site. This feature also makes XPG dangerous, in the way that it can cleave the template strand of the DNA and hence needs significant control and regulation (91). However, the lack of structural information on XPG and its activity as part of the NER complex makes it difficult to derive models about its substrate recognition mechanism.

1.2.4 GEN1

GEN1 is the prototypical Holliday junction (HJ) resolvase and is a bacterial equivalent of the RuvC resolvase (92, 93). GEN1 is a member of the 5’ XPG family of structure-specific nucleases, unlike RuvC which is related to the retroviral integrase family (5). HJs are four way junctions that join two duplex DNA molecules (94). These junctions arise during double strand breaks by homologous recombination, which is carried out by the interaction of two sister chromatids. Mechanistically both GEN1 and RuvC offer similar approach to resolving these HJ intermediates by coordinated action of introducing two nicks by two monomers of the enzyme. This action of dual incision results in the uncoupling of two DNA
molecules, which upon ligation revert back to normal heritable dsDNA structures (95).

In eukaryotic cells, there are two main pathways for processing HJs. Difference in these pathways primarily comes from whether or not they use nuclease-mediated hydrolytic cleavage (96). The non-nuclease mediated pathway proceeds by the helicase action of bloom helicase in BTR complex (BLM-TOPIIIα-RMI1-RMI2), by bringing two Holliday junctions close to each other and further uncoupling them by Topoisomerase3α (96, 97). This pathway is the most common approach to resolve HJs during homologous recombination and works by restoring the original genetic information by forming non-crossover products. This pathway is active throughout cell cycle and is responsible for the majority of HJ processing (96, 97). The HJs that escape this pathway are acted upon at a later stage of the cell cycle by two distinct pathways, both of which rely on the action of nucleases. The first of these nucleases is MUS81-EME1, which is activated in the prometaphase by the formation of a larger complex containing two other nucleases, the heterodimers SLX1-SLX4 and XPF-ERCC1 (96, 98). These three heterodimeric complexes together form a potent nuclease machine, called SMX trinuclease, that can resolve a large number of 5’- and 3’-intermediates left over by other DNA processing pathways including double HJs (98). The other pathway involves a sequential nuclease activity of the homodimeric 5’ endonuclease GEN1 (96, 97). GEN1 is a cytosolic protein and is only active after the breakdown of the nuclear envelope (97). Therefore, it is proposed to act as the last resort for resolving persistent HJs that escaped the BTR and the SMX pathways (99).
GEN1 is a member of the Rad2/XPG family of 5’ nucleases (29). Despite multiple past efforts to find the bacterial equivalent of RuvC, Human GEN1 and its Yeast equivalent YEN1 were discovered recently (92, 93, 100). GEN1 resolves HJs by introducing symmetrical cuts, thus making the resolved products immediately ligatable (94, 95). Structure of the human and fungal GEN1 catalytic cores have been determined. GEN1 contains classical 5’ nuclease features including N-terminal nuclease domain, an I-domain, followed by a conserved nucleic acid binding domains including H2TH motif and K+ (101, 102). Furthermore, GEN1 has an additional DNA binding domain in the form of a chromodomain (chromatin organization modifier domain) (102). The carboxy terminal of GEN1 is highly divergent in sequence and is predicted to be structurally disordered and might involve post translational modification and partner protein interaction sites (102).

GEN1/YEN1 can cleave a wide variety of DNA structures, including replication forks, splayed arms, 5’ flaps as well as its primary HJ substrate (94). Thus, the need to control its activity seems to be of paramount importance as its nuclease activity can cleave dsDNA replication/recombination intermediates. Control of the catalytic activity of GEN1/YEN1 is yet to be deciphered in full detail. Initially, it was proposed that the C-terminal of GEN1 may act as auto-inhibition domain of GEN1, reminiscent similar mechanism proposed for EXO1 nuclease (63, 92). Support for this seemed to stem from the fact that active nuclease isolated from fractionated HeLa extracts was C-terminal truncated. However, the recent characterization of full length GEN1 showed that it has equivalent robust activity.
compared to that of the truncated version (95). Control of YEN1 activity has been shown to be dependent on the phosphorylation status of the conserved CDK sites present in its carboxy-terminal region (96, 99, 103). Furthermore, even though human GEN1 also contains these CDK phosphorylation sites, effect of the phosphorylation status seems to have little effect on the catalytic activity of GEN1 (96). However, the presence of a nuclear export sequence (NES) present in the carboxy-terminal region in GEN1 seems to be an effective way to temporally restrict its access to genomic DNA after the breakdown of the nuclear membrane at mitosis (97, 99). Consequently, it has been proposed that majority of the recombination intermediates are processed by the SMX trinuclease complex and only persistent covalent bridges that escape the BTR/SMX pathways can be resolved to enable the chromosomal segregation. In fact, the artificial presence of nuclear localization signal sequence (NLS) and the deletion of NES in GEN1 lead to increased instances of sister chromatid exchange (99). These hypotheses leave open the role of the extended C-terminal part of GEN1, which seems to be highly disordered similar to EXO1 and FEN1, thus can be also effective to control the catalytic activity of GEN1 either via protein-protein interaction or due to other post translational modifications as is the case in FEN1 and EXO1. YEN1 is also controlled dually either via subcellular compartmentalization and/or nuclease activation (96).
1.3 Unifying features of 5’ nucleases

1.3.1 DNA bending

Selection of the cognate DNA substrate and licensing its cleavage is understandably the most important feature of 5’ nucleases. Any lax in control of their potent nuclease specificity can create havoc for genome integrity. DNA bending seems to be one of the most ubiquitous features of the 5’ nucleases studied thus far. Both FEN1 and EXO1 structures show the DNA bent at a 90° angle (51, 63). The structure of GEN1 in the product complex also showed the DNA bent to a similar degree as in FEN1 and EXO1, suggesting that GEN1 bends the HJ in a non-planar orientation (101, 104). There is no definitive information if XPG also changes the conformation of the bubble-containing DNA. This superfamily-unifying DNA bending intermediary step is invariably caused by the interaction of the superfamily conserved hydrophobic wedge at the ss/dsDNA junction. The bent DNA is stabilized by interactions at the down- and upstream of the junction that are located at a distance of one DNA helical turn apart (29).

DNA bending is not specific to 5’ nucleases but is a more universal feature to interrogate DNA contour. Early on, it was recognized that DNA bending plays an important role in diverse DNA transactional processes such as transcription, specific recombination and DNA packaging including the cases of integration host factor (IHF) mediated DNA bending and Histone like protein (HU) mediated DNA wrapping (105). DNA bending can be either intrinsic to the DNA itself, as a consequence of conformational variability due to changes in the normal Watson-
Crick base pairing, or it can be a result of an active bending process as the DNA interacts and aligns along the protein surface (106).

What role does DNA bending serve in these nucleases? dsDNA is a very stiff molecule. A worm-like chain model of B-DNA predicted that the amount of free energy required to bend 10bp of dsDNA by 50° is around 70 kJ/mol (107). However, the cost of bending a discontinuous DNA would be significantly less. This differential energy cost of DNA bending can serve as an important mechanism to test for the DNA discontinuity itself. Like other biochemical recognition processes, the origin of DNA bending has been subject to debate. One hypothesis proposes that any non-Watson-Crick base pairing would present itself in a bent conformer since it is already in a deformed structure. Thus, the protein would search for this bent DNA conformer and engage it in the appropriate pathway that would extrude this damage. This model can be thought of as a classical DNA conformational selection model. In another hypothesis, the protein actively sculpts the damaged DNA (108-111). This model is called DNA induced-fit model. In the case of 5’ nucleases, the issue of DNA bending being a result of conformational selection or induced-fit is still under debate. The implications of these two-contrary mechanisms bear fundamental relation to the substrate recognition mechanism in structure-specific nucleases owing to their indifference to the sequence of the DNA. Therefore, deciphering the basis of DNA bending in these nucleases would expectedly provide much needed understanding of the basis of substrate recognition and may help in the design of inhibitors for these proteins, which are targets for cancer therapy.
1.3.2 Disorder-to-order transition

All 5’ nucleases are known to undergo some degree of protein ordering upon substrate binding (29, 51, 63, 84, 101, 102). This is most clearly demonstrated in the case of FEN1 when comparing its structure in the apo- and substrate-bound form (51, 60). The helical arch region comprised of α4–α5 is disordered in the free protein. The ssDNA 5’ flap threads through this helical arch and is ordered in the process leading to the assembly of the active site. Furthermore, FEN1 in the apo enzyme form possesses one more disordered region in the α2–α3 loop that binds to the unpaired 3’ flap. This region becomes ordered in the presence of DNA substrate (51, 55, 60). In the case of EXO1, the large B-factor of the residues in the EXO1-DNA structure in the helical arch region also point to it being disordered in absence of DNA (63). Further support for this argument comes from recent time resolved crystallographic studies which showed that both the helical arch region and the DNA in the protein-DNA complex display a gradual increase in order as the complex moves from “binding” to the “transition” state (84). In both FEN1 and EXO1, the active site assembly is a consequence of this transition to lower degree of freedom in the protein helical arch region. The GEN1 structure also suggests that it undergoes a protein ordering step since part of the α4 region is disordered in the absence of DNA (102). In the case of XPG, the entire 600 amino acid spacer domains (R-domain) is predicted to be disordered, although not much is known about its functional mechanism (91). Collectively, the structural work on 5’
nucleases suggests that disorder-to-order transitioning is a common mechanism for controlling their active site assembly.

FEN1 cleaves gapped flap substrates, in which the 5’ flap contains short dsDNA that is separated from the nick junction by a short ssDNA (44). Similarly, it cleaves a DF substrate containing a bulky adduct in the 5’ flap so long as there is ssDNA between the adduct and the nick (48). Biochemical data also demonstrated that 5’ flap threading is a pre-requisite for catalysis (64). Given that a structured helical arch can’t accommodate a bulky adduct and dsDNA, these findings collectively suggest a model in which FEN1 initially threads the 5’ flap through the unstructured helical arch that subsequently transition to an ordered arch for active site assembly. Similarly, for EXO1, short flap ssDNA can thread through the helical arch and might also involve a disorder-to-order mechanism (82). What drives this ordering of the protein and its relevance is a question of intense research focus. One important function of this disorder-to-order mechanism is that it can potentially help in threading the ssDNA/gapped flap DNA through a restricted cap-helical gateway, which might be infeasible without an energy source if it were already structured. This might help FEN1 cleaves 5’ flap containing mismatches and secondary structure elements. Another possibility is that its ordering upon cognate substrate binding serves to rapidly select for the correct substrate in an induced-fit mechanism, thereby averting the catalysis of incorrect substrates. Also, a disorder-to-order mechanism can serve to decrease the entropic costs of ordering, until the enthalpic gains are met by cognate substrate binding. Disorder-to-order mechanism among different parts of the enzyme, triggered by different aspects of
substrate binding can also provide multi-dimensional validation before the enzyme commits to catalysis. Lastly, disorder-to-order can also serve as a mechanism of product release, thereby increasing the catalytic efficiency of the enzyme and acting as a signal for the next enzyme in the pathway to take over.

FEN1 was shown to cradle the unpaired 3’ flap nucleotide in a pocket, which itself undergoes a disorder-to-order transition. This raised the question of what drives the disorder-to-order transition in these two regions of FEN1. A possible clue to this question is that FEN1 recognizes and cleaves flap substrates bearing 1nt 3’ flap and no 5’ flap efficiently (44). The other clue comes from earlier biochemical work that shows significant decrease in FEN1 cleavage efficiency when removing the 3’ flap (44). The crystal structure of *Archaeoglobus fulgidus* FEN1 with only a 1-nt 3’ flap overhang showed a structured helical arch (55). Taken together, these results might suggest that binding of the 3’ flap is what drives the ordering on the helical arch. Whether there is coupling and synergistic control of both domains that undergo ordering is yet to be determined. Similar control of the catalytic activity by a single unpaired nucleotide has been shown in the case of HLTF (Helicase-like transcription factor) protein involved in protecting, reversing, and remodeling stalled replication forks (112). However, the lack of any such unpaired nucleotide in EXO1 and other 5’ nucleases, suggests some other mechanism that would regulate the ordering/conformational changes of the helical arch (63). GEN1 has been hypothesized to undergo conformational changes upon dimerization on the HJ, where the active sites of the two monomers face each other to coordinate their incisions in a sequential manner (102, 104). Coupling
ordering/conformational changes in the helical arch with DNA binding would provide a mechanism that controls the assembly of the active site. Logically, this coupling will provide an effective mechanism to verify if the nuclease bound the correct substrate or not before licensing its cleavage.

### 1.3.3 Selectivity of primary substrates and substrate promiscuity

All the 5' nuclease display a catalytic repertoire of cleaving a number of substrates at the 5' junction with either significant structural resemblance or no resemblance to the cognate substrate. They possess both endonuclease and exonuclease activity but have evolved to cleave their cognate substrate with greater efficiency than non-cognate substrates. It is unclear at what stage of substrate recognition this discrimination works. In the case of FEN1, it cleaves a variety of substrates including static DF, equilibrating DF, SF, Y-substrates as well as exonuclease substrates. But among these substrates, FEN1 has highest catalytic efficiency on static and equilibrating DF substrates (28). Similarly, EXO1 cleaves Exo substrates better than flap substrates (82). Both the origin and evolutionary significance of the promiscuous activity of these nucleases needs to be elucidated.

### 1.3.4 Protein-protein interactions and post translational modifications

All the 5' nucleases work in concerted molecular pathways. The need to find their substrates and catalyze their incisions within the larger framework of the respective
pathways make the channeling of a particular protein in space and time an ardent task. FEN1 is recruited to the replication fork by its interaction with PCNA, a trimeric sliding clamp, through its conserved PIP box motif (40). Additionally, it has been shown that FEN1 interacts with more than 20 proteins that alter its activity (Figure 1.6). Similarly, EXO1 interaction with RPA and Mutα/β activates its potent nuclease role in human mismatch repair (81). XPG also interacts with multiple proteins in the NER complex, and is known to even modulate the nuclease activity of XPF (86). Currently, no interacting partner proteins have been identified for GEN1, but it is predicted to have an extended disordered carboxy terminal region, which is usually the hub of interacting proteins. The obvious question that arises here is how these proteins modulate the nuclease activity. For example, PCNA has been shown to increase the binding efficiency of FEN1 to flap substrates, which can explain FEN1 catalytic stimulation (60). But other proteins which do not have any role as such for the recruitment of FEN1 can also increase its activity (28). However, whether any of these proteins work at a particular step during the catalytic cycle, particularly via influencing the disorder-to-order transition has not been shown. Finally, a significant body of study has shown how post translational modification of all these nucleases alters both activity and nuclear access but the precise mechanism by which these modifications influence the activities of these nucleases remains unresolved. (113).
1.4 Use of fluorescence to study nucleases

1.4.1 Principle of fluorescence

Fluorescence is an electronic phenomenon which occurs when an electron in a molecule absorbs the energy in form of photon and emits it back in form of photon with different wavelength. Fluorescence usually occurs in few nanoseconds time scale. The Jablonski diagram used to illustrate this fluorescence phenomena is shown in Figure 1.7. The photonic excitation results in shift of electronic state from
ground ($S_0$) to higher vibrational state of the excited state ($S_1$). This is accompanied by fast internal conversion or vibrational relaxation, which leads to an electron going to the lowest vibrational state in the $S_1$ electronic state. The reversion back to the ground state is accompanied by release of photon of lower energy. The energy loss via internal conversion and vibrational is best manifested in stokes shift in the emission profile to a longer wavelength relative the excitation profile (114).

The decay of the excited state does not always lead to fluorescence. The fluorophore can also return to the ground state via intersystem crossing. This results in fluorophore going into triplet state, which is observed as “blinking” of the dye in single molecule measurements (Figure 1.7). The basis of this blinking phenomenon is a change in the electron spin. The blinking state usually lasts on the order of milliseconds and during this time, the dye becomes desensitized for either excitation or emission (114).

The photo-physical properties of fluorescent dyes are quantitatively characterized by various parameters. The most common parameter to describe the fluorescent dye is its quantum yield ($\Phi_F$), which signifies the probability or efficiency of converting the absorbed photon to fluorescence emission. The other important parameter that describes the fluorescence of the dye is its fluorescence lifetime. Fluorescence lifetime measures the amount of time the dye spends in the excited state before it reverts back to the ground state. Both the absorption and emission of fluorophores are studied at the ensemble level and the properties of the population are deduced from the macroscopic properties of the process (114).
Figure 1.7. Energy landscape of a cyanine dye. (A) schematic representation of potential energy landscape for $S_0$ (ground singlet) and $S_1^*$ (first excited singlet) states as a function of single dihedral angle of the Cy3 polymethine bond, superposed on a Jablonski diagram exhibiting all theoretical transitions and effects. Some of the illustrated transitions are presented only for showing a complete Jablonski diagram, even though for Cy3 they were not observed experimentally, while other transitions were not experimentally tested at all. (B) Schematic representation of the Cy3 dye molecule and one of its dihedral angles, in fluorescent trans and dark cis configurations.
The behavior of the excited state population of fluorophores can be described by a simple rate equation as:

\[
\frac{dn^*(t)}{dt} = -kn^*(t) + f(t)
\]

Where \( n^* \) is the number of excited state population of fluorophore at time \( t \). \( k \) is the rate constant of all de-excitation processes with units as \( \text{s}^{-1} \) and \( f(t) \) is an arbitrary function of excitation process as function of time.

Assuming excitation states at \( t=0 \), the above equation reduces to:

\[
\frac{dn^*(t)}{dt} = -kn^*(t)
\]

This equation describes the decrease in the population of excited state molecules as a function of time. Integrating the equation leads to:

\[
n^*(t) = n^*(0)e^{-kt}
\]

The lifetime of the dye is equal to \( k^{-1} \). At \( t=1/k \), the above equation reduces to:

\[
n^*(t) = \frac{n^*(0)}{e}
\]

which is the definition of fluorescence lifetime. In other words, the lifetime is the time it takes for a population of excited molecules to decay by factor of \( 1/e \) or 36.8\% of the original excited state population. The de-excitation rate constant \( k \) is
summation of many rate constants that compete with fluorescence to de-excite the molecule from the excited singlet state.

\[ k = k_r + k_{ic} + k_{isc} + k_{ET} + k_{iso} + \ldots = k_r + k_{nr} \]

where \( k_r \) is the de-excitation rate via fluorescence, \( k_{ic} \) is the rate of internal conversion, \( k_{isc} \) is the rate of internal system conversion, \( k_{ET} \) is the rate of energy transfer as would be in a case such as FRET and \( k_{iso} \) is the rate of photoisomerization, which is applicable in the case of dyes with non-rigid heterocyclic construction (such as cyanine dyes) that are popular in smFRET experiments.

All these rates except for \( k_r \) do not lead to fluorescence and thus can be summed together in one term as \( k_{nr} \) which is the summation of all non-radiative rates.

The fluorescence lifetime (\( \tau \)) is the inverse of this de-excitation rate constant \( k \), as represented below:

\[ \tau = k^{-1} = (k_r + k_{nr})^{-1} \]

Thus, the fluorescence lifetime (\( \tau \)) would depend on the environmental state of the fluorophore due to the effect of the \( k_{nr} \) parameter in the equation.

The radiative lifetime, \( \tau_r = 1/ k_r \), is the natural or real lifetime of photonic emission, in the absence of any non-radiative deactivation processes. However due to the fact that it is impossible to untangle the radiative rates from other de-excitation rates in experimental measurements, the time characteristic of all the contributing
rates of the de-excitation processes is used instead. This is called fluorescence lifetime.

The quantum yield can be understood in terms of excited state decay rate as:

\[ \phi_F = \frac{k_r}{k + k_{nr}} = \frac{\tau}{\tau_r} \]

### 1.4.2 FRET theory

FRET was first identified by Cario, Franck and Perrin. The quantitative framework to describe FRET as energy transfer between two point dipoles was done by Förster and Oppenhiemer (115). FRET describes a non-radiative energy transfer from an excited electronic state of a donor fluorophore to the acceptor fluorophore (116). This energy transfer occurs via a distance-dependent dipole-dipole interaction of the electronic states of these fluorophores (117). This rate of energy transfer \( k_{ET} \) from the donor to the acceptor is given by:

\[ k_{ET} = \frac{1}{\tau_d} \left( \frac{R_0}{r} \right)^6 \]

where \( \tau \) is the fluorescence lifetime of the donor dye in the absence of an acceptor, \( R_0 \) (Förster distance) is the distance at which the energy transfer occurs with 50% efficiency and \( r \) is the distance between the donor and acceptor dyes (Figure 1.8). The rate is thus inversely proportional to the distance intervening the donor and acceptor dyes \( (r^6) \). Due to its distance-dependent nature, FRET could be used to report on the intervening distances between the donor and the acceptor as a proxy
for the change in the conformation of the biomolecule itself to which the dyes are attached. In that sense, FRET has been referred to as a “molecular ruler”. Förster distance of 1-10 nanometers is appropriate to study biological questions at molecular detail since it is in the size range of biomolecules or the distance between the subunits or binding partners (Figure 1.8). To use FRET as proximity indicator few criteria have to be satisfied including proximity of the donor and acceptor fluorophores and significant spectral overlap between the donor emission spectrum and the acceptor excitation spectrum. It is particularly important that the energy transfer between the dyes does not take place through the exchange of photons, i.e. it is non-radiative but rather happens via the exchange of energy between two oscillating dipoles that resonate with similar frequencies (114). Hence, one of the factors that affect the energy transfer in FRET is the orientation factor of the transition dipoles of the fluorophores, which is described by the \( \kappa \) (Kappa) factor. The rate of energy transfer in FRET can be described by the following relation:

\[
k_{ET} = \frac{Q_d \kappa^2}{\tau_d} \left( \frac{9000 \ln(10)}{128\pi^5 N n^4} \right) \int_0^\infty F_d(\lambda) \varepsilon_A(\lambda) \lambda^4 d\lambda
\]

where \( Q_d \) is the quantum yield of the donor fluorophore, \( \kappa^2 \) is the factor that describes the orientation transition dipoles of fluorophore, \( N \) is the Avogadro's number, \( n \) is the refractive index of the solution in which the energy transfer is taking place, \( F_d \) is the normalized donor emission spectrum between the
wavelength $\lambda$ and $\lambda + \Delta \lambda$ and $\varepsilon_A(\lambda)$ is the extinction coefficient of the acceptor dye (114). The integral in the above equation ($J(\lambda)$) represents the spectral overlap between the emission and excitation spectra of the donor and acceptor dyes, respectively, as follows:

$$J(\lambda) = \int_{0}^{\infty} F_d(\lambda) \varepsilon_A(\lambda) \lambda^4 d\lambda$$

Moreover, since the energy transfer is the ratio of the photons transferred from the donor to the acceptor to the total number of photons that are observed, FRET can be described as:

$$E = \frac{k_{ET}}{\tau_d^{-1} + K_{ET}}$$

This is the ratio of transfer rate of energy to the total rate of decay of the donor in the presence of FRET acceptor. Using the notation $k_{ET} = \frac{1}{\tau_d} \left(\frac{R_0}{r}\right)^6$ in the above equation, $E$ can be re-written as:

$$E = \frac{R_0^6}{R_0^6 + r^6}$$

The above equation shows that the energy transfer is dependent on the distance between the FRET pairs. The most common way to measure FRET efficiency is done by measuring the donor intensity in the absence and presence of an acceptor as follows:

$$E = 1 - \frac{F_{DA}}{F_D}$$
Similarly, FRET can also be calculated from either quantum yield or fluorescence lifetime data

\[
E = 1 - \frac{\tau_{DA}}{\tau_D} = 1 - \frac{\phi_{DA}}{\phi_D}
\]

**Figure 1.8. FRET as a molecular ruler.** Schematic showing the distance dependence of FRET efficiency. Red dashed lines represent the intrinsic distance (R\(_0\)) at which a particular FRET pair results in 50% transfer efficiency.
One of the valuable advantages of FRET is the fact that it is a ratiometric method; that is to say, it measures the internal distance between the fluorophores in the molecular frame not the laboratory frame. Thus, FRET is resistant to instrumental noise and drift (21). On the other hand, FRET has its particular requirements. For instance, not every pair of fluorophores is fit for a FRET experiment. The decision on the FRET pair should take into consideration the following requirements: a) the donor’s emission spectrum should overlap with the acceptor’s absorption spectrum, yet their emission spectra should be resolvable, b) the donor and the acceptor should be within a distance of 1-10 nm for efficient energy transfer to occur, and c) their dipole moments should not be perpendicular to each other or else they would cancel each other out (116). Single molecule FRET (smFRET) is possibly one the most common and flexible fluorescence technique for studying biological questions (118). However, similar to other single molecule fluorescence techniques, it is challenged by the need to detect single molecules. In order to detect single molecules, 1) there has to be at most one molecule per detection volume, and 2) the signal from each single molecule has to be significant as compared to the background, i.e., the signal-to-noise ratio should be considerably high (119). While the first concern can be taken care of by diluting the sample (to picomolar range), the second requirement can be met by enhancing the signal of the fluorophore and reducing the background. The signal can be enhanced by using photochemically stable and bright fluorescent tags, such as fluorescent proteins (GFP, YFP, and CFP), organic dyes (rhodamines, cyanines, oxazines, bodipies, and perylenes), and inorganic quantum dots (120).
On the other hand, reducing the background has been achieved by minimizing the excitation and detection volume through confocal microscopy or the more commonly used microscopy in FRET experiments, total internal reflection fluorescence (TIRF) microscopy (116).

### 1.4.3 Protein-induced fluorescence enhancement

The most common dyes used in smFRET studies include members of Cyanine family like Cy3, Cy5, Alexa Fluor 647 etc. These dyes owing to their non-rigid polymethine bond connecting two indole rings generate differential fluorescence intensities depending on the environment of the dyes (Figure 1.9). This phenomenon, which has been adapted as a single molecule assay to study the binding of proteins to dye-labeled nucleic acids and is termed as protein-induced fluorescence enhancement (PIFE) (Figure 1.9) (121, 122). PIFE has become popular due to the fact that it needs single rather dual fluorescent labeling as in the case of FRET and can complement other single molecule techniques such as magnetic tweezer and smFRET (123). Studying protein-DNA interactions by smFRET can be done by multiple ways. Both the donor and acceptor dyes could be placed exclusively on the DNA or protein. Alternatively, one label could be placed on the DNA and the other on the protein. In all cases, it requires a significant change in distance upon changing the protein and/or DNA conformations or having a labeled protein, which is a non-trivial task considering current labeling approaches (122). In these instances, PIFE presents a clear advantage of directly
monitoring protein binding without requiring a substrate conformational change or a labeled protein.

The use of PIFE as a single molecule assay has also been championed due to its short distance sensitivity, which supersedes that of FRET (< 4 nm) (122). The use of alteration in the fluorescence intensity of the dye due to the PIFE effect has been used to study dynamics of multiple nucleic acid binding proteins including helicases, polymerases, filament forming proteins such as RAD51, RecA and RadA (124). The photophysical basis of the PIFE effect has recently been the subject of intense research. It is hypothesized that the PIFE effect is due to change in the isomerization rate between fluorescent trans and non-fluorescent cis state in the excited singlet state upon binding of a protein in the proximity of the dye.

Figure 1.9. Schematic showing protein-induced fluorescence enhancement. Upon protein binding, the fluorescence of certain fluorophores (mainly cyanine dyes) increases and this increase can be used as a signal to probe the dynamics of protein interaction with DNA.
The effect was attributed to be analogous to the quantum yield change observed upon increasing the solution viscosity. It is proposed that local viscosity changes the isomerization rate in favor of the trans excited state (121). Thus, the convenient explanation for PIFE has been that protein binding results in the stabilization of the trans excited state, thereby increasing the fluorescence parameters including fluorescence lifetime, quantum yield or intensity. In support of this explanation, Cy3B dye with a rigid inter-heterocyclic construction is resistant to any fluorescence change upon protein binding near the dye (125). First reported use of PIFE like ensemble assay was done using stopped flow experiments that probed the dynamics of translocation of motor protein along the DNA as well as the binding of the T7 RNA polymerase to DNA using fluorescence intensity change as a marker (126). This has been followed by a significant increase in the number of studies that use PIFE to study variety of biological questions (122). However, the rational design of PIFE experiments and control of PIFE effect is still illusive and is more of an arbitrary nature. Additionally, the viscosity based explanation seems to be inadequate to explain the opposite effect of PIFE as protein binding cannot decrease the local viscosity.

1.4.4 Why single molecule spectroscopy

Enzymatic reactions are by nature stochastic, inhomogeneous as well as have inherent protein and substrate dynamicity. The correlation between the dynamicity of protein conformational changes and its influence on the enzymatic reaction is
the linchpin of modern biochemical enzymology. Most of the currently known biological reactions cannot be achieved without catalysts, thus the need of enzyme catalyzed biochemical reactions assumes significance. This has resulted in studying of enzymatic mechanism as a core activity in biological physics (127, 128). Complex biological systems show spatio-temporal disorder during the different intermediary steps of a reaction including substrate-enzyme binding, chemical reaction, and product release. In this respect, spectroscopy has been the backbone for studying these reaction coordinates (119).

Early on, it was realized that biochemistry follows the dynamic interconversions of conformational states that biomolecules undergo. In other words, the enzymes functionality is fundamentally linked to them undergoing conformational changes. Techniques such as X-ray crystallography, NMR and Cryo-EM can provide a picture of conformational snapshots of the biomolecules in a particular state. Furthermore, NMR can provide information related to the dynamics of a biomolecule at various time scales, whereas cryo-EM can be used to probe specific conformations of the biomolecule. Although both techniques can provide snapshots of the biomolecules dynamics, they both fail to present a full picture of the system in motion or provide any information related to the interconversion dynamics of the various states (115). Techniques such as NMR, electron paramagnetic resonance (EPR) as well as bulk fluorescence-based methods such as photoinduced electron transfer (PET), fluorescence anisotropy as well as bulk ensemble FRET are very limited in providing information about the dynamics of conformational changes. Even for a simple two-state system, an
appropriate mathematical model needs to be applied to interpret the experimental data and generate any meaningful conclusions (119, 129). The challenge is due to the inherent asynchronicity of the biochemical reactions, which results in an overall signal that measures the weighted average of the intrinsic signal of individual conformations. Consequently, the signal from the transient states is obscured by the signal from the more populated states (115). Other challenges are due to the fact that biological systems are stochastic and bear no correlation among different states and pathways. The situation is further complicated due to biological systems comprising of multi-step pathways. Each step of these pathways shows its own conformational diversity and leads to further decrease in the synchronicity of the signal (126, 130).

The predicament of the ensemble assays is depicted in Figure 1.10 using the example of fluorescence emission. In this example, the properties of ensemble molecules in different states yield an average value, which is biased towards the properties of those molecules/states that are present in highest concentration. Although signals generated from these ensemble assays have high signal-to-noise ratio, the individual behavior of molecules cannot be distinguished as such.
Figure 1.10. Ensemble versus single molecule detection. Schematic showing the difference between ensemble and single molecule detections. The ensemble measurements record the average behavior of all the molecules and present average behavior of all the molecules. Single molecule measurements record signals from individual molecules and present the behavior of that particular molecule.

Single molecule observation tackles many of these challenges by deconvoluting the signals from individual molecules. Early single molecule studies of biological molecules began when W.E Moerner detected the single molecules in condensed phase by absorption spectroscopy (131). Since then single molecule spectroscopy has moved leaps and bounds, where now these techniques have become routine assays in the study of biological questions. Single molecule spectroscopy presents a dynamic view of biology by analyzing the behavior of individual molecules in action, thus distinguishing that molecule in study from the
rest of the ensemble. For that individual molecule, the parameter of interest is recorded and the distribution of that parameter can be generated by recording signals from multiple molecules. The example presented in Figure 1.10 again illustrates as to how single molecule imaging obviates the ensemble averaging in bulk assays. In this example, fluorescence emission of single molecule is recorded, which can describe the fluorescence properties of that particular molecule. This distribution contains information that supersedes the information provided in average values generated by ensemble assays especially for the systems that have spatial and temporal dynamicity.

The extraordinary virtues of single molecule detection can be further understood by considering the points below:

1. Spatial and temporal sensitivity and resolution:

As discussed above, the fundamental way the single molecule detection differs from other techniques is by giving access to the time series of the spatial fluctuations that a biomolecule follows in its reaction cycle. With modern single molecule techniques, spatial resolution of 1 nanometer and temporal resolution of sub milliseconds can be achieved (132).

2. Removing the ensemble averaging and revealing the inherent distribution:

Collating the signal from multiple single molecule events allows the construction of frequency distribution of the intrinsic distribution of the studied parameter. In the heterogeneous systems like enzymatic catalytic cycle, it reveals further details of sub populations that make up the ensemble (130).
3. Molecular subspecies and transient events:

Single molecule methods can reveal the hidden sub-species of the system and quantify the dynamicity between the subspecies. Furthermore, due to time evolution signal generated in single molecule, transient events can be revealed, which are otherwise hidden in ensemble averaging (119).

4. Non-uniform system dynamics:

Studying the coordinates of an enzymatic reaction requires synchronization of catalytic intermediates. However, this synchronization is unstable in dynamic systems such as biological molecules. This is due to biological molecules, like enzymes, having their inherent stochastic and uncorrelated pathways. Single molecule again tackles this challenge by generating temporal signal of each individual molecule that can be correlated with the dynamics of the system (115).

5. Thermodynamics of biological reactions:

By revealing the frequency distributions of the conformations that a biomolecule populates, the thermodynamic landscape of these systems can be studied. For example, in disordered proteins, the relative distribution of individual and interchanging conformations can be understood thermodynamically by correlating each conformation with its energy landscape (133).

6. Mechanical information of chemo-mechanical biological systems:

Single molecule force experiments have been used to understand how biomolecules respond to various perturbations. More importantly, devoid of
ensemble averaging, single molecule force experiments reveal mechanistic information about the transient states and folding pathways such as in single protein folding studies (134).

The study of structure-specific nucleases presents a particular challenge, since their reactions combine conformational changes in both the protein and the DNA (5). smFRET can be powerful in this front since it can report on the conformational changes associated with the DNA and the protein along the catalytic cycle (115). However, since smFRET provides a single vector distance change information, the technique can be limiting when multiple distances changes and binding interactions are to be studied. Coupling to other techniques like PIFE, which adds no extra cost to the design of experiment, can provide an additional dimension to the data (123).

1.5 Motivation of study

This thesis was motivated by the idea that single molecule fluorescence tools can be used to study the intricacies of substrate recognition and catalysis of structure-specific nucleases. Here, I used smFRET to study FEN1, as an archetype of these nucleases, with different temporal resolutions. I further coupled smFRET with biochemical, biophysical and ensemble fluorescence assays to provide key mechanistic findings on the substrate recognition mechanism in 5’ nucleases. Furthermore, during the course of the FEN1 study using smFRET, I stumbled upon
a novel finding that this nuclease quenches the fluorescence of Cy3 by a novel mechanism that is opposite to the mechanism of PIFE. I followed up on this observation and provided a comprehensive study that deciphered the photophysical basis of both fluorescence enhancement and quenching in cyanine dyes. This study provides much-needed information on how to design the experiments using protein-induced fluorescence modulation of cyanine dyes.
Chapter 2

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

2.1 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 5’nuclease mechanisms of 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 substrate and suppresses off-target cleavage. FEN1 sculpts DNA with diffusion-limited kinetics to test DNA substrate. This DNA distortion mutually “locks” protein and DNA conformation and enables substrate verification with extreme precision. Strikingly, FEN1 never misses cleavage of its cognate substrate while blocking probable formation of catalytically competent interactions with noncognate substrates and fostering their pre-incision dissociation. These findings establish FEN1 has practically perfect precision and that separate control of induced-fit substrate recognition sets up the catalytic selectivity of the nuclease active site for genome stability.

2.2 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 single-stranded(ss)/double-stranded(ds)-DNA junctions (Figure 2.1A), including nicks, gaps, flaps, bubbles and four-way junctions (28, 136-138). This conserved cleavage site despite diverse structures operates by uniformly binding to a bent junction to place the scissile phosphate near the active site (51, 63, 139) (Figure 2.1B). Yet, the mechanism underlying this specificity remains unclear including the question of whether 5’nucleases actively distort the DNA or selectively bind to a DNA that bends spontaneously (140, 141). Such mechanistic knowledge not only pertains to biological understanding but also to strategies for the design of specific inhibitors as potential cancer drugs (142). Catalysis is proposed to require changes in the protein conformation that assembles the active site (51, 60, 63, 102, 139, 143) and movement of the scissile phosphate closer to the catalytic metals (51, 63, 139). Although possible steps in the substrate selection and cleaving process have been described (51, 60, 63, 102, 139, 143), much of the control mechanisms by DNA and protein conformational changes that lead to exquisite catalytic selectivity and efficiency remain controversial and largely undetermined.

Flap endonuclease 1 (FEN1) and its substrate and product complexes provide a prototypic system for unveiling the extreme catalytic selectivity of structure-specific 5’nucleases. Whereas sequence-based specificity partially
explains the secret of replication fidelity, key information is missing about the basis for the structure-based excision required at ~50 million Okazaki fragment sites during human DNA replication. Strikingly, 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 (44). FEN1 recognizes dsDNA bearing a double-flap (DF) nick junction consisting of short ssDNA or ssRNA 5’flaps and strictly one nucleotide (nt) ssDNA 3’flap (Figure 2.1A) (42, 44). DF intermediates are produced during Okazaki fragment synthesis on the lagging strand and during long-patch base excision repair (144, 145). Mutations that reduce FEN1 expression or alter its activity are linked to cancers and genetic diseases (25, 27, 78, 138, 146). In cells, the 5’flap is complementary to the template strand, enabling the junction to equilibrate and form a single nt 3’flap (Figure 2.1A). Upon 5’flap cleavage, the 3’flap complements the newly unpaired template base to create a DNA ligase 1 sealable nick (Figure 2.1A). FEN1 contacts duplex DNA from both sides of the flap junction through a 100° bend stabilized by the interaction of the superfamly-conserved hydrophobic wedge with the junction (Figure 2.1B) (51). A superfamly-conserved helical gateway covered by a unique FEN1 helical cap forms a narrow cavity at the DNA junction. This gateway is suitable to select for threading ss 5’flaps with a free end (Figure 2.1B) (50, 51, 64, 141). Alternatively, clamping the 5’flap away from the active site is a proposed selection mechanism (63). A small cavity makes contacts with the 3’flap and may impose specificity for the single nt 3’flap (Figure 2.1B). Part of the cap-helical gateway, which contains catalytically indispensable residues,
and the 3’flap-binding pocket appear disordered without DNA (60), but it is ordered when bound to the 3’flap (Figure 2.1B) (51, 55), suggesting DNA-induced ordering of the cap-helical gateway. Comparison of substrate and product complexes shows that the scissile phosphate nucleotide is fully paired at the nick junction and away from the active site’s metal ions, suggesting that unpairing flanking nucleotides may move ssDNA into the active site (51).

By building upon these strong static structural and ensemble biochemical results, we reasoned that single-molecule experiments could resolve mechanistic unknowns by deconvoluting DNA bending, protein disorder-to-order transitioning, active-site assembly and incision. Like other 5’nucleases, FEN1 displays maximum catalytic efficiency for its cognate substrate but it is only residually active on substrates that vary only slightly (42, 44). To define the mechanism for this catalytic efficiency and selectivity, we used single-molecule (sm)FRET at a millisecond to sub-millisecond temporal resolution to simultaneously measure in real time DNA conformational changes and catalysis when FEN1 encounters cognate or noncognate flap substrates as well as when the disorder-to-order transition or the active site is perturbed.

2.3 Results

2.3.1 FEN1 actively bends the DNA

A major question in DNA damage recognition is whether the DNA distortion observed in protein–DNA complexes occurs spontaneously and is captured by the protein (termed conformational selection) or if the protein actively sculpts the DNA.
into the distorted conformation (Figure 2.1C). To determine which is the case for FEN1, we started by establishing the conformational state of DF substrates alone using an ideal non-equilibrated (NonEQ) DF substrate containing 6 nt ssDNA 5’flap and 1 nt ssDNA 3’flap primers with no complementarity with the template strand (NonEQ DF-6,1). DNA bending was monitored by placing an Alexa Fluor-647 acceptor 12 nt into the upstream dsDNA and a Cy3 donor at the 5’flap end (NonEQ DF-6,1Flap) (Figure 2.1D) or 15 nt into the downstream dsDNA (NonEQ DF-6,1dsDNA) (Figure 2.1E). The substrates contained biotin to allow for immobilization on a polyethylene glycol-coated coverslip via biotin-NeutrAvidin linkage (Figure 2.1D,E). 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 (147) as the primary method or the confocal mode for higher temporal resolution.

The single-molecule time traces of the substrate alone showed a single FRET state with no transition from this state (Figure 2.1D,E); FRET efficiency histograms generated from multiple single molecules fit to a single Gaussian (Figure 2.1.1A,B). The conformer of the duplex arms of the substrate was insensitive to variation in the concentration of divalent metal ions (Mg^{2+}) (Figure 2.1.2A,B), type of metal ion (Mg^{2+} versus Ca^{2+}) (Figure 2.1.1A,B and Figure 2.1.2A,B) or 5’flap length (Figure 2.1.2A,B). The flap-labeling scheme was sensitive to variation in the Mg^{2+} ion concentration when the 5’flap length exceeded 6 nt (Figure 2.1.2B). This could explain why previous work suggested that a double-flap substrate is dynamic structure (140). To test for short-lived alternative
conformers, we used confocal-based smFRET to increase the temporal resolution to 5 ms on surface-immobilized DNA and to sub-ms on freely diffusing DNA in solution. Importantly, we found that substrate remained as a single conformer (Figure 2.1.2C,D). Potential mean force molecular dynamics (MD) simulations showed that extended DNA (~165°) was the most energetically favorable conformer in DF-6,1 (Figure 2.1F and Figure 2.1.3A,B). Base stacking at the nick junction and between the 3’flap base and the first base on the 5’flap stabilized this extended conformer (Figure 2.1.3B). The energetic cost required to bend the DNA up to ~140° was low (Figure 2.1F) and similar to that in dsDNA (148). This was followed by a rapid increase in the energy required to surpass a significant barrier of ~14 kcal/mole to break the base stacking and bend the DNA (Figure 2.1F). These data suggest that DF substrate remains in an extended form that must be actively bent.

In fact, adding FEN1 to DNA NonEQ 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 2.1D,E). 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 NonEQ DF-6,1_{Flap} and NonEQ DF-6,1_{dsDNA}, respectively (Figure 2.1.1A,B). This dissociation constant agreed with the nM range of \(K_m\) from bulk cleavage assays (Figure 2.1.4A) (44) and the DNA binding dissociation constant (\(K_{d\text{-binding}}\)) of FEN1 and NonEQ DF-6,1 as determined by surface plasmon resonance (SPR) (Figure 2.1.4B). The change in FRET in both NonEQ DF-6,1_{Flap}
and NonEQ DF-6,1 dsDNA was confirmed by time-resolved bulk FRET measurements (Figure 2.1.4C).

To mimic the in vivo junction, we used equilibrated (EQ) DF-6,1 (Figure 2.1A). FEN1 actively bent EQ DF-6,1 dsDNA to a similar extent and similar $K_{\text{d-bending}}$ as DF-NonEQ 6,1 dsDNA (Figure 2.1.1B,C). Nonetheless, time traces showed multiple transitions between bent and unbent states (Figure 2.1G). 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 ($\tau_{\text{bending}}$) and unbent ($\tau_{\text{unbending}}$) states at increasing FEN1 concentrations indicated that the apparent first-order rate constant for DNA bending ($k_{\text{bending}} = 1/\tau_{\text{bending}}$) increased linearly while that for DNA unbending ($k_{\text{unbending}} = 1/\tau_{\text{unbending}}$) remained constant (Figure 2.1G). This is the trend 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 ($1.4\pm0.03\times10^8 \text{ M}^{-1} \text{ s}^{-1}$), and the average value of $k_{\text{unbending}}$ ($k_{\text{off-unbending}}$) was $0.45\pm0.05 \text{ s}^{-1}$ (Figure 2.1G).
Figure 2.1 Active junction bending by structure-specific 5’ nucleases. (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. FEN1 alone (1ULI.pdb) (60) and in complex with bent DNA (3Q8L.pdb) (51), highlighting the various structural features of FEN1 and the regions that undergo through disorder-to-order transition upon DNA binding. (C) Active DNA versus DNA conformational capturing models for forming the FEN1 complex with the bent DNA conformer. Monitoring DNA
bending of FEN1 and non-equilibrated DF-6,1 using the flap-labeling scheme (NonEQ DF-6,1Flap) (D) and internal labeling-scheme (NonEQ DF-6,1dsDNA) (E). For each labeling, a schematic of the donor and acceptor positions (upper panel) and smFRET time traces of the substrate alone (middle panel) and in presence of FEN1 (lower panel) are shown; change in FRET upon DNA bending in each labeling scheme is highlighted. (F) Analysis of the structure of NonEQ DF-6,1 by MD simulations. The effective free energy profile (PMF) from adaptive biasing force calculations is shown. (G) Bending of equilibrated DF-6,1 (EQ DF-6,1dsDNA) by FEN1. smFRET time traces of EQ DF-6,1dsDNA alone (upper panel) and in the presence of FEN1 (middle panel) 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. $K_{d-bending} = k_{off-unbending}/k_{on-bending}$. (H) Bending of nicked substrate using the internal labeling scheme (NickdsDNA) by EXO1. A schematic of the donor and acceptor positions (upper panel), smFRET time traces of NickdsDNA alone and in the presence of EXO1 (middle panels) and analysis of its $k_{on-bending}$, $k_{off-unbending}$ and $K_{d-bending}$ (lower panel) is presented. Donor and acceptor are at identical positions to those in DF-6,1dsDNA in Figure 2.1E. $k_{on-bending}$, $k_{off-unbending}$ and $K_{d-bending}$ were calculated as in 1G. All TIRF-smFRET experiments were acquired at 100 ms. The following figure supplements are available for Figure 2.1: Figure 2.1.1, Figure 2.1.2, Figure 2.1.3, Figure 2.1.4, Figure 2.1.5.
Figure 2.1.1 Bending of equilibrated and non-equilibrated DF-6,1 by FEN1. (A) TIRF-based single-molecule FRET (smFRET) of FEN1 and NonEQ DF6,1\textsubscript{Flap}. Histograms of NonEQ DF6,1\textsubscript{Flap} alone (high FRET) and in the presence of FEN1 (low FRET) (upper panel). $K_{d\text{-bending}}$ (lower panel) is calculated by fitting the percentage of the bent conformer from the FRET histograms at various concentrations of FEN1 with a non-linear least squares regression; the percentages of unbent and bent DNA are calculated by fitting two Gaussians. The uncertainty in calculating the percentage of bent DNA and $K_{d\text{-bending}}$ correspond to the standard deviation of triplicate measurements and the non-linear least squares regression fit, respectively. (B) $K_{d\text{-bending}}$ of FEN1 on NonEQ DF-6,1\textsubscript{dsDNA}. $K_{d\text{-bending}}$ is calculated as described in Figure 2.1.1A. (C) $K_{d\text{-bending}}$ of FEN1 on EQ DF-6,1\textsubscript{dsDNA}. $K_{d\text{-bending}}$ is calculated as described in Figure 2.1.1A.
Figure 2.1.2 Flap substrates exist as a stable extended conformer. (A) Effect of various Mg$^{2+}$ concentration on DF-6,1. TIRF-smFRET histograms of NonEQ DF6,1$_{dsDNA}$ (left panel) and NonEQ DF6,1$_{Flap}$ (right panel) with increasing concentration 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 NonEQ DF12,1$_{dsDNA}$ (left panel) and NonEQ DF12,1$_{Flap}$ (right panel) with increasing concentration 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. (C) Burst confocal-smFRET histograms from freely diffusing DNA in solution of NonEQ DF6,1\textsubscript{dsDNA} (0.5 nM) (upper panel) and NonEQ DF6,1\textsubscript{Flap} (0.5 nM) (lower panel) acquired at sub-ms temporal resolution. No enrichment of other FRET conformers was observed upon increasing the temporal resolution from 160 ms (shown in Figure 2.1.1A,B) to sub-ms. (D) Confocal-smFRET time traces of surface-immobilized NonEQ DF6,1\textsubscript{Flap} (upper panel) and NonEQ DF6,1\textsubscript{dsDNA} (lower panel) with 5 ms temporal resolution.

**Figure 2.1.3** MD simulations of the conformational states and DNA bending energy of nick and various flap structures. (A) Definition of the DNA bending angle. The
bending angle is calculated between two vectors. One vector was defined by using the center of the masses of two blocks of nucleotides: block I (green) and block II (green) and one vector was defined with the center of mass of block III (blue) and block VI (blue). (B), (C), (D), and (E) are the averaged structure (upper panel) and histogram of the DNA bending angle (lower panel) for NonEQ DF-6,1, Nick, EQ DF-6,1 and SF-6,0, respectively, taken from the MD simulations. These MD simulations demonstrated that the extended DNA structure was the most energetically favorable conformer in all tested substrates and that these substrates could sample bent conformers up to 140°, which are equivalent to those observed for dsDNA (148).
Figure 2.1.4 Ensemble assays support single molecule results. (A) Cleavage of NonEQ DF-6,1\textsubscript{dsDNA} (0.5 nM) by bulk assays. A plot of initial rates (\(v_0\), nM.min\(^{-1}\)) in relation to the FEN1 concentration fitted with a generalized non-linear least-squares regression using a Michaelis-Menten model. The values for \(v_0\) were estimated using bulk cleavage assays with different time intervals as described in the Methods. This plot was used to determine the steady-state \(K_m\). Uncertainty in \(K_m\) corresponds to the standard error of the fit. (B) SPR binding studies of FEN1 and NonEQ DF-6,1\textsubscript{dsDNA} (left panel), SF-6,0\textsubscript{dsDNA} (middle panel) and DF-30,1\textsubscript{blocked-dsDNA} (right panel). The sensorgram of binding of increasing concentrations of FEN1 is shown (left panel). The maximum response units (RU) reached at each FEN1 concentration were fitted using the steady-state affinity model to obtain the equilibrium dissociation binding constant (\(K_d\)-binding) (lower panel). The uncertainty corresponds to the standard deviation of \(N = 2\) runs. 170 RU, 70 RU and 90 RU of NonEQ DF-6,1\textsubscript{dsDNA}, SF-6,0\textsubscript{dsDNA} and DF-30,1\textsubscript{blocked-dsDNA} were immobilized on the surface, respectively. (C) Bulk-FRET measurements of NonEQ DF-6,1. The FRET efficiency of the donor-acceptor pair in NonEQ DF-6,1\textsubscript{Flap} at different FEN1 concentrations (left panel) and relative fluorescence lifetimes (\(\tau_{\text{donor(enzyme)}}/\tau_{\text{donor}}\)) at different FEN1 concentrations without acceptors (middle panel) are shown. The results show that FEN1 binding influenced the fluorescence intensity of the donor in the flap-labeling scheme. \(K_{d\text{-bending}}\) after correcting for the effect of the donor was similar to that obtained from uncorrected apparent FRET (Figure 2.1.1C); \(K_{d\text{-bending}}\) was calculated using a standard quadratic equation for simple bimolecular association as described in Methods and the uncertainty corresponds to the standard error of the fit. The uncertainty corresponds to the standard error of the fit. Bulk-FRET measurements of NonEQ DF-6,1\textsubscript{dsDNA} (right panel). The relative fluorescence lifetime (\(\tau_{\text{donor(enzyme)}}/\tau_{\text{donor}}\)) of DF-6,1\textsubscript{dsDNA} at different FEN1 concentrations without acceptors (right panel), showing no effect on the donor fluorescence intensity upon FEN1 binding in the internal-labeling scheme. (D) Burst confocal-smFRET histograms from freely diffusing substrate missing the 3'flap but containing the 5'flap (SF6,0\textsubscript{dsDNA}) (0.5 nM) in solution acquired at sub-ms temporal resolution (upper panel). \(K_{d\text{-bending}}\) (lower panel) calculated as described in Figure 2.1.1A.
Figure 2.1.5 Active bending of nicked DNA by EXO1. TIRF-smFRET histograms of Nick\(_{\text{dsDNA}}\) alone and in the presence of EXO1 (left panel) and \(K_{d\text{-bending}}\) (right panel). \(K_{d\text{-bending}}\) is calculated as described in Figure 2.1.1A. The donor and acceptor in Nick\(_{\text{dsDNA}}\) are at identical positions to those in NonEQ DF6,1\(_{\text{dsDNA}}\).

It is unclear what caused the much higher \(K_{d\text{-bending}}\) reported in our earlier work (141), but we suggest that both slower association and faster dissociation rates were influences. Nonetheless, the FRET states of NonEQ DF-6,1\(_{\text{Flap}}\) alone and when bent by FEN1 and the relative comparison of bending the cognate with the noncognate substrates are similar under low and high \(K_{d\text{-bending}}\) conditions as shown below.

To see if active bending of the ss/ds-DNA junction may be a conserved feature in 5’nucleases, we tested human mismatch repair exonuclease 1 (EXO1), which recognizes an ideal junction of either a nick or a 3’ overhang (63). EXO1 actively bent a DNA nick with diffusion-limited \(k_{\text{on-bending}}\) (Figure 2.1H); the donor and acceptor had identical positions to those in NonEQ DF-6,1\(_{\text{dsDNA}}\). Free MD simulations showed that the nick behaved similarly to flap substrates for bending.
angles in the 140°−180° range (Figure 2.1.3C). The bent conformer of EXO1 had similar FRET to that of FEN1 (Figure 2.1.1B and Figure 2.1.5), consistent with the structures of their DNA complexes (51, 63).

2.3.2 FEN1 never misses cleavage of its correct substrate

To examine 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 2.2A). Time traces indicated that FEN1 always bent NonEQ DF-6,1\(^{\text{Flap}}\) before cleaving the Cy3-containing 5’flap; remarkably every DNA bending event led to a successful cleavage reaction (Figure 2.2A, Figure 2.2.1A). We confirmed DNA bending before cleavage by the clear anti-correlated change in the donor and acceptor intensities (Figure 2.2A). 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 2.2.2A,B). This confirms that the loss of donor particles is due to 5’flap cleavage and not due to donor photobleaching. Analysis of FRET values before cleavage from individual time traces showed that FEN1 cleaved NonEQ DF-6,1\(^{\text{Flap}}\) from a fully bent state (Figure 2.2B) and remained bent for 160±7 ms prior to cleavage (Figure 2.2C). FEN1 cleavage generates two products: 5’flap ssDNA and nicked dsDNA (Figure 2.2.1A). 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 (44, 149). Consistent with
these findings, we also observed that the lag time before cleavage is not influenced by the presence of excess 5’flap ssDNA (Figure 2.2.2C). Furthermore, SPR showed only residual transient binding of FEN1 to ssDNA at concentrations that were orders of magnitude above $K_{d\text{-}bending}$ of DF-6,1 (Figure 2.2.2D). Our single-molecule cleavage measurement is not inhibited by lack of 5’flap ssDNA product release. Therefore single-turnover $k_{cat}$ ($k_{STO}$) could be determined directly from the lag time prior to cleavage ($k_{STO} = 1/\tau_{before \text{ cleavage}}$). However, since 5’flap release would still contribute to the dwell time before cleavage, our single turnover should be treated as an apparent value. Notably, our $k_{STO}$ (6.3±0.2 s$^{-1}$) was comparable to that determined by bulk cleavage assays ($k_{STO}$ 12.3 s$^{-1}$→5 s$^{-1}$) (150, 151), with the slight difference explained by the lower reaction temperature in the single molecule assays. 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.

The lag time distribution prior to cleavage shows a rise and decay (Figure 2.2C), suggesting that the underlying catalytic mechanism after the diffusion-limited DNA bending step involves two or more steps, as a single-step process will have a single exponential decay. We reasoned that these steps likely include 3’flap-induced disorder-to-order transitioning and cleavage chemistry. To test this idea, we employed glycerol and sucrose 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 2.2D and Figure 2.2.1B); a similar effect was observed with
sucrose (Figure 2.2.1C). Yet, $k_{STO}$ was unaffected by polyethylene glycol-8000, a high-molecular-weight viscogen that is too large to interfere with local protein conformational changes (Figure 2.2.1D). $k_{STO}$ 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. The shape of the histograms in the presence of viscogen however remains a rise and decay (Figure 2.2.1B), in contrast to the prediction of collapsing into a single exponential decay should protein-ordering acts in a single rate-limiting step. This suggests that the 3’flap-induced protein ordering likely involves multistep processes that are being slowed down by the presence of viscogen and/or these multisteps control different rate-limiting steps during catalysis such as DNA unpairing and/or DNA shifting into the active site. Biologically relevant, the cleavage behavior from the first DNA bending and the $k_{STO}$ were similar whether there was a deliberate mispaired 3’flap (NonEQ DF-6,1; Figure 2.2A,C and Figure 2.2.1A) or an equilibrating 3’flap (EQ DF-6,1; Figure 2.2E,F and Figure 2.2.1E), consistent with bulk cleavage reactions (149).
Figure 2.2. Cleavage of cognate substrate by FEN1. (A) Cleavage of NonEQ DF-6,1\textsubscript{Flap}.
Schematic showing the simultaneous monitoring of DNA bending and 5’flap cleavage at the single-molecule level (upper panel). A representative smFRET time trace with a zoomed-in view showing the cleavage of NonEQ DF-6,1\textsubscript{Flap} in which FEN1 never misses the opportunity to bend the DNA and cleave it (lower panel). (B) FRET of the bent state before cleavage of NonEQ DF-6,1\textsubscript{Flap} fitted with a Gaussian distribution from multiple cleavage events. (C) Dwell times of the bent state prior to cleavage of NonEQ DF-6,1\textsubscript{Flap} fitted with a gamma distribution to calculate average dwell time ($\tau$\textsubscript{avg}) from number of independent experiment $N = 3$; the uncertainty corresponds to the standard error of the mean. Single turnover $k_{cat}$ ($k_{STO}$) is $= 1/\tau$\textsubscript{avg}. Cleavage was performed at 50 ms temporal resolution. (D) Effect of low molecular weight viscogen (glycerol) on disorder-to-order transitioning in FEN1. Graph showing relative $k_{STO}$ of NonEQ DF6,1\textsubscript{Flap} cleavage upon addition of glycerol at increasing relative viscosity from $N = 2–3$ fitted with a linear regression to calculate the slope of the curve; the error corresponds to the standard error of the mean. $k_{STO}$ was determined as in Figure 2.2C. (E) A representative smFRET time
trace showing the cleavage of EQ DF-6,1_{Flap} in which FEN1 never misses the opportunity to bend the DNA and cleave it. (F) A histogram showing the distribution of dwell times of the bent state prior to cleavage of EQ DF6,1_{Flap}. \( \tau_{avg} \) from \( N = 3 \) is calculated as in Figure 2.2C at a temporal resolution of 50 ms. The following figure supplements are available for Figure 2.2: Figure 2.2.1, Figure 2.2.2.
Figure 2.2.1 Single-molecule cleavage of cognate substrates and probing conformational changes to FEN1 by viscogens. (A) Representative TIRF-smFRET
time traces showing the cleavage of NonEQ DF6,1\textsubscript{Flap} in which FEN1 always bends and cleaves the DNA. (B) Histograms showing the distribution of dwell times of bent state prior to the cleavage of NonEQ DF6,1\textsubscript{Flap} by FEN1 in the presence of varying concentrations of glycerol. For each concentration, the histogram was fitted and $\tau_{\text{avg}}$ was calculated from N = 4 as described in Figure 2.2C. (C) and (D) Histograms showing the distribution of dwell times of the bent state prior to the cleavage of NonEQ DF6,1\textsubscript{Flap} by FEN1 in the presence of 40% sucrose or 10% PEG-8000, respectively. The histogram was fitted and $\tau_{\text{avg}}$ was calculated from N = 4 as described in Figure 2.2C. (E) Representative TIRF-smFRET time traces showing the cleavage of NonEQ DF6,1\textsubscript{Flap} in which FEN1 always bends and cleaves the DNA.
Figure 2.2.2 Controls for assigning donor loss to the cleavage and immediate departure of the 5’flap. (A) (Left panel) bar chart comparing the Cy3 donor signal lost in cognate substrate due to photobleaching (+Ca\(^{2+}\)) or to both photobleaching and incision (+Mg\(^{2+}\)) in presence of FEN1. The donor loss increases significantly in the presence of Mg\(^{2+}\) as compared to what would be expected for its loss in the presence of Ca\(^{2+}\). (Right panel) The detailed classification of all single-molecule time traces in the field of view for cognate substrate in the presence of Mg\(^{2+}\) as explained in Methods (right panel). This shows that most of the traces exhibited donor loss instantaneously only upon FEN1 bending, demonstrating that donor loss in the presence of Mg\(^{2+}\) is due 5’flap cleavage by FEN1. The uncertainty corresponds to the standard deviation between multiple movies in the presence of either CaCl\(_2\) or MgCl\(_2\). (B) Time-dependent loss of Cy3 molecules under binding (Ca\(^{2+}\)) or cleavage (Mg\(^{2+}\)) conditions, showing that sudden loss of Cy3 coincides with FEN1 introduction into the focal volume in the case of Mg\(^{2+}\), while Cy3 loss in the case of Ca\(^{2+}\) is minimal and not abrupt. (C) Histograms showing the distribution of dwell times of the bent state prior to the cleavage of NonEQ DF6,1\(_{\text{Flap}}\) by FEN1 in the presence of excess 10 nt ssDNA as a competitor, showing no effect on the cleavage dwell time when compared to cleavage of NonEQ DF6,1\(_{\text{Flap}}\) by FEN1 (Figure 2.2C). The histogram was fitted and \(\tau_{\text{avg}}\) was calculated from N = 2 as described in Figure 2.2C. (D) SPR binding of FEN1 to 22 nt ssDNA, showing only residual binding of FEN1 at extreme concentrations of FEN1. (E) Similar analysis to that described in Figure 2.2.2A (left panel), showing that most of the loss of the donor molecules in the noncognate substrate SF-6,0 resulted from cleavage by FEN1 in the presence of Mg\(^{2+}\).

2.3.3 FEN1 actively creates a 3’flap in equilibrated DF substrate

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 NonEQ DF while
maintaining its 5'flap (a substrate termed single flap (SF)) decreased FEN1 cleavage activity by 34-fold (44). Time traces on surface-immobilized SF-6,0Flap accessed at 5 ms using confocal-based smFRET showed that FEN1 actively bent SF (Figure 2.3A). 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 NonEQ DF-6,1 (Figure 2.3A). However, the $k_{\text{on-bending}}$ remained limited by diffusion and similar to that of EQ DF-6,1 dsDNA (Figure 2.3A). $K_{d\text{-bending}}$ was 50-fold higher than for NonEQ DF-6,1 (Figure 2.3A), consistent with that observed from confocal-based smFRET with burst analysis of freely diffusing SF-6,0 dsDNA (Figure 2.1.4D) and the markedly increased $K_{d\text{-binding}}$ by SPR (Figure 2.1.4B). These results show that a 3'flap is not required for DNA bending but it is critical for DNA binding stability.

We next established whether the equilibrated junction existed as a SF or DF by comparing the FRET states of substrates containing only 5'flap, nonequilibrated 5'- and 3'-flap, or equilibrated 5'- and 3'-flap. Interestingly, we found that the FRET values of the substrate alone in EQ DF-6,1 dsDNA and in SF-6,0 dsDNA were similar (E~0.27), but less than that in NonEQ DF-6,1 dsDNA (E~0.34) (Figure 2.3B; Figure 2.3.1). The geometry of the equilibrated DF and SF was slightly less extended than that of dsDNA (E~0.23) (Figure 2.3B; Figure 2.3.1). This suggests that the equilibrated junction existed as a SF, which was shown by free MD simulations not to equilibrate to a DF (Figure 2.1.3D,E).

The finding that FEN1 cleaves equilibrated and non-equilibrated DF substrates with equal activity (149) prompted us to propose that FEN1 actively converts the equilibrated junction from a SF to a DF by pulling out its 3'end to
create a 3′flap. The nick position relative to the donor and acceptor would differ by one base pair if EQ DF-6,1\textsubscript{dsDNA} existed in a SF form compared to that in NonEQ DF-6,1\textsubscript{dsDNA} (Figure 2.3B; Figure 2.3.1). Creating a 3′flap would thus move the junction back to the identical position of that in NonEQ DF-6,1\textsubscript{dsDNA} (Figure 2.3B; Figure 2.3.1). In a control experiment, we showed that we were able to detect junction movement by one base pair because a deliberate difference of one base pair in the nick position of EQ DF-6,1\textsubscript{dsDNA} would result in a detectable difference in the FRET value of their bent conformers (E\textasciitilde0.46 versus E\textasciitilde0.54) (Figure 2.3B; Figure 2.3.1). The bent conformer in EQ DF-6,1\textsubscript{dsDNA} had identical FRET to that in NonEQ DF-6,1 (E\textasciitilde0.54) (Figure 2.3B; Figure 2.3.1), demonstrating that the equilibrated nick junction must have moved by one base pair. These data therefore suggest that FEN1 actively sculpts the 3′end of its \textit{in vivo} equilibrated nick junction to create a 3′flap and to drive its ordering. The active DNA bending and its subsequent use to induce protein ordering through active formation of a 3′flap suggest that there is an induced-fit mechanism between FEN1 and DNA that functions in a mutual way.
Figure 2.3. Active sculpting of the 3’end of a nick Junction creates a 3’flap and drive protein ordering. (A) Confocal-based smFRET time traces of surface-immobilized SF6,0\textsuperscript{flap} alone (left panel) and in the presence of FEN1 (middle panel) acquired at 5 ms,
showing rapid transitions from high to low FRET upon DNA bending. $k_{\text{on-bending}}$ and $k_{\text{off-unbending}}$ of SF6,0Flap by FEN1 (right panel) calculated as in Figure 2.1G. (B) FEN1 actively creates a 3'flap at the nick junction of cognate and noncognate substrates. Determining the status of the 3'flap in equilibrated DF and SF junctions by comparing the FRET states of various nick-junction positions in NonEQ DF6,1dsDNA, EQ DF6,1dsDNA and SF6,0dsDNA in the absence or presence of FEN1. 0.5nM DNA and saturating concentrations of FEN1 were used (1000 nM for SF6,0dsDNA and its one base pair shift construct and 200 nM for the remaining constructs). FRET values were determined by fitting the burst confocal-smFRET histograms from freely diffusing DNA at sub-ms temporal resolution with a Gaussian. FRET is reported as percentage and the uncertainty corresponds to the standard deviation of $N = 3$. The following figure supplements are available for Figure 2.3: Figure 2.3.1.
Figure 2.3.1 All histograms corresponding to the data shown in Figure 2.3B. Histograms were fit with Gaussian distribution to get the mean of the distribution. Uncertainties correspond to the error of the fit.

2.3.4 Substrate verification by 3’flap-induced protein ordering

The ability of FEN1 to actively create a 3’flap at the nick junction of its equilibrated DF substrate raises the possibility that it could also create a 3’flap at nick junctions of noncognate substrates. This mechanism would explain why the cleavage site is shifted by 1 nt in SF versus DF substrates (44). We found that the FRET value of the bent conformer in SF-6.0dsDNA was also similar to that of NonEQ DF-6.1dsDNA (Figure 2.3B; Figure 2.3.1), demonstrating that the nick junction must have moved by one base pair. In a control experiment we showed that a shift of one base pair in the nick position in SF-6.0dsDNA resulted in a detectable difference in the FRET of the bent conformer as observed in the case of EQ DF-6.1dsDNA (Figure 2.3B; Figure 2.3.1). This indicates that FEN1 creates 3’flaps at noncognate nick junctions, suggesting that there is another requirement during substrate validation.

In an *in vivo*-equilibrating junction, the nick structure would be maintained, while in noncognate substrates, a one-nucleotide mismatch would be added at the junction (Figure 2.3B; Figure 2.3.1). FEN1 discriminates against such a structure with 33-fold reduced activity (152). Here, the $K_{d-bending}$ of DF-6.1 containing a one-nucleotide mismatch at the junction (termed DF-7,1mismatch(1nt)-Flap) increased by seven-fold (Figure 2.4A), with the time traces showing a less-stable bent conformer (Figure 2.4B and Figure 2.4.1A). Since FEN1 forms 3’flaps for both cognate and
noncognate substrates, only the junctions that are fully paired are therefore stably bent.

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 α2 in the gateway and stacks with K128 on α5 in the cap (51) (Figure 2.4C). Mutating R47 to A (FEN1-R47A) disabled FEN1’s cleavage on DF substrate to a similar extent as wild-type (wt)-FEN1’s cleavage on SF-6,0 (51). To test this allosteric signaling idea, we maintained the 3’flap binding using NonEQ DF-6,1dsDNA while altering R47 using FEN1-R47A. The defects in $K_{\text{d-bending}}$ and $k_{\text{off-unbending}}$ resemble those in wt-FEN1 on SF-6,0 (Figure 2.4A,B and Figure 2.4.1B). We determined that FEN1-R47A engaged the 3’flap because $K_{\text{d-bending}}$ and $k_{\text{off-unbending}}$ increased when SF-6,0dsDNA was used rather than NonEQ DF-6,1dsDNA (Figure 2.4A,B).

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 (50, 64, 141) or that the 5’flap may be clamped away from the active site for catalysis (63). To test possible threading 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-30,1\textsubscript{blocked}-dsDNA) impaired DNA bending to comparable $K_d$-bending and $k_{off}$-unbending of SF-6,0 (Figure 2.4A,B and Figure 2.4.1C), consistent with the markedly reduced $K_d$-binding captured by SPR (Figure 2.1.4B). Notably, the bent DNA in the unthreaded substrate was distorted, but it did not reach the same final FRET state as when the 5’flap was not blocked (Figure 2.4.1D). This significant DNA distortion was masked in our previous experiment that relied on a flap labeling scheme to infer to the geometry of the blocked-threaded complex (141). Importantly, our new results indicate that 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 $K_d$-bending of the unthreaded substrate upon removal of its 3’flap (SF-30,0\textsubscript{blocked}-dsDNA) (Figure 2.4A) 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.

Collectively these results demonstrate that FEN1 bends both cognate and noncognate substrates and that $K_d$-bending is higher for noncognate substrates. This is consistent with our previous findings under high $K_d$-bending conditions (141). 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. Our observation of FEN1’s ability to significantly bend the DNA in the blocked-threaded complex challenges our previous conclusion that 5’flap threading is strictly required to induce DNA bending (141). Our new 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.

Figure 2.4. Verification of the bent DNA conformer by the 3’flap-induced protein ordering. (A) Bar chart comparing $K_{d\text{-bending}}$ 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 biotin-NeutrAvidin on 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). (B) Bar chart comparing $k_{off\text{-unbending}}$ for FEN1-WT or FEN1-R47A on various non-equilibrating flap substrates using the internal labeling scheme. The lower estimate of $k_{off\text{-unbending}}$ for FEN1-WT on DF-6,1 corresponds to the 60 s acquisition time where transitions were rarely detected. $K_{d\text{-bending}}$ and $k_{off\text{-unbending}}$ are calculated as in Figure 2.1.1A and Figure 2.1G, respectively. $k_{off\text{-unbending}}$ 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. The smFRET technique and temporal resolutions used in Figure 2.4A,B are described in Figure 2.4.1. (C) 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-caps the α2 in the gateway.
(highlighted in green) and stacks with K128 on α5 in the cap (highlighted in purple) (3Q8L.pdb) (51). The following figure supplements are available for Figure 2.4: Figure 2.4.1.

Figure 2.4.1 Bending kinetics of various noncognate substrates. (A) TIRF-smFRET time traces of DF-7,1 mismatch(1nt)-Flap alone and in the presence of FEN1. (B) Surface-immobilized confocal-smFRET time traces of NonEQ DF-6,1 Flap in the presence of FEN1-R47A at 5 ms temporal resolution. (C) The effect of blocking 5'flap threading on DNA bending by FEN1. 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 of DF-30,1 blocked-dsDNA) prior to the addition of FEN1 (upper panel).
Surface-immobilized confocal-smFRET time traces of DF-30,1\textsubscript{blocked-dsDNA} alone and in the presence of FEN1 at 5 ms temporal resolution (lower panel). The substrate was immobilized by surface-coated-NeutrAvidin via the biotin group on the 5’flap. (D) A bar chart comparing final bent FRET states of DF-30,1\textsubscript{trapped-dsDNA}, DF-30,1\textsubscript{blocked-dsDNA}, SF-30,0\textsubscript{trapped-dsDNA} and SF-30,0\textsubscript{blocked-dsDNA} using burst confocal-smFRET histograms from freely diffusing substrates acquired at sub-ms temporal resolution. FEN1 concentrations were 5000 nM for SF-30,0\textsubscript{internal-blocked}, 1000 nM for DF-30,1\textsubscript{internal-blocked}, 1000 nM for SF-30,0\textsubscript{internal-trapped} and 200 nM for DF-30,1\textsubscript{internal-trapped}. 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. (E) TIRF-smFRET time traces of DF6,2\textsubscript{Flap} alone and in the presence of FEN1.

2.3.5 FEN1 avoids off-target DNA cleavage in the DNA lockdown step

To test the mechanism for assembly of catalytically competent active sites for cognate substrate incision, we compared the lifetime of the bent conformer to the lag time for cleaving correct versus incorrect substrates. In SF-6,0, the lifetime of the bent conformer was ~3.5 fold shorter than the required lag time prior to cleaving the cognate substrate (Figure 2.3A and Figure 2.2C). Traces of single-molecule cleavage showed that SF-6,0 underwent multiple cycles of DNA bending and unbending before a successful DNA bending event led to 5’flap cleavage (Figure 2.5A; Figure 2.5.1A; Figure 2.2.2E). These abortive DNA bending events are masked in bulk cleavage assays, which leads to underestimation of both $k\text{STO}$ and the accuracy of FEN1 cleavage. Following the FEN1 cleavage reaction at the single-molecule level clearly leads to additional information. Similar results were observed in the cleavage of DF-6,1 by FEN1-R47A (Figure 2.5B and Figure
These results show that destabilizing the bent DNA intermediate to rates that are limiting for catalysis reduces the probability of assembling catalytically competent active sites.

Interestingly, we also observed similar bending behavior without cleavage of FEN1 on noncognate substrates under conditions in which the lifetime of the bent conformer does not limit catalysis. DF-7,1\text{mismatch(1nt)}-Flap and DF containing 2 nt 3’flap (DF-6,2\text{Flap}) exhibited $K_{\text{off-unbending}}$ that was \textasciitilde13-15 fold slower than that of SF-6,0 and \textasciitilde3-4 fold longer than $k_{\text{STO}}$ of the cognate substrate (Figure 2.4B and Figure 2.2C), yet FEN1 still bent these substrates multiple times without cleaving them (Figure 2.5C,D and Figure 2.5.1C,D). FEN1 therefore likely has intrinsic mechanisms that block the probable formation of catalytically competent active sites in noncognate substrates to inhibit off-target incision.

We reasoned that there are two possible mechanisms for controlling incision that can be tested experimentally. 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 $k_{\text{STO}}$ 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 $k_{\text{STO}}$ 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 $k_{\text{STO}}$ of FEN1 in
all tested noncognate substrates was similar and comparable to that in the cognate substrate (Figure 2.5A-D and Figure 2.2C). This indicates 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 on how it prevents cleavage in noncognate substrates.

Figure 2.5. Cleavage of noncognate substrates by FEN1. (A), (B), (C) and (D) Representative smFRET time traces showing the cleavage of FEN1 on SF6,0\textsubscript{Flap}, FEN1-R47A on NonEQ DF6,1\textsubscript{Flap}, FEN1 on DF-7,1\textsubscript{mismatch(1nt)}\textsubscript{Flap} and FEN1 on DF-6,2\textsubscript{Flap}, respectively, and the distribution of dwell times of the bent state prior to cleavage ($\tau_{avg}$). Cleavage occurs after a random number of missed cleavage opportunities from a bent conformer as illustrated in blue arrows in Figure 2.5A. $\tau_{avg}$ is calculated from $N = 4$ as in...
Figure 2.2C at a temporal resolution of 50 ms. The following figure supplements are available for Figure 2.5: Figure 2.5.1.
Figure 2.5.1 FEN1 cleavage of various noncognate substrates. (A) Representative TIRF-smFRET time traces showing the cleavage of SF6,0Flap in which FEN1 cleaves the DNA after multiple unsuccessful bending events. (B) Representative TIRF-smFRET time traces showing the cleavage of NonEQ DF6,1Flap by FEN1-R47A in which FEN1-R47A cleaves the DNA after multiple unsuccessful bending events. (C) Representative TIRF-smFRET time traces showing the cleavage of DF6,2Flap in which FEN1 cleaves the DNA after multiple unsuccessful bending events. (D) Representative TIRF-smFRET time traces showing the cleavage of DF7,1mismatch(1nt)-Flap in which FEN1 cleaves the DNA after multiple unsuccessful bending events. (E) Representative TIRF-smFRET time traces showing the cleavage of NonEQ DF6,1Flap by FEN1-Y40A in which FEN1-Y40A cleaves the DNA after multiple unsuccessful bending events. (F) Bar chart comparing $k_{\text{off-unbending}}$ of the DNA bending events that did not lead to 5’flap cleavage in presence of Mg$^{2+}$ in the cases of FEN1-WT on noncognate substrates and FEN1-Y40A and FEN1-R47A on NonEQ DF-6,1. The $k_{\text{off-unbending}}$ Values were similar to those reported under no cleavage condition in the presence of Ca$^{2+}$ that are presented in Figure 2.4B. The slight decrease in rates in the cases of FEN1-WT on SF-6,0 and FEN1-R47A on NonEQ DF-6,1 in the presence of Mg$^{2+}$ relative to Ca$^{2+}$ is a result of averaging due to the lower temporal resolution in TIRF-based smFRET (Mg$^{2+}$ case) versus immobilized confocal-based smFRET (Ca$^{2+}$ case).

2.3.6 Role of active site residues in positioning the 5’flap and junction

Given that FEN1 actively bends the DNA, it is unclear how FEN1 positions the junction before it locks down the DNA conformation. We therefore tested the role of key active-site residues. Mutating individual gateway residues Y40, K93, R100 or one of the metal-coordinating aspartic acidic residues (D181) to alanine markedly reduced the bent conformer’s stability (Figure 2.6A). These results revealed a direct role for active-site residues in stabilizing the bent DNA conformer. Interestingly, $k_{\text{on-bending}}$ was reduced up to 11-fold in gateway mutants, with R100 and to a lesser extent Y40 being critical residues (Figure 2.6B). This result is
surprising, given that these gateway residues appear to be disordered prior to DNA binding. Thus, active-site residues evidently contribute to active positioning of the junction while the 5'flap is being threaded through the unstructured cap-helical gateway. Moreover, D181A had only a minor effect on $k_{on\text{-}bending}$, implying that the metal ions do not interact with the phosphates during DNA bending and 5'flap threading. Importantly, this could provide a mechanism that protects the 5'flap from nonspecific cleavage.

FEN1-Y40A is the only active-site mutation that retains activity, albeit with a 100-fold reduced $k_{STO}$ (151). Y40 is proposed to play various roles in substrate positioning, including placing FEN1 at the junction, at the 5'flap in the cap-helical gateway and at the scissile phosphate in the active site (51, 151). We observed that FEN1-Y40A cleaves DF-6,1 after multiple cycles of DNA bending and unbending (Figure 2.6C, Figure 2.5.1E). This highlights the extreme selectivity of FEN1 to local variation in positioning of the junction and 5'flap for proper DNA lockdown. Unlike noncognate substrates or FEN1-R47A, the $k_{STO}$ in FEN1-Y40A increased by about three fold (Figure 2.6C); nonetheless, $k_{off\text{-}unbending}$ was still about two-fold slower than its $k_{STO}$ (Figure 2.6A,C). This indicates the presence of another step after proper DNA lockdown that involves active positioning of the scissile phosphate for incision.
Figure 2.6. Role of active-site residues in active positioning of the 5’flap and the junction. (A) Bar chart comparing $k_{\text{off}}$-unbending of WT-FEN1 and the FEN1 active-site mutants Y40A, K93A, R100A and D181A on NonEQ DF6,1Flap. $k_{\text{off}}$-unbending is calculated as described in 1G. (B) Bar chart comparing $k_{\text{on}}$-bending of wild-type FEN1 on EQ DF6,1Flap and the FEN1 mutants Y40A, K93A, R100A and D181A on NonEQ DF6,1Flap. $k_{\text{on}}$-bending is calculated as described in Figure 2.1G. (C) A representative smFRET time trace showing the cleavage of NonEQ DF6,1Flap by FEN1-Y40A after multiple trials of DNA bending (left panel) and a histogram showing the distribution of dwell times of the bent state prior to cleavage (right panel). $\tau_{\text{avg}}$ from $N = 4$ is calculated as in Figure 2.2C at a temporal resolution of 50 ms. (D) Model for control of catalytic selectivity by the DNA mutual induced-fit mechanism in FEN1. DNA sculpting: FEN1 actively bend 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 and drive 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, mispair junctions) and FEN1 mutants (R47A, K93A and R100A). DNA release or catalysis: the DNA will shift or unpair 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.

2.4 Discussion

Critical cellular processes such as DNA replication and repair are regulated by molecular properties encoded in interacting macromolecules whereby distinct dynamic conformations correspond to different functional outcomes. 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’. Findings so far have suggested that many processes are regulated by conformational change before substrate binding by the substrate’s selective binding to the active form of the enzyme, indicating that functional or conformational selection is in play (153, 154). 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?

With the FEN1 single-molecule results, a picture emerges of induced conformational changes to both substrate and 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 2.6D). 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 diffusion-limited on-rates in FEN1 and EXO1. We propose that different members of the 5’nuclease family 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.

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 a protein-conformational change (Figure 2.6D). Substrate distortion by ~90° and DNA-induced conformational
changes in proteins are features that extend beyond 5’nucleases (155-157). 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. Here, nuclease precision in replication comes from induced-fit that regulates the compatibility of the distorted DNA conformer with active-site assembly and its off rate to allow cleavage of cognate but not noncognate substrates (Figure 2.6D).

The diffusion-limited bending of cognate substrate by FEN1 and its cleavage from the first encounter represents a practically perfect precision reaction whose rate is 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. In cases in which an enzyme encounters noncognate substrates and cleaves them after multiple trials, results on substrate specificity from classical biochemical techniques that monitor product formation become misleading. 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 (51, 60, 63, 155-157). 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 (78, 158). 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.
2.5 Methods

Figure 2.7. Sequences of DNA constructs used in this study. Schematic of the substrates with internal-labeling and flap-labeling schemes used in this study including their sequences. A legend of the symbols used is shown.
2.5.1 Protein purification

Human FEN1 (amino acids: 2–380) was cloned into a pE-SumoPro expression vector (Lifesensors), which encodes an N-terminal His<sub>6</sub>-Tag followed by SUMO protein. This clone was used for recombinant expression FEN1 in BL21(DE3) E. coli strain. FEN1 was purified by multiple chromatographic steps involving two sequential Ni-NTA columns interspersed by sumo protease cleavage. Purity was further increased by running FEN1 on heparin column and Hioload superdex-75 gel filtration column. FEN1 was dialyzed against buffer (50 mM Tris-HCl pH 7.5, 50% glycerol, 300 mM NaCl, 10 mM BME), flash frozen and stored at -80°C.

Human EXO1 catalytic domain (amino acids: 2-352; referred to as EXO1 in this study) was cloned into a pE-SumoPro expression vector (Lifesensors). EXO1 was expressed and purified using a similar protocol to that described for FEN1. The purified EXO1 was dialyzed against buffer (50 mM Tris-HCl pH 7.5, 50% glycerol, 300 mM NaCl, 10mM BME), flash frozen and stored at -80°C.

2.5.2 DNA substrate preparation

DNA oligos, modified and unmodified, were purchased from Integrated DNA Technologies (IDT). All sequences used to make the substrate structures are shown in Figure 2.7. To prepare the DNA substrates, we mixed equimolar concentrations of oligos in TE-100 and annealed by heating to 95°C followed by slow cooling to room temperature. Properly annealed substrates were purified on
non-denaturing PAGE gel and extracted using the crush and soak method followed by ethanol precipitation.

### 2.5.3 TIRF-based smFRET

Experiments were performed in a microfluidic flow chamber made by sandwiching a polyethylene spacer (100 μm thick polyethylene double-sided spacer SA-S-1L from Grace Biolabs) between a quartz slide and a glass coverslip with inlet and outlet tubing channels. The glass coverslip was functionalized and passivated by a combination of 1:100 molar ratio of biotinylated polyethylene glycol (Biotin-PEG-SVA MW 5,000) and polyethylene glycol (mPEG-SVA MW 5000) (Laysan Bio Inc.). The flow chamber was incubated with NeutrAvidin (0.2 mg/ml) just prior to the experiment for 2 min and then washed excessively with reaction buffer (50 mM HEPES pH 7.5, 1 mM DTT, 5% glycerol and 0.1 mg/mL BSA, 100 mM KCl and 10 mM CaCl$_2$); in the cleavage assays, CaCl$_2$ was replaced with 10 mM MgCl$_2$. This was followed by incubation with biotin-labeled DNA substrate (100 pM) until a sufficient surface coverage of fluorescent-labeled substrate was achieved. This was followed by washing with imaging buffer (described below) containing the appropriate divalent metal ion, CaCl$_2$ for bending or MgCl$_2$ for bending and cleavage.

To minimize the effect of photobleaching and photo-blinking, we used an oxygen scavenging solution as described earlier (159) leading to enzymatic removal of oxygen by a 6 mM proto-catechuic acid (PCA) (Sigma-Aldrich, P5630)
and 60 nM protocatechuate-3,4-dioxygenase (PCD) system. Trolox (Sigma-Aldrich) was added at 2 mM concentration to reduce the photo-blinking by quenching the triplet state. The imaging buffer contained the reaction buffer and the aforementioned oxygen scavenging solution.

The experiments were performed on a custom-built TIRF-FRET setup as described earlier (141). For data analysis, the spatial mapping of the donor and acceptor emission channels was first calibrated using fluorescent beads that were imaged in TIRF mode. This generated a transformation matrix file, which was then used in the subsequent analysis of fluorescent molecules to map the donor and acceptor positions. The fluorescent molecules were registered as Gaussian point spread functions (PSF) around the brightest pixel in both channels and aligned with each other using the transformation matrix file. Donor and acceptor intensities were extracted by using software as described previously (160) and the apparent FRET efficiency was subsequently calculated. Any molecules with aberrant emission in brightness were excluded from further analysis. Histograms of FRET efficiencies were obtained from alternating the excitation of donor and acceptor 2c-ALEX as described previously (161). The vbFRET package implemented in Matlab (162) was used for dwell time analysis. The association ($\tau_{\text{bending}}$) and disassociation ($\tau_{\text{unbending}}$) dwell times were generated by idealizing and fitting the single molecule traces with two FRET states modeled by vbFRET (bent and unbent states). Histograms that were generated from dwell times in each state were fit with a single exponential decay function to generate $k_{\text{bending}} (1/\tau_{\text{unbending}})$ and $k_{\text{unbending}} (1/\tau_{\text{bending}})$, respectively.
2.5.4 Confocal-based smFRET

The experiments were performed on a custom-built confocal epifluorescence microscope setup (163). 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 532nm Cobalt 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 W cm\(^{-2}\) or 255 W cm\(^{-2}\) 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 recollimated 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 bandpass 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.

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 CaCl₂ 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.

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.

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/harrjp/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 (164). 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 FRET (i.e., bent) and high FRET (i.e., unbent) states. These lists were imported into OriginPro, histogrammed and fitted to a single exponential decay.

2.5.5 Single-molecule cleavage assays

Cleavage experiments were performed by TIRF-based smFRET at a temporal resolution of either 50 or 100 ms. The surface-immobilized substrate was pre-incubated in the flow chamber with imaging buffer containing 10 mM MgCl₂. In the case of cleavage of NonEQ DF6,1_Flap and EQ DF6,1_Flap by wild-type FEN1, image acquisition started before FEN1 reached the microfluidic chamber. In all other cleavage experiments, imaging started after the protein had reached the microfluidic chamber but before it reached the focal volume. This delay in acquisition was to reduce particle loss due to acceptor photobleaching since the waiting time before cleavage markedly increased under these suboptimal conditions. Selective loss of the donor signal was confirmed by direct excitation of the acceptor at the end of the cleavage experiment.
Due to short lag time before cleavage, manual counting of frames in the bent state was used to calculate the cleavage dwell time for each trace. All particles in the field of view were grouped into the following five different categories as shown in Figure 2.2.2A. a) Molecules with aberrant intensity that suffered from strong noise, photoblinking, step bleaching or deviation from the average intensity. These particles were excluded from further analysis. b) Molecules that had acceptor photobleaching before loss of the donor signal. These molecules do not influence dwell time analysis, which depends only on the donor signal. They were therefore excluded. c) Molecules that went to the low FRET bent state followed by a single-step loss in the donor signal. This formed the bulk of the traces of cognate substrates. d) Molecules that lost their donor signal without going into the low FRET bent state. This excluded minority population could result from donor photobleaching events and/or 5’ flap cleavage that occurs at faster rate than the acquisition time. e) Molecules that stayed in the unbent high-FRET state within our imaging time. These molecules were also excluded from the cleavage analysis as they exhibit no FEN1 binding. The following criteria were used to perform dwell time analysis for the selected particles. a) A minimum FRET change of 0.2 between unbent and bent frames was applied as a filtering criterion before selecting traces for dwell time analysis. b) Each selected trace was checked for anti-correlated behavior between the donor and acceptor upon change in the FRET efficiency. c) With noncognate substrates, the dwell time was calculated by counting the number of frames spent in the lower FRET state before donor signal was lost in the last bent step. MATLAB was used to calculate the mean of the
cleavage dwell time by fitting with gamma distribution function and the error in the mean by bootstrap. In the viscosity measurement, a falling ball viscometer (Gilmont) was used to calculate absolute viscosity. The density of the solution was measured from the mass of 1 ml of the same solution as used to calculate the viscosity.

2.5.6 Time-resolved bulk FRET measurements

The measurements were performed on a QuantaMaster 800 spectrofluorometer (Photon Technology International Inc.) coupled with a supercontinuum fiber laser source. The fluorophore lifetime was determined by the time-correlated single-photon counting (TCSPC) method. The excitation was carried out at 535 nm and emission was collected at 568 nm using 5 nm bandwidths for the excitation and emission. A longpass filter with 550 nm cut-on was placed on the emission side to prevent scattered light. The instrument response function (IRF) was determined using a colloidal silica suspension. The decay time traces were acquired at 10,000 counts. The measurements were performed at room temperature using the same 5’ flap constructs and buffer composition as in the smFRET experiments. The determination of the lifetimes was done using IRF reconvolution and a multi-exponential decay function incorporated in the FluoFit software package (PicoQuant). The donor lifetime curve was fitted to two-exponential decay. The best fit was selected based on reduced chi-square and randomness of the residuals.
The FRET efficiency, \( E_{FRET} \), refers to the conformational change resulting from the action of the enzyme on the substrate. \( E_{FRET} \) is calculated from the measured lifetime of the donor in the donor-only and donor-acceptor substrates at the respective enzyme concentration using the following Equation:

\[
E_{FRET} = 1 - \left( \frac{\tau_{DA-Enzyme}}{\tau_{D-Enzyme}} \right),
\]

where \( \tau_{D-Enzyme} \) and \( \tau_{DA-Enzyme} \) are the amplitude-weighted average lifetimes of the donor excited-state in the donor-only and donor-acceptor substrates in the presence of the enzyme, respectively.

The dissociation bending constant \( (K_{d-bending}) \) was calculated by fitting the data to a standard quadratic equation for simple bimolecular association (S+F \( \rightleftharpoons \) SF) under equilibrium conditions:

\[
E = E_0 + (E_f - E_0) \left( \frac{(S+F+K_{d-bending})^2 - \sqrt{(S+F+K_{d-bending})^2 - 4SF}}{2S} \right),
\]

where \( E \) is the FRET value at any protein concentration, \( E_0 \) and \( E_f \) are the initial and final FRET values, and S and F represent the substrate and FEN1 concentrations, respectively.

### 2.5.7 Surface plasmon resonance (SPR) binding

SPR binding was performed on a Biacore T100 (GE Healthcare Inc.). Biotinylated DNA substrates were immobilized on S-series streptavidin sensor chips in HBS-
EP buffer according to the manufacturer’s recommendations. The response unit (RU) of the immobilized substrate is stated in the figure legends. FEN1 was dialyzed overnight at 4˚C against the smFRET reaction buffer containing CaCl₂ (no oxygen-scavenging solution was added). Serial dilutions of FEN1 were made using the same reaction buffer. 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 protein sample injection at a flow rate of 20 µL/s for 90 s for DF-6,1 or 20 µL/s for 120 s for the other substrates. 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.). The maximum RUs reached at each FEN1 concentration were fitted using the steady-state affinity mode to obtain the equilibrium dissociation constant (\(K_{d\text{-binding}}\)) for each DNA substrate.

2.5.8 Steady-state bulk-cleavage assays

Reaction mixtures containing 0.5 nM Cy5-labeled NonEQ DF-6,1 dsDNA (Cy5 was placed 15-nt away from the nick junction on the downstream dsDNA on the 5’flap primer) in 1X reaction buffer (50 mM HEPES-KOH pH 7.5, 100 mM KCl, 0.1 mg/ml BSA, 5%(v/v) glycerol, 10 mM MgCl₂ and 1 mM DTT) were pre-incubated at 37˚C before the initiation of the cleavage reaction with the addition of varying concentrations of FEN1. Each reaction mixture was incubated further at 37˚C and equal aliquots were removed and quenched by equal volumes of 2X denaturing buffer (90% deionized formamide, 100 mM EDTA) at the following time intervals.
(0, 0.17, 0.5, 1, 1.5, 2, 5, 10 mins). These samples were run on 20% denaturing PAGE gels, which were imaged using a Typhoon TRIO Variable Mode Imager (GE Healthcare, Life Sciences). The product formation was quantified using the ImageJ gel analysis tool. For each FEN1 concentration, the concentration of the product formed was plotted against time to estimate the initial rate ($v_0$, nM.min$^{-1}$) by taking the slope of the linear part. These $v_0$ values were plotted against the FEN1 concentration and $K_m$ was determined by nonlinear least-squares fitting using a Michaelis-Menten model.

### 2.5.9 Molecular dynamics simulations

All simulations were performed with the AMBER 15.0 molecular dynamics package using the Parm14 force field with parmbsc0 nucleic acid modifications (165, 166). In total, four substrate DNA models (denoted NonEQ DF-6,1, SF-6,0, EQ DF-6,1 and nicked DNA) were generated. A dsDNA 47-mer with sequence d(5’-TGACCGTGTGTGACGGTGAGGAGGAAAGTTCCCCTCTCTACGGCAG-3’)•d'(5’-CTGCCGTAGGAGGAATTTCTCTCTGCTA(25)C(26)A(27)CGACCGTCAAACAACTGGGTCA-3’), identical to the one used in the smFRET experiments, was first constructed in a canonical B-DNA conformation with the UCSF CHIMERA program (167). The model NonEQ DF-6,1 was built by adding 5’ ssDNA flap (5’-TTTTTA-3’) and a single-nucleotide 3’ flap (G-3’) at the junction of the two DNA duplexes (between bases C26 and T25). The SF-6,0 and EQ DF-6,1 models were constructed by adding 5’ flap ssDNA (5’-TTTTTA-3’) at base C26 and 5’ flap ssDNA
(5’-TTTAC-3’) at base A27, respectively. Each system was solvated with TIP3P water (168) with a minimum distance of 15.0 Å from the DNA to the edge of the periodic simulation box. The systems were then neutralized by the addition of Na+ counterions. Additionally, 100 mM NaCl concentration was introduced to mimic physiological conditions. First, the water and ions were subjected to 3000 steps of steepest descent and 1500 steps of conjugate gradient minimization while restraining all DNA atoms with a force constant of 2 kcal/mol Å². All restraints were then released. The particle mesh Ewald (PME) method (169) was used to treat the long-range electrostatic interactions. The cutoff for non-bonded interactions was set to 10 Å. All bonds involving hydrogen atoms were constrained using the SHAKE algorithm. We imposed a 1-fs simulation time step during equilibration. The temperature of the simulated systems was then gradually increased to 300 K over 50 ps in the NVT ensemble. Subsequently, equilibration dynamics was carried out in the NPT ensemble (p = 1 atm and T=300 K) for an additional 50 ns. Then, 800-ns production runs were carried out for each simulation system in the isothermal isobaric ensemble (p = 1 atm and T = 300 K). We utilized hydrogen mass repartitioning (HMR) as a method to increase the simulation time step to 4 fs during the production runs (170). The substrate-bending angle for each system was defined and computed as described in Figure 2.1.3A. Data were analyzed with the CPPTRAJ code in AMBER15 (165) and TCL scripts in VMD (171).

The effective free-energy profile (potential of mean force; PMF) for bending the NonEQ DF-6,1 DNA substrate was estimated using the adaptive biasing force (ABF) method (172, 173) with the COLVARs module of NAMD 2.11 (174). ABF is
a widely used enhanced sampling approach, which computes average forces along a predefined reaction coordinate (RC) and then applies an adaptive biasing potential to flatten the underlying free energy landscape. As a result, all points along the RC can be sampled efficiently. In our case, the RC was defined as the bending angle between the two-dsDNA fragments of the substrate. The exact definitions of the two vectors and the fragments are shown in Figure 2.1.3A.

First, we carried out a 20-ns targeted molecular dynamics (TMD) simulation (175), with a force constant of 100 kcal/mol Å² applied to all nucleic acid heavy atoms. This simulation transformed the DNA from a straight conformation (bending angle of ~180°) to the bent conformation observed in the FEN1/DNA complex (bending angle of ~90°). The target configuration of the bent DNA was directly taken from the FEN1/DNA X-ray structure 3Q8L (51). In the ABF simulations, the reaction coordinate was segmented into nine discrete windows with a confining wall potential ($k = 50$ kcal/mol) placed at the boundaries. Snapshots collected from the TMD trajectory were used to seed the ABF windows. Each window was further subdivided into small 0.2° bins. Force averages were then accumulated into the bins and continuously updated in the course of the ABF simulation. Cancellation of the averaged forces through the gradual introduction of an adaptive bias led to enhanced sampling and overcoming of energy barriers along the RC. Since the instantaneous forces may fluctuate considerably, the application the adaptive bias was delayed until an adequate number of force samples was collected (2000 samples). PMF reconstruction was then accomplished by integration of the averaged forces from the bins.
This chapter contains data published in “Rashid, F. et al. Single-molecule FRET unveils induced-fit mechanism for substrate selectivity in flap endonuclease 1. Elife6, e21884 (2017)”.

Authors:
Fahad Rashid, Paul D. Harris#, Manal S. Zaher#, Mohamed A Sobhy, Luay I Joudeh, Chunli Yan, Hubert Piwonski, Susan E. Tsutakawa, Ivaylo Ivanov, John A. Tainer, Satoshi Habuchi, and Samir M. Hamdan

# These authors contributed equally to this work.

Contributions:
F.R. and M.A.S. established the single-molecule TIRF experiments. F.R. designed and performed the single-molecule TIRF experiments and purified proteins. P.D.H., H.P., F.R. and S.H. designed the confocal FRET experiments. P.D.H. performed the confocal FRET experiments. M.A.S. performed the time-resolved bulk FRET experiments. M.Z. performed the EXO1 experiments, bulk cleavage assays and supported F.R. in optimizing and analyzing the single molecule TIRF experiments and protein purification. L.I.J. performed the SPR experiments. C.L. and I.I. performed MD simulations. F.R, S.E.T., J.A.T, S.H. and S.M.H. designed the study. F.R. and S.M.H. supervised the study and wrote the manuscript. F.R, S.E.T., J.A.T, S.H. and S.M.H edited, revised and proofread the manuscript. All authors analyzed the data, discussed the results and commented on the manuscript.
Chapter 3

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

3.1 Abstract

RNA-DNA hybrid primers synthesized by low fidelity DNA polymerase α to initiate eukaryotic lagging strand synthesis must be removed efficiently during Okazaki fragment (OF) maturation to complete DNA replication. In this process, each OF primer is displaced and the resulting 5’-single-stranded flap is cleaved by structure-specific 5’-nucleases, mainly Flap Endonuclease 1 (FEN1), to generate a ligatable nick. At least two models have been proposed to describe primer removal, namely short- and long-flap pathways that involve FEN1 or FEN1 along with Replication Protein A (RPA) and Dna2 helicase/nuclease, respectively. We addressed the question of pathway choice by studying the kinetic mechanism of FEN1 action on short- and long-flap DNA substrates. Using single molecule FRET and rapid quench-flow bulk cleavage assays, we showed that unlike short-flap substrates, which are bound, bent and cleaved within the first encounter between FEN1 and DNA, long-flap substrates can escape cleavage even after DNA binding and bending. Notably, FEN1 can access both substrates in the presence of RPA, but bending and cleavage of long-flap DNA is specifically inhibited. We propose that FEN1 attempts to process both short and long flaps, but occasional missed cleavage of the latter allows RPA binding and triggers the long-flap OF maturation pathway.

### 3.2 Introduction

All processes involving breaks in DNA, including replication, repair, recombination, transcription and related cell cycle progression, have the potential to cause genome instability and lead to disease states (177). Hence, proteins involved in these multi-step DNA metabolic pathways often function in a coordinated manner to minimize and resolve DNA breaks, and maintain genome stability (178, 179). One such process is Okazaki fragment (OF) maturation, required to complete discontinuous lagging strand DNA replication (180). In eukaryotes, DNA polymerase α (Pol α) synthesizes an 8-12 nucleotide long initiator RNA (iRNA) and extends it with ~20-30 deoxyribonucleotides (nt) to create an iRNA-DNA primer that is handed off to polymerase δ (Pol δ), which in turn extends it to an ~200 nt OF. Approximately 10 million OFs are generated per cell cycle, all of which have to be linked efficiently during OF maturation to generate a continuous lagging strand. During this process the iRNA must be removed and the DNA fragments ligated; moreover, since Pol α lacks proofreading activity, misincorporated bases in the DNA primer need to be corrected. Thus, proper OF maturation is critical for maintaining the integrity of the genome (180, 181).

OF maturation begins when Pol δ reaches the end of a nascent OF and displaces the 5’ end of a previous fragment to generate a 5’ single-stranded iRNA/DNA flap of varying lengths protruding from a nick junction (Figure 3.1A and...
1B). Complementarity between the 5’ flap and the template strand could potentially equilibrate the nick junction to form a one-nucleotide 3’ flap (Figure 3.1B) (42, 135, 182). This double flap (DF) structure is recognized by structure-specific 5’ endonucleases and cleaved one nucleotide into the junction to form a nick that is sealed by DNA Ligase 1 (Figure 3.1B). Flap Endonuclease 1 (FEN1) is primarily responsible for the flap cleavage reaction, although other pathways have been proposed depending on the flap length (for reviews, refer to (46, 180, 181, 183)). The primary pathway involves FEN1 recognizing and cleaving short 5’ flaps while the secondary, less frequent pathway is invoked for processing longer flaps (Figure 3.1A). Pol δ has limited strand displacement activity, and can idle back and forth between polymerase and exonuclease activity on the primer, thus limiting flap length. Indeed, Garg et al. (145) and Stodola & Burgers (150) have suggested that the main product of FEN1 cleavage is a monoribonucleotide resulting from an active hand-off mechanism between Pol δ and FEN1 (nick translation), in which incremental strand displacement by Pol δ and cleavage by FEN1 are tightly coupled. Nevertheless, long flaps do occur as evident from both in vitro and in vivo studies (184-188). In the secondary pathway proposed for long flaps, the iRNA/DNA strand is bound by Replication Protein A (RPA), FEN1 cleavage is inhibited, and Dna2 helicase/nuclease is recruited to the site (189, 190). Dna2 displaces RPA (191) and progressively cleaves the flap to make it shorter. At this point, Dna2 dissociates due to its lower affinity for the short flap (185) or FEN1 disengages Dna2 (192, 193), and the DNA becomes a substrate for FEN1 again. Notably, despite the relative infrequency of long-flap processing events, Dna2 is
an essential enzyme in *Saccharomyces cerevisiae* while FEN1 is not (194, 195). This rather counterintuitive finding can be explained by the presence of backup endonucleases such as RNaseH2 and Exo1 that can substitute for FEN1 on short flaps (196, 197). On the other hand, there appears to be no equivalent substitute for Dna2 activity on long flaps, and since OF maturation defects have drastic consequences, Dna2 is essential.

Deletion mutations of Pif1 and the third subunit of Pol δ (Pol32) in *S. cerevisiae*, as well as post-translational modifications (especially acetylation) of human FEN1, Dna2 and Pol δ shed some light on how and why long flaps might occur during DNA replication. Pif1, a 5'-3' helicase, has been shown to augment the strand displacement capacity of *S. pombe* Pol δ by unwinding the previous OF end (198). A similar finding has been reported for *S. cerevisiae* Pol δ in the presence of Pol32 as well, compared to the 2-subunit Pol δ (199). Hence, both Pol32 and Pif1 can increase the likelihood of generating longer flaps. Indeed, deletion of either protein rescues the lethality caused by deletion of Dna2 in *S. cerevisiae*, although ΔPol32 has a lesser effect than ΔPif1 (199), suggesting a lower requirement for Dna2 activity when there are fewer long flaps. Acetylation of FEN1, Dna2 and Pol δ by histone acetyltransferase p300 also provides information on the generation and processing of long flaps. While complete acetylation of FEN1 reduces its cleavage activity by about 90% (200), the same modification stimulates by many fold Dna2 helicase and nuclease activities (201) and is suggested to enhance Pol δ strand displacement activity (181, 202). Thus, acetylation of these lagging strand DNA replication proteins favors the long-flap
pathway. One might consider the long-flap pathway an option for removing the full error-containing iRNA/DNA primer synthesized by low fidelity Pol α. On the other hand, processing a large fraction of OFs by this multi-protein, multi-step pathway that requires significant DNA re-synthesis at each OF by Pol δ might result in untenable delays in OF maturation. It is possible that the long-flap pathway is operational in select regions of the genome (reviewed by Balakrishnan and Bambara (181)). Since p300 acetylase exhibits a preference for chromatin regions with active gene expression, acetylation of replication/repair proteins may invoke the long-flap pathway at these sites (181).

Given that longer flaps have been detected in vivo and in vitro, we set out to understand how FEN1 acts on them, and whether FEN1 actions influence the choice between short- and long-flap pathways. Specifically, we asked whether FEN1 bends long flaps and catalyzes their cleavage in a similar manner as short flaps, and whether its mechanism is affected by RPA, which can compete with FEN1 for binding long flaps.

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 (Figure 3.1C) (51, 135, 140, 141). 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 (Figure 3.1C). Blocking 5’ flap threading with biotin/streptavidin at the end of the 5’ flap demonstrates that threading is a prerequisite for catalysis (64). Single molecule
experiments with such substrates show that FEN1 achieves a weak bent state that cannot support cleavage (135). Furthermore, the conformer of the blocked 5’ flap (141) and the degree of DNA bending (135) vary from that in the threaded complex. These single molecule findings are consistent with time-resolved crystallography data on human exonuclease 1 (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 (203). 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 (66, 203). Threading can also be viewed in the context of the regulation and 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.
Figure 3.1. FEN1 actions in Okazaki fragment maturation. (A) Short- and long-flap pathways for processing Okazaki fragments (OF). Left: schematic of the major short-flap pathway. A short 5’ flap (1-6nt) created by limited strand displacement activity of DNA
polymerase δ is recognized and cleaved by FEN1 to generate a nick that is sealed by DNA ligase 1. Right: schematic of the minor long-flap pathway. If a longer flap is formed, it is tightly bound by RPA, which inhibits FEN1 and necessitates Dna2 involvement. Dna2 displaces RPA and cleaves the long flap progressively until it is too short to maintain Dna2 binding and becomes a perfect substrate for FEN1 in the short-flap pathway. (B) Recognition and cleavage of short flaps by FEN1 in vivo involves flap equilibration. Strand complementarity between the downstream and upstream primers enables the nick junction to equilibrate. FEN1 binds this substrate and sculpts a 1 nt 3’ flap, resulting in a double flap (DF) structure that is incised 1 nt inside the junction to generate a sealable nick. (C) Crystal structure of FEN1 in complex with a 5’ flap substrate. FEN1 makes extensive interactions with the upstream and downstream duplex regions flanking the nick junction. The DNA is bent at the junction by ~100°, and the 5’ flap is threaded through a cap-helical gateway structure that oversees the active site. Binding of the 3’ flap transitions the gateway from a disordered to ordered form, providing a mechanism to control active site assembly (PDB accession code: 5UM9) (66).

Recently, using double flap substrates (DF) containing short 5’ flaps, we showed that the DNA exists in an extended conformer that is actively bent by FEN1 in diffusion-limited kinetics (135). This bending appears to induce ordering and assembly of the FEN1 active site associated with binding of the 1-nt 3’ flap in the 3’ flap-binding pocket (Figure 3.1C) (51, 55, 60, 135). This DNA-protein induced-fit mechanism leads to a catalytically competent state in the first encounter between FEN1 and a cognate DF substrate. However, in the case of a non-cognate substrate, the probability of forming the catalytically competent state is lower, which favors DNA dissociation prior to flap cleavage (135). In this study, we examined equilibrated DF substrates of different 5’ flap lengths by single molecule FRET (smFRET) and rapid quench-flow bulk cleavage assays to determine whether and how the mechanism of action of human FEN1 might vary with flap...
length. We found that the efficiency with which FEN1 binds and bends a longer flap containing DNA is not significantly affected, and its single turnover cleavage rate is slightly decreased. Importantly, as flap length increases, the probability of flap cleavage within the first DNA binding/bending encounter decreases. In other words, FEN1 has a higher chance of dissociating from the substrate without cleaving as flap length increases. These missed cleavage events create an opportunity for RPA to bind the flap and inhibit FEN1, thus necessitating involvement of Dna2 and initiation of the long-flap pathway.

3.3 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.

3.3.1 Effect of 5’ flap length on the DNA bending activity of FEN1
All DF substrates used in this study 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 single flap or double flap substrate (Figure 3.1B). Equilibrated (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 (135, 149). We used two DNA labeling schemes for the smFRET assays, namely "flap-labeling" and "internal-labeling". In the flap-labeling scheme (Figure 3.2A), 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 3.2A). These two labeling schemes provide complementary views of DNA binding, bending, cleavage and product release steps in the reaction, as described below.

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 the Materials and Methods section. The FRET peaks were centered around 0.30 for all unbent DF substrates and around 0.52 for the FEN1-bent substrates (Figure 3.2B-E). This result indicates that the flap length does not affect the basic backbone structure of the duplex portion of the substrate, whether free or bound to FEN1. 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 3.2B-E. The data yielded the following DNA bending dissociation constants ($K_d$-
bending): EQ DF-6,1\textsubscript{Internal} = 4.8±0.6 nM; EQ DF-29,1\textsubscript{Internal} = 3.3±0.4 nM; EQ DF-50,1\textsubscript{Internal} = 4.1±0.5 nM; and EQ DF-60,1\textsubscript{Internal} = 17.6±3.1 nM (Figure 3.2B-E). 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 (135). While EQ DF-6,1\textsubscript{Internal} and EQ DF-29,1\textsubscript{Internal} substrates transitioned between bent and unbent states in the presence of FEN1, EQ DF-50,1\textsubscript{Internal} and EQ DF-60,1\textsubscript{Internal} displayed higher stability in the FEN1-bound bent state, including a significant portion of bent particles that did not undergo any transition (Figure 3.2D and E). This higher stability observed with EQ DF-50,1\textsubscript{Internal} and EQ DF-60,1\textsubscript{Internal} implies that FEN1 has a lower rate of binding/bending given that the $K_{d,bending}$ does not change in case of EQ DF-50,1\textsubscript{Internal} and increases slightly in case of EQ DF-60,1\textsubscript{Internal}. EQ DF-6,1\textsubscript{Internal} and EQ DF-29,1\textsubscript{Internal} displayed diffusion-limited association rates; (1.57±0.47) x10\textsuperscript{8} and (1.33±0.01) x10\textsuperscript{8} M\textsuperscript{-1}s\textsuperscript{-1}, respectively, with similar dissociation rates.
Figure 3.2. Effect of 5' flap length on FEN1 DNA bending activity. (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. (B) FEN1 bending efficiency of internally-labeled equilibrated DF-6,1 (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\textsubscript{internal} 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 B\textsubscript{max}≤100 and yielded the DNA bending dissociation constant (K\textsubscript{d-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\textsubscript{bending} (s\textsuperscript{-1}) and k\textsubscript{unbending} (s\textsuperscript{-1}) versus FEN1 concentrations (nM). At each FEN1 concentration, time traces were idealized and fit by vbFRET to calculate the dwell times τ\textsubscript{bending} and τ\textsubscript{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\textsubscript{bending} (1/τ\textsubscript{bending}) and k\textsubscript{unbending} (1/τ\textsubscript{unbending}). The association rate constant (k\textsubscript{on-bending}) was calculated from the slope of the linear regression fit of k\textsubscript{bending} versus FEN1 concentration. The dissociation rate constant (k\textsubscript{off-unbending}) was calculated as the mean of k\textsubscript{unbending}. The error bars correspond to the standard error of the exponential fit of k\textsubscript{bending} and k\textsubscript{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\textsubscript{d-bending} = k\textsubscript{off-unbending}/k\textsubscript{on-bending}. (C) FEN1 bending efficiency of EQ DF-29,1\textsubscript{internal} as described in B. (D) FEN1 bending efficiency of EQ DF-50,1\textsubscript{internal} as described in B. (E) FEN1 bending efficiency of EQ DF-60,1\textsubscript{internal} as described in B.
3.3.2 A kinetic mechanism for short 5’ flap recognition and cleavage by FEN1

Our next goal was to assess the catalytic efficiency of FEN1 on substrates with varying flap lengths. We first investigated the kinetic mechanism of FEN1 on a short flap substrate to establish a basis for detailed comparison of the different DNAs. Addition of FEN1 to EQ DF-6,1\text{Flap} (Cy3 donor-labeled 5’ flap) in the presence of catalytically competent Mg\textsuperscript{2+} ions resulted in FEN1 bending the substrate, followed by a brief period with the complex in bent conformation and then flap cleavage (Figure 3.3A), as reported previously (135). We showed previously that FEN1 always cleaves the 5’ flap in DF-6,1 within the first DNA bending event, without a missed opportunity (135). The diffusion limited association rate (Figure 3.2B) (135) and the productive catalysis from the first bending event (135) demonstrate that FEN1 cleavage is not limited by substrate binding/bending. Since the steady state cleavage kinetics ($k_{\text{cat}}$) is significantly slower than the single turnover kinetics ($k_{\text{STO}}$) (44, 149), FEN1 is widely accepted to be product-inhibited. The reaction generates two products, a 5’ flap and a nicked DNA duplex, and based on 5’flap and nicked product competition studies (44, 135), only the latter was competitor. Therefore, it has been proposed that 5’ flap release is fast following cleavage whereas nicked duplex release limits the $k_{\text{cat}}$. The cleavage kinetics were determined by monitoring the bending step just before loss of the donor signal. Thus, the experiment with donor-labeled 5’ flap DNA reports the $k_{\text{STO}}$ rate, which includes DNA bending, protein ordering for active site assembly, chemistry, and flap release (135). By fitting the dwell time spent in the bent state ($t_{\text{bending-flap}}$) for each cleavage event with a gamma distribution, we
obtained an average $T_{\text{bending-flap}}$ of 155±30 ms and calculated the $k_{\text{STO}}$ ($1/T_{\text{bending-flap}}$) to be 6.5±1.2 s$^{-1}$ for the EQ DF-6,1$^{\text{Flap}}$ substrate (Figure 3.3A); this result is consistent with our previous single molecule measurements (135). Rapid quench-flow bulk cleavage experiments at 35:1 of FEN1:DNA ratio determined the $k_{\text{STO}}$ as 21±0.9 s$^{-1}$ (Figure 3.3B), which is consistent with previous reports (44, 149). We attributed the ~3 fold slower $k_{\text{STO}}$ in single molecule experiments to the difference in temperature; 22 °C in single molecule versus 37 °C in bulk.

Figure 3.3. Single molecule and bulk cleavage kinetics of FEN1 on short DF-substrates. (A) Flap-labeling smFRET cleavage assay. Top: schematic of the assay. EQ DF-6,1$^{\text{Flap}}$ is labeled as described in Figure 3.2A. In the presence of Mg$^{2+}$, FEN1 binds and
bends EQ DF-6,1\textsuperscript{Flap}, decreasing FRET from \(~0.7\) to \(~0.5\). Upon cleavage, the Cy3-labeled flap is released and the signal is lost. The assay follows the time spent by DNA in bent state before loss of signal. Bottom left: a representative single molecule time trace showing FEN1 bending and cleaving the substrate before the 5’ flap is released; the inset zooms in on the cleavage event preceded by a brief bending step showing clear anti-correlation between donor and acceptor intensities. Bottom right: distribution of the dwell times spent in bent state \(t_{\text{bending-flap}}\) for \(N=113\) cleavage events fitted to a gamma distribution. The average \(t_{\text{bending-flap}}\) is reported with the standard error of the mean. The cleavage reaction was performed at 50 ms temporal resolution. More representative traces are shown in Figure 3.1B. (B) 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 \(k_{\text{cat}}\) \((\text{slope}/[\text{FEN1}])\). Error bars correspond to the variation of the two replicates, and the error of the fit is reported. (C) Internal-labeling smFRET cleavage assay. Top: schematic of the assay. EQ DF-6,1\textsuperscript{internal} is labeled as described in Figure 3.2A. In the presence of Mg\textsuperscript{2+}, FEN1 binds and bends EQ DF-6,1\textsuperscript{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 (Figure 3.3.1A). 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\textsuperscript{internal}, bent EQ DF-6,1\textsuperscript{internal} and unbent nicked product, respectively. Bottom right: distribution of the dwell times spent in bent state \(t_{\text{bending-internal}}\) for \(N=64\) cleavage events fitted to a gamma distribution. The average \(t_{\text{bending-internal}}\) is reported with the standard error of the mean. The cleavage reaction was performed at 50 ms temporal resolution. More representative traces are shown in Figure 3.3.1C.

The flap-labeling scheme follows the cleavage reaction up to the 5’ flap release step but offers no information about the nicked DNA product; therefore, we
used a complementary assay based on the internal-labeling scheme to monitor steps subsequent to flap release. This assay detects the time spent by the substrate in bent state at a high FRET of $E \sim 0.52$ before the signal decreases to $E \sim 0.25$, which we interpret as the nicked product in an unbent state (Figure 3.3C). This interpretation is supported by the following information. The unbent EQ DF-6,1$_{\text{Internal}}$ substrate exhibits a difference of $E \sim 0.05$ from the unbent nicked product, as evident from the DNA-only histograms of DF-6,1 and nicked DNA (Figure 3.2B and Figure 3.3.1A). Thus, if DNA bending is not followed by 5’ flap cleavage, the FRET should change from 0.3 (unbent substrate) to 0.52 (bent substrate) and back to 0.3 (unbent substrate), whereas in the case of DNA bending followed by 5’ flap cleavage, the FRET should change from 0.3 (unbent substrate) to 0.52 (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 $K_{d\text{-bending}}$ of DF-6,1$_{\text{Internal}}$ (4.8 nM; Figure 3.2B) but well below $K_{d\text{-bending}}$ of the nicked product (lower estimate of 580 nM; Figure 3.3.1A). Furthermore, since FEN1 always cleaves DF-6,1 after the first encounter (135), 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 3.3.1. FEN1 bending of the nick DNA supports internal-labeling cleavage. (A) smFRET bending efficiency of FEN1 on nick internal. Left panel shows smFRET histograms of nick internal DNA-only (top) with a single peak centered around 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 around 0.41. Increasing the binding affinity of FEN1 to Nick_internal by lowering KCl from 100 mM to 40 mM, removes the averaging effect and clearly shows that the FRET state of Nick_internal upon binding FEN1 is similar to that of DF-6,1_Internal (E ~0.5; see Figure 3.4.1A). Right top: a representative single molecule time trace of Nick_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 a one-site binding model with Bmax≤100 yields $K_{d\text{-bending}}$. As described for Figure 3.7E, the $K_{d\text{-bending}}$ of FEN1 on Nick_internal is a lower estimate. This is due to the averaging effect 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 % bent substrate at each concentration and thus explicit determination of $K_{d\text{-bending}}$. (B) Representative single molecule time traces showing cleavage of EQ DF-6,1_flap using the flap-labeling smFRET cleavage assay as in Figure 3.3A. (C) Representative single molecule time traces showing cleavage of EQ DF-6,1_internal using the internal-labeling smFRET cleavage assay. FEN1 bends and cleaves the substrate almost always within the first bending event, as in Figure 3.3C.

The data show that the time spent by the internal-labeled substrate in bent state ($\tau_{\text{bending-internal}}$) is 270±70 ms (Figure 3.3C), while that of the flap-labeled substrate ($\tau_{\text{bending-flap}}$) is 155±30 ms (Figure 3.3A). The difference between $\tau_{\text{bending-flap}}$ and $\tau_{\text{bending-internal}}$, which reports the dwell time of the nicked product in bent state ($\tau_{\text{product-bent}}$), is only 115±75 ms. This value yields an apparent rate of 8.7±5.7 s$^{-1}$ (1/$\tau_{\text{product-bent}}$) 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$^{-1}$ (Figure 3.3A), and is much higher than the $k_{cat}$ measured in bulk under similar reaction conditions (1.4±0.1 s$^{-1}$; Figure 3.3B), 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 (44). 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 3.4A). 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 (Kd = 62 nM and 12 nM, respectively; Figure 3.4.1A and B). A control experiment with flap-labeled EQ DF-6,1 showed that kSTO is not affected by lower KCl (τbending-flap is 155±30 ms at 100 mM KCl and 180±40 ms at 40 mM KCl; Figure 3.3A and Figure 3.4.1C, respectively). Since FEN1 bends (Figure 3.2B Bottom Right) and cleaves EQ and Non-EQ substrates with similar kSTO (135, 149), Non-EQ DF-6,1Internal 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 τbending-internal from 270±70 ms at 100 mM KCl (Figure 3.3C) to 570±115 ms at 40 mM KCl (Figure 3.4B), indicating that the dwell time of bent product, τproduct-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 ($T_{\text{product-unbent}}$), before FEN1 dissociation followed by fast rebending re-established the bent state at $E \sim 0.52$ (Figure 3.4B). These results could be interpreted such that product release by FEN1 occurs in two steps: $T_{\text{product-bent}}$ wherein FEN1 briefly holds the product in bent state for $390 \pm 120$ ms after 5’ flap departure and $T_{\text{product-unbent}}$ wherein FEN1 remains bound to the unbent product for $2100 \pm 420$ ms before dissociating into solution. Therefore, the actual $T_{\text{release}}$ could be the sum of the two dwell times, which yields a $K_{\text{release}}$ of $0.40 \pm 0.07$ s$^{-1}$ for the nicked product. The resulting $k_{\text{cat}}$ of $0.37 \pm 0.06$ s$^{-1}$ ($1/(T_{\text{bending-flap}} + T_{\text{release}})$) is in line with rates determined by bulk experiments at 1:800 of FEN1:DNA (Figure 3.4.1D and Figure 3.3B at 40 and 100 mM KCl, respectively) and previous reports (44, 149). Taken together, these results suggest that FEN1 turnover is limited by release of the nicked DNA product, mainly from an unbent state.
Figure 3.4. Kinetic scheme for short 5’ flap recognition and cleavage by FEN1. (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 (135); 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 ($t_{\text{bending-flap}}$). (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 ($t_{\text{product-bent}}$) is the difference between the bent state dwell times in the internal-labeling ($t_{\text{bending-internal}}$) and flap-labeling ($t_{\text{bending-flap}}$) assays. (4) FEN1 dissociation: FEN1 remains bound to the unbent product (E ~0.25) for some time ($t_{\text{product-unbent}}$) before dissociating into solution. Thus, product release occurs in two steps and $t_{\text{release}}$ is the sum of $t_{\text{product-bent}}$ and $t_{\text{product-unbent}}$. (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 (Figure 3.4.1B). (B) smFRET cleavage of Non EQ DF-6,1Internal. Top: representative single molecule time trace showing cleavage of DF-6,1Internal and exhibiting the substrate and product dynamics described in (A). The FRET state and the substrate/product conformer in each step is illustrated. $t_{\text{bending-internal}}$ is highlighted in red and $t_{\text{product-unbent}}$ is highlighted in blue on the time trace. The distributions of $t_{\text{bending-internal}}$ (bottom left) and $t_{\text{product-unbent}}$ (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. More representative traces are shown in Figure 3.4.1E. (C) Bulk and single molecule PIFE experiments. Left: bulk time-resolved fluorescence lifetime measurements of Non EQ DF-6,1PIFE 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. As described in Methods, the lifetimes are determined using a 2-exponential decay fit, and show 35% fluorescence enhancement with FEN1. Right: representative time trace showing a smPIFE cleavage experiment with Non EQ DF-6,1PIFE. The substrate/product conformer in each state is illustrated. The time spent in the enhanced-fluorescence state $t_{\text{PIFE}}$ is highlighted in green. The distribution of $t_{\text{PIFE}}$ for N=77 cleavage events was fitted to gamma distribution, with the mean and standard error of the mean reported. The cleavage reaction
was performed at 100 ms temporal resolution. Additional traces are shown in Figure 3.4.1F.

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 3.4C) and observed 35% enhancement of Cy3 fluorescence with Non EQ DF-6,1PIFE in the presence of FEN1 (Figure 3.4C). 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 (τPIFE) 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, τPIFE 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 (135), and at 100 mM KCl and 250 nM FEN1 when no product rebending is observed, as shown in Figure 3.3C 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 τPIFE for N=77 cleavage events with a gamma
distribution yields a lengthy average $\tau_{\text{PIFE}}$ of 2210±500 ms (Figure 3.4C) 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.
Figure 3.4.1. Lower salt concentration stabilizes FEN1 bending of its products. (A) smFRET bending kinetics of Nick\textsubscript{Internal} by FEN1 at 40 mM KCl in the presence of Ca\textsuperscript{2+}. Top: a schematic showing the Nick\textsubscript{Internal} structure. Left: smFRET histograms of Nick\textsubscript{Internal} alone and upon addition of 500 nM FEN1. Unbent DNA-alone peak (shown in magenta) is centered around 0.24 and bent peak (in blue) is centered around 0.5. Right Top: a representative single molecule time trace of Nick\textsubscript{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 $K_{\text{d-bending}}$ are determined as described for Figure 3.2B. (B) smFRET bending kinetics of $\text{Product}_{\text{Internal}}$ by FEN1 in the presence of Ca$^{2+}$. Top: a schematic showing the $\text{Product}_{\text{Internal}}$ structure with a fixed non-equilibrated 1 nt 3’ flap mimicking the cleavage product of a Non EQ DF substrate. Left: bending kinetics of $\text{Product}_{\text{Internal}}$ at 40 mM KCl, with the middle panel showing a single molecule time trace of $\text{Product}_{\text{Internal}}$ upon addition of 20 nM FEN1 exhibiting dynamic behavior, and the bottom panel showing the association and dissociation rate constants as described in (A). Right: bending kinetics of $\text{Product}_{\text{Internal}}$ at 100 mM KCl. At higher salt, $\text{Product}_{\text{Internal}}$ displays faster transitions with a significantly elevated $k_{\text{off-unbending}}$, and hence $K_{\text{d-bending}}$. However, $k_{\text{on-bending}}$ is not influenced by salt concentration. (C) smFRET cleavage of EQ DF-6,1$\text{Flap}$ by FEN1 at 40 mM KCl. Distribution of the dwell times spent in the bent state ($t_{\text{bending-flap}}$) for N=72 cleavage events. The average $t_{\text{bending-flap}}$ at 40 mM KCl is comparable to that obtained at 100 mM KCl (Figure 3.3A). 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 3.3B, 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 3.3B). (E) Representative time traces of smFRET cleavage of Non EQ DF6,1$\text{Internal}$ by FEN1 at 40 mM KCl. The traces show similar behavior to that shown in Figure 3.4B. (F) Representative time traces of smPIFE cleavage of Non EQ DF-6,1$\text{PIFE}$ by FEN1 at 100 mM KCl. The traces show similar behavior to that shown in Figure 3.4C.

Finally, based on our interpretation of the smFRET data from experiments with both flap- and internal-labeled DNAs, we introduced a comprehensive reaction mechanism of FEN1 action on short flaps (Figure 3.5). To summarize, FEN1 binds and actively bends DF-6,1 by diffusion-limited kinetics, cleaves the 5’ flap rapidly after protein ordering and active site assembly (135), followed by slow product release that occurs in two steps, highlighting the relatively high affinity and stability of the FEN1-nicked product complex.
**Figure 3.5. Kinetic scheme for the 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. $k_{\text{bending}}$ and product release rates are shown for Non EQ DF-6,1 $k_{\text{STO}}$ at 40 mM KCl is shown for EQ DF-6,1 $k_{\text{STO}}$ in Figure 3.4.1C; $k_{\text{STO}}$ is independent of KCl concentration and it has been shown to be similar for both EQ DF-6,1 $k_{\text{STO}}$ and Non EQ DF-6,1 $k_{\text{STO}}$ (135). The substrate and product on/off rates are determined in presence of Ca$^{2+}$; substrate EQ DF-6,1 at 100 mM KCl (Figure 3.2B) and product at 40 mM KCl (Figure 3.4.1B). Based on data with the product (Figure 3.4.1B), we do not anticipate a significant change in $k_{\text{on-bending}}$ of the substrate with lower KCl concentration.
3.3.3 Effect of 5’ flap length on the FEN1 kinetic mechanism

After establishing the FEN1 kinetic mechanism on a short 6 nt flap substrate, we tested its activity on longer flaps (29, 50 and 60 nt). In this case, we performed experiments mainly with internal-labeled substrates since the higher flexibility of long flaps increased the noise level with flap-labeled substrates, complicating identification of real bending and cleavage events (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 internal; Figure 3.2C-E), but lower than the $K_{d-bending}$ of nicked DNA (lower estimate of 580 nM; Figure 3.3.1A). As with DF-6,1 internal, 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 $\tau_{bending-internal}$ for each DF substrate as shown in Figure 3.6A-C. Interestingly, there were no significant differences in $\tau_{bending-internal}$ 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) (Figure 3.3C). Given that $k_{release}$ for the common nicked product is not expected to vary with 5’ flap length, similar $\tau_{bending-internal}$ values mean that $k_{STO}$ is similar across the flap lengths as well.

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 3.6D). 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 3.6.1E, 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 3.6E). 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 ($\tau_{\text{bending-missed}}$) showed that $\tau_{\text{bending-missed}}$ 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 $\tau_{\text{bending-missed}}$ (~270 ms) was in the same range as $\tau_{\text{bending-internal}}$ (Figure 3.6E and Figure 3.6A-C, 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\(^{-1}\) (135) than the ~3.7 s\(^{-1}\) rate \((1/\tau_{\text{bending-missed}})\) measured here (Figure 3.6E). It is also possible that these events result from trapping of partially threaded complexes in a fashion similar to that observed with Exo1 (203).
Figure 3.6. Effect of 5’ flap length on FEN1 cleavage activity. (A) smFRET cleavage of EQ DF-29,1\textsubscript{internal}. Top: distribution of the dwell times spent in bent state (\(t_{\text{bending-internal}}\)) for N=98 cleavage events fitted to a gamma distribution. Average \(t_{\text{bending-internal}}\) 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-29,1\textsubscript{internal}, bent DF-29,1\textsubscript{internal} and unbent nicked product. The cleavage reaction was performed at 50 ms temporal resolution. More representative traces are shown in Figure 3.6.1A. (B) smFRET cleavage of EQ DF-50,1\textsubscript{internal} as described in (A). More representative traces are shown in Figure 3.6.1B. (C) smFRET cleavage of EQ DF-60,1\textsubscript{internal} as described in (A). More representative traces are shown in Figure 3.6.1C. (D) FEN1 misses cleavage on longer DF substrates. FEN1 almost always cleaves a short-flap substrate (EQ DF-6,1\textsubscript{internal}) in the first bending event (135). 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 vbFRET-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. More representative traces are shown in Figure 3.6.1D. (E) 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 ($\tau_{\text{bending-missed}}$) by the different DF substrates. The reported N is the number of missed events, not number of particles, accounting for particles with multiple missed events. (F) 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 Figure 3.3B.

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 3.6.1F). 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 (135).
Figure 3.6.1. Further characterization of missed cleavage events. (A) Representative time traces showing cleavage of EQ DF-29,1_{Internal} wherein FEN1 cleaves the substrate.
within the first bending event as in Figure 3.6A. (B) Representative time traces showing cleavage of EQ DF-50,1\textsubscript{Internal} wherein FEN1 cleaves the substrate within the first bending event as in Figure 3.6B. (C) Representative time traces showing cleavage of EQ DF-60,1\textsubscript{Internal} wherein FEN1 cleaves the substrate within the first bending event as in Figure 3.6C. (D) Representative time traces showing cleavage of long DF substrates with missed opportunities as shown in Figure 3.6D. (E) Assignment of a bending event as missed cleavage or actual cleavage. Time traces that exhibit multiple bending events in EQ-DF-50,1\textsubscript{Internal} 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. (F) 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.

Finally, we conducted complementary rapid quench-flow bulk cleavage experiments with the same EQ DF substrates to validate the finding that $k_{\text{STO}}$ and $k_{\text{cat}}$ do not vary significantly with increasing 5' flap length. As described previously for DF-6,1, $k_{\text{STO}}$ was determined at 35:1 ratio of FEN1:DNA, while $k_{\text{cat}}$ was determined at 1:800 ratio of FEN1:DNA. The $k_{\text{STO}}$ was slightly higher for 30 nt (32.5±1.4 s\textsuperscript{-1}; Figure 3.6F) compared to 6 nt (21±0.9 s\textsuperscript{-1}; Figure 3.3B) and slightly lower for 50 nt (12.6±0.7 s\textsuperscript{-1}) and 60 nt flaps (9.9±0.5 s\textsuperscript{-1}). The $k_{\text{cat}}$ also followed the same trend as $k_{\text{STO}}$, showing a small change with flap length relative to DF-6,1 (2-fold maximum; Figure 3.6F), 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 $k_{STO}$) (149). 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., (149) 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 $T_m = 51.9^\circ$C). The $k_{STO}$ of FEN1 is significantly reduced on DF substrates containing 5’ flap hairpins (44, 182); 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. 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 (Figure 3.5). 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.

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. 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 (Figure 3.6.2). 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 single flap RNA substrates (204). 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.

Figure 3.6.2. Kinetic parameters of FEN1 substrate bending and catalytic efficiency on DNA versus RNA EQ DF substrates.

<table>
<thead>
<tr>
<th></th>
<th>EQ DF-6,1</th>
<th></th>
<th>EQ DF-29,1</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DNA</td>
<td>RNA</td>
<td>DNA</td>
<td>RNA</td>
</tr>
<tr>
<td>(K_{d\text{-bending}}) (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_{\text{on}}) (x10^8 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_{\text{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_{\text{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>
3.3.4 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 (181); 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 3.1), is lethal in S. cerevisiae whereas deletion of the FEN1 orthologue (Rad27) is tolerated, albeit with a severe mutator phenotype (194, 195). 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 3.7A). This result can be explained by the dual single stranded DNA-binding modes of RPA: a weak, transient 8-nucleotide mode ($K_{d-binding}$ ~50 nM) and a stronger, more stable 30-nucleotide mode ($K_{d-binding}$ ~0.5 nM) (205), 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\textsubscript{Internal} and DF-29,1\textsubscript{Internal}) in the presence of Ca\textsuperscript{2+}. 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 3.7B); 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 \(K_{d(binding)}\) of both binding modes. As shown in Figure 3.7C, 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 3.7C) as with FEN1 alone (Figure 3.2B). 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 \(K_{d(bending)}\) constants (4.8±0.6 nM and 8.6±1.2 nM in the absence and presence of RPA, respectively; Figure 3.7E). 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 3.7D) when compared with FEN1 alone (Figure 3.2C); 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 $K_{d-bending}$ for FEN1 in the presence of RPA, nevertheless, a lower estimate of 318.0±30.7 nM (Figure 3.7E) was ~100 fold higher than that in the absence of RPA (3.3±0.4 nM; Figure 3.2C). 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 3.7D). 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 (135).
Figure 3.7. 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).
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\textsubscript{Internal} and EQ DF-29,1\textsubscript{Internal}. Left: EQ DF-6,1\textsubscript{Internal} DNA-only histogram (top) with FRET centered around 0.33, and upon addition of 200 nM RPA (bottom) with FRET centered around 0.31. Right: corresponding histograms for EQ DF-29,1\textsubscript{Internal} substrate. (C) smFRET bending of EQ DF-6,1\textsubscript{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 3.2B). 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\textsubscript{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 3.2C). 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). (F) smFRET cleavage in the presence of RPA. 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 $t_{bending-flap}$ for N=112 cleavage events in the presence of RPA was fitted with a gamma distribution and the mean with standard error is reported. More representative traces are shown in Figure 3.7.1A.
Finally, the effect of RPA on FEN1 cleavage activity was measured under single turnover conditions using flap-labeled DF-6,1Flap and DF-29,1Flap 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 (τbending-flap = 190±40 ms; Figure 3.7F) or absence of RPA (τbending-flap = 155±30 ms; Figure 3.3A). We do not anticipate any effect of RPA on kcat with DF-6,1 since Kd-bending of the product was not affected by RPA (Figure 3.7.1B). 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 3.7D), 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 (206). 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 (207). 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.
Figure 3.7.1. RPA does not affect FEN1 product release. (A) Representative time traces showing cleavage of EQ DF-6,1\textsubscript{Flap} in the presence of 100 nM RPA; FEN1 cleaves the substrate within first bending event as in Figure 3.7F. (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\textsubscript{Internal} in Figure 3.2B. RPA does not display any significant effect on FEN1 bending efficiency with Product\textsubscript{Internal}, and hence is unlikely to affect its product release kinetics.

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.
3.4 Discussion

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 (66, 208, 209). 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 (25, 74, 179, 210).

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 (145, 150). 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 (184); deletion of FEN1 in *S. cerevisiae* results in duplications that indicate flap lengths as long as 100 nt as well (186, 211). In this case, the evidence indicates that RPA and Dna2 helicase/nuclease are involved in flap removal in addition to FEN1 (189, 190, 212, 213). Processing of long flaps has the benefit of removing both the iRNA 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 (206, 214). 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 (181, 200, 201, 215).

In this study, we examined FEN1 activity on short and long flaps in order to elucidate the events leading to OF maturation by the short- versus long-flap pathways. Interestingly, we 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 (Figure 3.6F and 3.3B, respectively), 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 (Figure 3.2B-E). 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 DNA, not the flap (51, 66). In smFRET single turnover cleavage experiments, again we found little difference in the rates of multiple steps in the reaction with increasing flap length (Figure 3.3C and 3.6A-C). 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 (Figure 3.6E). 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 (Figure 3.3C and 3.6A-C versus 6E). 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.

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 (Figure 3.7 and Figure 3.7.1). 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 (Figure 3.7D). 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 helicase/nuclease. 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 an RPA-bound long-flap substrate
could facilitate substrate transfer to FEN1 as Dna2 shortens the 5’ flap and
displaces RPA. Moreover, our finding that the nicked product is released in two
steps indicates a product hand-off mechanism. Since FEN1 holds on to the DNA
in unbent form, we speculate that perhaps it contacts part of the duplex while
allowing DNA ligase to access the newly formed nick and complete OF maturation.

To summarize, in this study, we defined the kinetic mechanism of FEN1 in
greater detail, especially regarding recognition and cleavage of long 5’ flap-
containing DNA substrates. 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 structure/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 of this system.

3.5 Materials and Methods

3.5.1 Proteins expression and purification
Human FEN1 (amino acids: 2–380) was cloned and expressed as described
previously for single molecule (135) and ensemble experiments (216). Human
RPA clone (pET11d-tRPA) was a generous gift of Professor Marc S. Wold, and
the protein was purified as described (217). Briefly, the plasmid was expressed in
BL21 (DE3) Escherichia coli cells and RPA was purified on a HiTrap blue HP
column, followed by desalting and concentration on a hydroxyapatite column, and
finally ion exchange chromatography on a MonoQ column. Protein fractions were pooled, flash frozen and stored at -80°C.

### 3.5.2 DNA substrates
DNA oligos were synthesized and HPLC purified by Integrated DNA technologies (IDT). Substrates were annealed by mixing template: 5’ flap: 3’ flap strands in 1: 3: 5 molar ratio 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 MgCl₂ was added to the annealing buffer. Substrates were purified to >80% purity by non-denaturing 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. A list of these substrates is illustrated in Figure 3.8A.
Figure 3.8. Substrates used in this study. Schematic illustrating the sequences as well as modifications of the substrates used in this study. A legend of the symbols used is included. (A) Substrates used in single molecule bending and cleavage assays. (B) Substrates used in bulk stopped-flow single turnover and steady state cleavage assays. (C) Substrates used to assess RPA effect on FEN1 cleavage efficiency in bulk assays.

3.5.3 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 one inlet and one outlet tubing attached for exchange of buffer.

For immobilization of the double-labeled biotinlated 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 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 (159), which relies on the enzymatic elimination of oxygen through a 6 mM protocatechuic 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.

All single molecule experiments were performed using a custom-built TIRF-FRET setup (147). 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 excitation through green laser (CW) 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 (160) 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 the apparent FRET.

### 3.5.3.1 Ca2+- based DNA bending assays

These experiments were performed under the conditions described above in the presence of CaCl₂. 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 (161). 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 $K_{d-bending}$ was estimated using the constraint $B_{max} \leq 100$. For dwell time analysis of the time traces, CW movies were analyzed using the vbFRET package implemented in Matlab (162), where the time traces were idealized and fitted to two FRET states (bent and unbent). vbFRET is based on Hidden Markov Modeling where the most likely FRET values are determined based on probability alone. The dwell times spent in each state were plotted in histograms and fitted with a single exponential decay yielding $k_{bending}$ ($1/T_{bending}$) and $k_{unbending}$ ($1/T_{unbending}$). $k_{bending}$ was plotted versus different FEN1 concentrations to obtain $k_{on-bending}$ from the slope of the linear fit. $k_{off-bending}$ was obtained from the average of $k_{unbending}$ at different FEN1 concentrations.

3.5.3.2 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 MgCl$_2$. For experiments with RPA, the DNA-immobilized surface was pre-incubated with sufficient RPA before co-injection of RPA and FEN1; 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, we followed the protocol described in our previous study (135). Briefly, a cleavage event was identified as a transition from the bent to unbent 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. For the internal-labeling scheme, a cleavage event was identified by following 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 CaCl\(_2\) experiments. The dwell times 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 by gamma distribution 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 rather than the unbent nicked duplex 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.

### 3.5.4 PIFE bulk and single molecule 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-$6,1_{PIFE}$ in the absence and presence of 1 μM FEN1 were performed at room temperature in time-correlated single-photon counting (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-$6,1_{PIFE}$ 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.

### 3.5.5 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 ratio, 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$_2$). A list of the substrates is shown in Figure 3.8B.

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$_2$, 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 ($\lambda_{Ex} = 488$ 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 hFEN1 was mixed with 800 nM DNA substrate (10x$K_M$; (44) 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]).

### 3.5.6 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 (substrates shown in Figure 3.8C)
were used to assess FEN1 cleavage efficiency in the absence and presence of RPA. For direct comparison, the assay was performed as described previously for yeast proteins (206). 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% 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).

**Authors:**
Manal S. Zaher, Fahad Rashid#, Bo Song#, Luay I. Joudeh, Mohamed A. Sobhy, Muhammad Tehseen, Manju M. Hingorani and Samir M. Hamdan.

#These authors contributed equally to the paper as second authors.

**Contributions:**
M.S.Z., F.R. and S.M.H. conceived and designed the experiments. M.S.Z and F.R. established single molecule assays. M.S.Z performed single molecule experiments and bulk cleavage assays in the presence of RPA. F.R. purified human FEN1. B.S., M.S.Z., M.M.H. and S.M.H. designed ensemble kinetic assays. B.S. performed ensemble kinetic assays. L.I.J, M.A.S, M.T. helped with some experiments. M.S.Z. and S.M.H. wrote the manuscript. M.S.Z., M.M.H. and S.M.H. edited, revised and proofread the manuscript.
Chapter 4

4. Initial state of the DNA-dye complex sets the stage for on demand fluorescence modulation upon protein binding*

4.1 Abstract

Protein-induced fluorescence enhancement (PIFE) is a popular tool for characterizing protein-DNA interactions. PIFE has been explained by an increase in local viscosity, due to the presence of the protein residues. This explanation denies the opposite effect of fluorescence quenching. This work offers a new perspective for understanding PIFE’s mechanism and reports the first observation of a novel phenomenon that we name ‘protein-induced fluorescence quenching’ (PIFQ), which exhibits an opposite effect to PIFE. A detailed characterization of these two fluorescence modulations reveals that the initial fluorescence state of the labeled mediator (DNA) determines whether this mediator-conjugated dye undergoes PIFE or PIFQ, upon protein binding. This key role of the mediator DNA, characterized for the very first time in this study, provides guide to obtain either PIFQ or PIFE, ‘on-demand’. This makes the arbitrary nature of the current experimental design obsolete, allowing for proper integration of both PIFE and PIFQ with existing bulk and single molecule fluorescence techniques used to investigate the binding mechanisms of proteins to nucleic acids.

* This chapter is to be published in: Rashid F, Raducanu VS, Zaher MS, Tehseen M, Habuchi S, Hamdan SM. (2018) Initial state of the DNA-dye complex sets the stage for on demand fluorescence modulation upon protein binding (Under Review in Nature Communications).
4.2 Introduction

In recent years, protein-induced fluorescence enhancement (PIFE) has gained popularity, as a stand-alone or complementary single molecule (SM) imaging assay. This popularity mainly stems from its simple labeling requirements involving the use of a single fluorophore (218) and its ability to measure changes in a distance-dependent manner that supersedes the FRET range (125). PIFE has been used to study a variety of protein-DNA interactions including, but not limited to, polymerase binding, translocation of DNA motor proteins, and catalysis in structure-specific endonuclease (125, 126, 176, 218-227). It has been also used to determine the directionality of the filament-forming proteins RAD51, RecA, and DnaA (125, 228, 229) that were previously inaccessible by smFRET. PIFE has also been used to complement other SM imaging techniques, such as smFRET or DNA flow stretching, to provide an extra observable that opens a new dimension of information (140, 224, 230-232).

As a photophysical phenomenon, PIFE occurs in environmentally sensitive fluorophores of the cyanine family such as Cy3, Cy5, Alexa555 and Alexa647 (Figure 4.1.1). PIFE experiments with cyanine dyes coupled to DNA or to protein illustrate that neither the dye coupling nor the presence of the interacting molecule(s) induce a bathochromic or hypsochromic shift (230). This spectral stability implies that the energy levels remain unperturbed, and therefore the explanation of the environmental sensitivity of cyanine dyes is ascribed to the cis-trans photoisomerization of their two indole cyclic groups around the polymethylene
bond (Figure 4.1B and Figure 4.1.1) (121, 233). Even though both isomers can absorb light upon excitation, only the trans S\textsuperscript{1}*-to-trans S\textsubscript{0} transition results in fluorescence emission. However, several lines of evidence support that the ground state is primarily trans, with some studies even considering it to be all trans (234, 235). First, a conversion pathway from cis to trans, but not the opposite, occurs in the ground state (236). Second, the trans isomer has a larger absorption cross-section, reflecting its higher thermodynamic stability, higher symmetry and larger absorption transition dipole moment (233, 237). Finally, the energy barrier between the cis and the trans isomers in the ground state is relatively high (238). Collectively, the stability of the trans ground state indicates that the role of photoisomerization in the environmental sensitivity of cyanine dyes must act primarily on the excited state.

It is well-accepted that the isomerization in the excited state goes through a metastable 90°-twisted intermediate state during the rotation (Figure 4.1A) (121, 230, 238-240). Considerable research emphasizes some important characteristics of this 90°-twisted intermediate state that sheds light on how it comes into play within the context of photoisomerization. First, it is very tightly coupled to the ground state, and therefore leads to a fast de-excitation (121, 230, 238-240). Second, this de-excitation is non-radiative due to the non-planar structure of the 90°-twisted excited state (237). Third, the energy barrier between the 90°-twisted intermediate state and the excited trans conformer is relatively low (238). Finally, the excited cis isomer, itself, can isomerize to the 90°-twisted intermediate state (233), and it mostly de-excites non-radiatively with negligible quantum yield (121,
Therefore, it has been proposed that the rate of photoisomerization from trans $S_1^*$ to twisted $S_1^*$ (Figure 4.1A) dictates the fluorescence properties of the dye in a particular configuration or environment. This rate is affected by changes in the local environment of the fluorophore and is in competition with the rate of fluorescent de-excitation. This leaves open the possibility of changing these rates by affecting the local environment of the fluorophore, by changing the viscosity for example. Within this paradigm, PIFE has been explained by an analogy to an increase in the local viscosity due to the presence of the protein residues decreasing the rate of photoisomerization from trans $S_1^*$ to twisted $S_1^*$. 

**Legend**

- Absorption
- Internal conversion
- Vibrational relaxation
- Fluorescence
- Non-radiative relaxation (incl. collisional quenching)
- Intersystem crossing
- Phosphorescence
- Delayed fluorescence
- Energy transfer to oxygen molecule in triplet state
- Cis-trans isomerization
- Ground-state complex formation (static quenching)
- Exciplex formation
- Spin and vibrational states
* Denotes an excited state
Figure 4.1. Energy landscape of a cyanine dye. (A) Schematic representation of potential energy landscape for $S_0$ (ground singlet) and $S_1^*$ (first excited singlet) states as a function of single dihedral angle of the Cy3 polymethine bond, superposed on a Jablonski diagram exhibiting all theoretical transitions and effects. Some of the illustrated transitions are present only for showing a complete Jablonki diagram, even though for Cy3 they were not observed experimentally, while other transitions were not experimentally tested at all. (B) Schematic representation of the Cy3 dye molecule and one of its dihedral angles, in fluorescent trans and dark cis configurations.
Figure 4.1.1. Structures of the different fluorophores used in this study: (A) 3'-iCy3 (B) 5'-iCy3 (C) internal iCy3 (D) Cy3-NHS (E) Cy3B (F) iCy5 (G) Alexa Fluor 555 (H) Alexa Fluor 647 and (I) DyLight 633.

Existing models rely on the idea that proteins influence the rate of photoisomerization from trans S1* to twisted S1* through the concepts of ‘steric hindrance/restriction’ (244) and ‘specific contact with certain residues’ (121, 233).
However, these models would only predict fluorescence enhancement and deny the existence of an opposite phenomenon, since the presence of the protein can only increase the local viscosity. Moreover, this enhancement necessitates immediate local contact between the protein residues and the dye. If the dye is in contact with the same protein residues, these models assume that the change in fluorescence will not depend on the initial fluorescence of the DNA-Dye (DNA structure and sequence used). Collectively, these observations suggest that in PIFE, the protein is primarily responsible for the fluorescence enhancement and the role of the DNA, itself, is trivial.

To expand the current models of PIFE, one would have to consider the mediator that propagates the interactions and to which the dye is attached. It has been shown that coupling fluorophores to DNA or to a protein, via at least one linker, enhances its fluorescence, mostly by a partial rigidization of the fluorophore (235, 237, 239, 245-248). These studies also defined some of the parameters contributing to the fluorescence properties of DNA-coupled Cy-dyes, including DNA sequence, dye position and the overall DNA structure. With that in mind, an extra dimension has to be considered in PIFE experiments, as the DNA itself can restrict the fluorophore's photoisomerization through various interactions, therefore dictating the initial state of fluorescence. However, current characterizations of PIFE and DNA-coupled Cy-dyes neither offer a systematic or quantitative correlation between the initial fluorescence of the DNA-Dye system and the final fluorescence, after binding of the protein, nor they provide detailed guidelines to design and control PIFE.
To characterize the relationship between the initial and final fluorescence states of DNA-coupled Cy-dyes upon protein interaction, we investigated a new phenomenon that we referred to as ‘protein-induced fluorescence quenching’ (PIFQ), simply because it seemed to show the opposite effect of PIFE, upon binding of the proteins to certain DNA-Dye complexes. It is worth mentioning that several mechanisms are known to quench fluorescence, such as static quenching, collisional quenching, photoinduced electron transfer (PET), iron-sulfur clusters, and FRET (249-251). However, none of these mechanisms can explain the PIFQ effect observed here. Therefore, we present a new perspective to the understanding of PIFE and PIFQ by investigating the change in fluorescence induced by disrupting the initial DNA-Dye structure upon protein binding. In this mechanism, the initial DNA-Dye structure is the determinant factor for the initial state of fluorescence, which in turn sets the stage for the fluorophore to experience either PIFE or PIFQ in the presence of an external modulator (e.g. bound protein, annealed complementary DNA strand, specific DNA structure induced by metal ions, among others). However, the final state of fluorescence is dictated by the structure and interactions of the dye within the final complex. Our perspective does not exclude the possibility of ‘steric hindrance/restriction’ and/or ‘specific contact with certain residues’ due to the presence of the protein contributing to the final state, but rather adds one additional variable, that is, the initial state of fluorescence of the DNA-Dye system. Additionally, we show that, similarly to PIFE, PIFQ is suitable in both ensemble and single molecule assays.
4.3 Results

4.3.1 A novel DNA-mediated bidirectional fluorescence modulation observed through FEN1/DF system

We recently used smFRET to characterize the mechanism of double flap (DF) substrate recognition by DNA replication and repair Flap Endonuclease 1 (FEN1) (252). Briefly, FEN1 cleaves excess 5’flaps from the DNA to restore its heritable duplex form. In this process, FEN1 threads the 5’flap into a capped helical gateway structure where it forms specific contacts that promote catalysis (252, 253). In-vitro, DF is made by annealing three different oligonucleotides to create a nick bearing a 5’flap of single-stranded DNA (ssDNA) of variable lengths and a 1 nucleotide (nt) 3’flap; these substrates are named ‘DF-(length of 5’flap), (length of 3’flap)’. In our studies, we have used several schemes for the labeling of the DF substrate (141, 176, 252). One of them used a Cy3, attached through phosphoramidite linkage (referred to, from here onwards, as ‘iCy3’) at the tip of the 5’flap (Figure 4.2A). In this scheme, we were puzzled by a photophysical protein-induced effect that was manifested by iCy3 quenching beyond FRET (252). Steady state fluorescence spectra of iCy3 in DF-6,1 showed a major fluorescence quenching (up to 40%) upon addition of FEN1, without causing any spectral shift (Figure 4.2.1A). This quenching depends on the length of the 5’flap, starting with 52% in DF-2,1 and reaching almost zero in DF-18,1 (Figure 4.2.1B). Hence, iCy3 quenching is distance-dependent, similar to PIFE (125). None of the known quenching mechanisms can explain this observed effect, particularly since FEN1
does not contain any iron sulfur cluster and carbocyanine dyes cannot be quenched, via PET, by tryptophan residues or guanosine nucleobases (254, 255). We thus conclude that this iCy3 quenching is novel and refer to it as ‘protein-induced fluorescence quenching’ (PIFQ), analogous to PIFE.

Next, we used the FEN1/DF system to characterize the nature of PIFQ and the factors that affect it. We investigated whether the observed PIFQ was mediated by the DNA itself or by a direct effect of FEN1 on the photophysics of iCy3. In our experiments, we measured the time-resolved fluorescence lifetime of iCy3, for each flap length, in three contexts: 1) the 5'-iCy3-labeled oligo, 2) the iCy3-labeled DF substrate and 3) the iCy3-labeled DF in the presence of FEN1 (Figure 2A). The measurements, in the case of DF-6,1 for example (Figure 2B), confirmed the quenching behavior upon FEN1 binding; the lifetime of iCy3 in the DF was ~2 ns and decreased to ~1.2 ns upon addition of FEN1. Surprisingly, the lifetime of iCy3 in the oligo alone was ~1.5 ns. This shows that, upon oligo annealing within the DF structure, the fluorescence of iCy3 was enhanced dramatically (40 %) and that FEN1 binding reduced this enhancement, back to a similar level as in the oligo alone. The same trend was observed across the different flap lengths (2-18 nt long) (Figure 2C). Interestingly, although the different oligos had similar initial fluorescence lifetimes, their levels of enhancement, upon annealing to form the DF substrate, decreased as the flap length increased. A plot of the percentage of change in lifetimes versus flap length shows an anti-correlated behavior, characterized by an increase in fluorescence upon formation of the substrate that is counteracted by FEN1 binding to the DF substrates (Figure 2D). Therefore, the
inherent fluorescence increase upon DF substrate formation is driving the subsequent quenching effect. Thus, the observation can be better understood from the perspective of the DNA as a mediator of fluorescence change rather than the protein being the active re-configurator of fluorescence.

To account for the fluorescence enhancement upon formation of the DF substrate and the subsequent quenching upon protein binding, various possibilities were considered. We first ruled out any ground state-induced static quenching by measuring the absorbance of DF substrates either with or without FEN1. These measurements showed no significant difference in the absorption between the two conditions (Figure 4.2.1C). Furthermore, both fluorescence lifetime and quantum yield (Figure 2C and Figure 4.2.1E, and F) showed a similar trend, suggesting that nonlinear effects were not the cause for such changes in fluorescence either. To probe whether the fluorescence modulations in FEN1/DF system are analogous to those in PIFE, particularly with respect to the modulation of photoisomerization, we exchanged the iCy3 fluorophore with a Cy3B in our 3-context system of oligo, DF, and DF/FEN1 for flap lengths of 2, 4 and 6 nt. Cy3B is an analog of Cy3 dye, but with a rigid inter-heterocyclic construction rather than a rotationally flexible polymethine bond (Figure 4.1.1). This rigidity renders Cy3B incapable of any photoisomerization and unresponsive to any dynamics that may affect this isomerization, thus making it (125, 230) environmentally insensitive, and therefore an ideal control for PIFE/PIFQ. In our study, the fluorescence lifetimes of the Cy3B-bearing oligos alone and their corresponding DF substrates either with or without FEN1 showed no difference (Figure 2E). Moreover, these fluorescence lifetimes
were comparable to that of the free Cy3B dye. The value of the total non-radiative excited-state lifetime loss obtained theoretically anti-correlated with the fluorescence lifetime of iCy3 in the oligo, DF, and DF/FEN1 system (Figure 4.2.1D). This loss was found significantly higher than that of Cy3B, since it also accounted for a photoisomerization loss. The Cy3B experiments, along with the total non-radiative lifetime loss, show that for the iCy3-labeled system, both PIFE and PIFQ stem from the excited cis-trans photoisomerization.

Our results also suggest that the isomerization of iCy3 in the DF substrate system is likely influenced by its interactions with the neighboring DNA nucleotides. This proposition is supported by the decrease in iCy3 fluorescence enhancement upon formation of the DF substrate as the 5’flap length increases (Figure 2C and D). To directly support this hypothesis, we fixed the length of the 5’flap but varied its sequence, and showed that the lifetime of iCy3 upon formation of the DF substrate was sequence-dependent (Figure 2F). Interestingly, the lifetimes of the DF substrate in the presence of FEN1 were nearly similar, consistently with the similar final states observed when varying the length of the 5’flap using polydT sequence. This supports our findings that PIFQ is mediated by the DNA and that FEN1 simply reverses this effect.

Shedding some light on the nature of PIFQ, we pursued further characterizations by exploring other variables that can affect it. First, we questioned whether changing the fluorophore attachment chemistry could have any effect on its fluorescence properties. Hence, we changed the DNA-Cy3 attachment chemistry from phosphoramidite to a longer linker, using Cy3 coupled
via NHS chemistry (Cy3N) for three flap oligos with variable lengths (2, 4 and 6 nt-long flaps). We observed that Cy3N behaved similarly to iCy3, but that the magnitude of fluorescence enhancement, upon formation of the DF substrate and its subsequent quenching by FEN1, was significantly reduced (Figure 2G). Additionally, the fluorescence modulations in the case of Cy3N seemed to have a shorter distance effect than iCy3 (Figure 2G). The change in the magnitude of PIFQ in Cy3N relative to iCy3 also supports our hypothesis that PIFQ is influenced by the interactions of the dye within the DNA. The difference between the modulations of iCy3 and Cy3N may also result from the presence/absence of sulfonate groups (Figure 4.1.1).

Following that, we investigated if PIFQ occurs in different fluorophore types. Using DF substrates of variable 5’flap lengths (2, 4 and 6), we examined the change in fluorescence, upon addition of FEN1, in the case of Alexa555 and Alexa647 coupled via NHS. We also examined the change in fluorescence for iCy5 coupled through phosphoramidite at the tip of the 5’flap. It is worth noting that, similarly to Cy3, Alexa555, Cy5 and Alexa647 contain flexible polymethine bonds (Figure 4.1.1). We showed that iCy3, Cy3N and iCy5 exhibited a significant PIFQ, while Alexa555 and Alexa647 only exhibited a moderate to low PIFE (Figure 2G). This result underscores the fact that the observed quenching effect is not a Cy3-specific phenomenon. The difference in fluorescence modulation among these fluorophores can be attributed to the difference in the length of their polymethine bonds as well as their overall charge.
In our next step, we investigated the potential impacts from changing the position of the fluorophore in the 5’ flap. Coupling iCy3 to the tip of the 5’ flap restricted its rotational freedom as shown by its higher lifetime relative to the free Cy3 dye (Figure 2C). We then restricted iCy3 further from both sides by placing it internally at position 4 in the 5’ flap while gradually adding 1 nt at a time up to 4 nts on its 5’ side. We observed that both the enhancement upon DF substrate formation and PIFQ upon FEN1 addition remained constant after the addition of one dT (DF-5,1). Nonetheless, these modulations dropped abruptly for the following addition (DF-6,1) and plateaued at a lower level with the extra additions (Figure 2H). We therefore postulate that iCy3 when restricted only from one side or slightly from the other side (1 dT), may be capable of forming interactions with the neighboring DNA nucleotides, and consequently reducing the rate of photoisomerization from the excited trans conformer. However, when the fluorophore is significantly restricted from both sides, we believe that some of these interactions may be lost, leading to a reduced effect of both the enhancement upon substrate formation and PIFQ.

Collectively, these experiments suggest that the interactions of iCy3 with the neighboring DNA are the main cause of the enhancement upon substrate annealing and that the binding of FEN1 merely breaks these interactions leading to the observed PIFQ. Moreover, they suggest that the photophysical properties of Cy3 in a particular substrate are affected by the overall DNA-dye structure and the interactions within.
Figure 4.2. A novel DNA-mediated bidirectional fluorescence modulation observed through FEN1/DF system. (A) Schematic showing the oligo bearing iCy3 at its 5’ tip, annealed to two other oligos to form a double flap structure (DF) that serves as a substrate for FEN1 binding. The ssDNA flap part of the DF is threaded through a narrow pathway in FEN1 for cleavage to happen. (B) Time-resolved fluorescence lifetime decays of Cy3 in oligo alone (grey), upon making the DF substrate (blue) and upon addition of FEN1 to the
DF substrate (cyan) with a 6 nt-long 5' flap. (C) Bar chart showing time-resolved fluorescence lifetime of Cy3 in oligo alone, upon making the DF substrate and upon addition of FEN1 to the DF substrate for flap lengths 2-18 (color scheme is same as B). Fluorescence lifetime of free Cy3 dye is included (in purple) for comparison. (D) Graph showing percentage of change in fluorescence lifetime, upon annealing of the DF substrate and FEN1 binding to the annealed substrate as a function of flap length. (E) Bar chart graph showing the time-resolved fluorescence lifetime of Cy3B-labeled oligos alone, upon making the DF substrate and upon addition of FEN1 to the DF substrate for flap lengths 2, 4 and 6. Cy3B is linked to the DNA via a 13-carbon linker. (F) Bar chart showing the time-resolved fluorescence lifetime of Cy3 in the oligo alone, upon making the DF substrate, and upon addition of FEN1 to the DF substrate for 6 nt-long 5' flap with different flap sequences. (G) Graph showing the percentage of change in fluorescence, based on lifetime measurements in the DF substrates for different fluorophores and different linkers, with and without FEN1. iCy3 is linked to DNA via a 3-carbon linker. Other fluorophores used are: Cy3B, Cy3N, iCy5, Alexa555, and Alexa647. (H) Graph showing the percentage of change in fluorescence for Cy3 upon annealing of the substrate and addition of FEN1, when iCy3 is incorporated inside the ssDNA flap.
Figure 4.2.1. PIFQ likely stems from modulation of photoisomerization rates (A)
Steady state emission spectra of the DF substrate with flap length, with and without FEN1 showing that the Cy3 fluorescence of the DF substrate is quenched by FEN1. (B) Graph showing the flap length dependence of FEN1-induced quenching of DF substrates, with flap lengths varying from 2 to 18. (C) Graph showing percentage of change in ground state absorption of the DF substrate, with and without FEN1. (D) Bar chart graph showing the loss of the excited-state lifetime due to non-radiative de-excitation, including the photoisomerization from trans*. This loss is calculated, as described in Materials and
Methods section, for Cy3 (purple) and Cy3B (grey) dyes, in the following four forms: free dye, bound to oligo, and in DF substrate with and without FEN1. The horizontal dashed line indicates the average total non-radiative loss in Cy3B for the four forms. Further loss (above the dashed line) in Cy3 is attributed to the photoisomerization loss from trans* excited state. (E) Bar chart graph showing the fluorescence quantum yield of Cy3 in oligo alone, upon making the DF substrate and upon addition of FEN1 to DF substrate for flap lengths 2-18. (F) Graph showing percentage of change in quantum yield, upon annealing of the DF substrate and FEN1 binding to the annealed substrate as a function of flap length.

4.3.2 The initial lifetime of Cy3-DNA sets the stage for fluorescence enhancement or quenching in the presence of RPA.

We next explored the versatility of PIFQ and its adaptability to other systems. We opted to investigate the interaction of the human ssDNA-binding protein RPA with ssDNA (Figure 4.3A). This system removes the peculiarity of the DNA-mediated fluorescence enhancement seen in the case of DF DNA and sets the stage for a general DNA-protein binding system. Since ssDNA-binding proteins have been shown to induce PIFE on labeled ssDNA (231), we questioned whether the binding of RPA could induce PIFE, PIFQ, or no change at all, and which factors may affect the outcome. To answer this, we designed a library of iCy3-labeled ssDNA oligos with varying sequences, fluorophore positions (3’, 5’, or internal) (Figure 4.6) and fluorophore types. Throughout these experiments, the oligo length was kept constant at 22 nts, which supports 1:1 binding of ssDNA:RPA, via the known tight binding mode (256).

Our library included 16 3’-iCy3-labeled oligos with varying sequences (Figure 4.3B). The initial fluorescence lifetimes of these oligos ranged between 0.5
to 1.5 ns. Upon addition and binding of RPA, all oligos exhibited an increase in fluorescence to a varying degree. However, the final fluorescence lifetimes for the various oligos in the presence of RPA seemed to average around a threshold of 1.67 ns, regardless of the initial lifetimes. In general, the degree of fluorescence enhancement seemed to depend on the initial fluorescence lifetime of the ssDNA oligo itself, with a Pearson coefficient $\rho = -0.72 \pm 0.13$.

On the 5' side, the initial fluorescence lifetimes of 21 iCy3-labeled oligos ranged from 0.7 to 2 ns (Figure 4.3C). This library of oligos exhibited a more versatile modulation of fluorescence where the binding of RPA produced either enhancement, quenching or insignificant effect on the oligos tested. Overall, we observed that the initial fluorescence lifetime of the oligos seemed to dictate the protein-induced modulation of the fluorescence. In most cases, oligos with lifetimes of 1 ns or below showed a PIFE effect whereas oligos with lifetimes of 1.3 ns or higher showed a PIFQ effect; oligos with lifetimes in-between did not show any significant change. Nevertheless, irrespective of the type of effect, the modulation of the RPA appeared to bring the oligos fluorescence lifetime to a threshold of 1.24 ns. The anti-correlation between the initial lifetime and the fluorescence change, upon RPA binding, was found significant, with a Pearson coefficient $\rho = -0.80 \pm 0.16$.

The library of 15 internally-iCy3-labeled oligos displayed diverse initial lifetimes that ranged from 1 to 2.8 ns, a higher range than both 3' and 5'-iCy3-labeled oligo libraries (Figure 4.3D). Surprisingly, O-328, with an initial lifetime of 2.81 ns, displayed an even longer fluorescence lifetime than that of Cy3B of 2.42
ns (Figure 4.3D and Figure 2E). This library shows versatile fluorescence modulations, despite the majority showing PIFE. Similarly to the 3'- and 5'-libraries, the RPA-induced fluorescence modulation seemed to bring the oligos fluorescence lifetime to a defined threshold, albeit a higher one (1.85 ns) than that observed with both 3'- and 5'-libraries. Again, the Pearson coefficient $\rho = -0.89 \pm 0.07$ between the initial lifetime and the change in fluorescence indicates a strong anti-correlation.

Distributions of the initial fluorescence lifetimes of the iCy3-labeled oligos, along with the final lifetimes upon addition of RPA, were plotted in box charts for the three libraries (Figure 4.3E). For the 3'-library, we found that the addition of RPA induced an overall PIFE effect with a clear shift of the distribution towards longer lifetimes. Yet, RPA binding did not significantly affect the variation of the distribution, as indicated by the similarity between its standard deviation and the deviation in the dataset containing oligos only. In contrast, in the 5'- and internal-libraries, RPA binding narrowed down the variation of the distribution of lifetimes by ~2 fold. However, these two libraries displayed an opposite overall shift of their distributions. The fluorescence modulations of all tested oligos were confirmed by steady state fluorescence measurements, and appeared to be broadly in agreement with those determined by time-resolved measurements, as indicated by the Pearson correlation coefficient of the two datasets (Figure 4.3.1B).
Figure 4.3. The initial lifetime of Cy3-DNA sets up the stage for the fluorescence enhancement or quenching in the presence of RPA. (A) Schematic describing the effect of RPA binding to an iCy3-labeled oligo with different sequences and different fluorophore positions, leading to either quenching or to a fluorescence enhancement. (B) Fluorescence lifetimes of the 16 3’-iCy3 labeled oligo library with different sequences, either with RPA (blue) or without (grey). The dashed line shows the average fluorescence lifetime in the presence of RPA. The Pearson coefficient of the correlation between the initial fluorescence lifetime (in the absence of RPA) and the change in fluorescence (%) is reported with its standard error. (C) Library of 22 5’-iCy3-labeled oligos with their fluorescence lifetimes in the absence (grey) and presence (green) of RPA is shown. The
horizontal dashed line represents the mean fluorescent lifetime of the library, in the presence of RPA. The Pearson coefficient and its standard error (as described in B) are reported. The green vertical dashed line delimits oligos that show an overall PIFE effect, upon addition of RPA, whereas the red vertical dashed line delimits those showing an overall quenching effect. (D) Bar chart showing the fluorescence lifetimes of a library consisting of 15 internally-iCy3-labeled oligos in the absence (grey) and presence of RPA (purple). Horizontal dashed line, Pearson coefficient, green and red vertical dashed lines are as described in C. (E) Box chart summarizing the distribution of the lifetimes for the 3 libraries described above, with and without RPA. (F) Representation of the effect of the fluorophore type and position (6 different positions as illustrated in Figure 4.3.1A). The change in fluorescence (%), upon the addition of RPA are shown for the 6 oligos with different fluorophore types and linking chemistries (iCy3 in cyan, Cy3N in green, Cy3B in black, Alexa555 in magenta, Alexa647 in blue, iCy5 in red and DyLight633 in orange). The horizontal dashed line represents the zero line.

The next step of our study involved adding one more dimension to the diversity of our ssDNA library by varying the type of fluorophore and its position. We tested 7 different fluorophores, namely iCy3, Cy3N, Cy3B, Alexa555, Alexa647, iCy5 and DyLight633, each placed at 6 different positions within a 22 nt-long oligo (Figure 4.3.1A). Except for Cy3B, all fluorophores contained a polymethine bond capable of undergoing cis-trans photoisomerization (Figure 4.1.1). It is worth noting that the sequence of these oligos is, in principle, preserved; however, changing the position of the fluorophore entails changing the local sequence sensed by the fluorophore. RPA induces PIFE in most fluorophores at different positions with the exceptions of Cy3B, 5’end-iCy3, and 5’end-iCy5. PIFE was found significant for iCy3, Cy3N and Alexa647 (20-80%) but highly elevated for Alexa555 (80-160%). These findings could have a substantial impact on the
experimental design, when a significant PIFE effect is desired. The variations in
the fluorescence change, observed for different dyes, maybe be attributed to the
different lengths of the polymethine bond, overall charge of fluorophore as well as
the position and chemistry of its linker. These parameters can influence the way
the fluorophore interacts with the DNA and the protein, in the initial and final states.

Figure 4.3.1. Steady state measurements correlate with time-resolved results. (A)
Schematic illustrating the different fluorophore positions in oligos O-278 to O-283 from 5'
to 3' end. (B) Plot showing the change in fluorescence (%) upon RPA binding to various
oligos from all 3 libraries shown in Figure 4.3B-D. The fluorescence change determined
by time-resolved fluorescence measurements (red) and steady-state fluorescence
measurements (blue) are both shown with significant agreement between the two
measurements as indicated by Pearson coefficient.
4.3.3 Insights into the structural properties of the DNA-dye complexes

This set of experiments is designed to further investigate the initial fluorescence state of a library of oligos, by probing the role of different external modulators on the overall DNA-Dye structure and rotational freedom in the absence of protein. Assuming that PIFE is an analogous effect to the presence of a viscogen, we investigated the effects of varying the glycerol concentration on the fluorescence lifetime for several DNA constructs (Figure 4.4A). It is clear that most of our constructs, except Cy3B, responded to an increase in glycerol concentration by increasing their lifetime. Qualitatively, it can be inferred that the range of this change is proportional to the fluorophore’s photoisomerization degree of freedom in a specific configuration. Accordingly, a more dynamic range reflects higher freedom. Our tested constructs, labeled with a Cy3 fluorophore exhibited a wide range of initial lifetimes (lifetime at 0% glycerol) and a varied response to an increase in viscosity. They mainly fell within three classes: free from both ends (free Cy3 - the most sensitive to environmental changes), linked from one side to DNA (Cy3N - less sensitive to environmental changes) and linked from both sides (iCy3 - the most insensitive to environment). However, the final lifetimes at 100% glycerol for all constructs seemed to approach a similar value, which was expected, as the photoisomerization freedom is fully inhibited at a high viscogen concentration.

For a more quantitative approach, the glycerol percentage (0-60% range) was first converted to dynamic viscosity (257, 258) (Figure 4.4.1A); it was then plotted against the fluorescence lifetimes for all the various DNA constructs (Figure
4.4.1B). The result curves were fitted with a Michaelis-Menten-like hyperbola, where $K_{1/2}$ represents the dynamic viscosity at which half of the maximum fluorescence lifetime is achieved. We found that constructs with a broader dynamic range of lifetimes resulted in higher $K_{1/2}$ values, due to the fact that their initial lifetimes were significantly different from the maximum lifetime (Figure 4.4.1B). When the initial lifetimes are plotted against the inverse of the $K_{1/2}$ values, which can be viewed as a measure of the restriction of photoisomerization imposed on a fluorophore by viscosity, a striking linear dependence is observed (Figure 4.4B). This linear dependence also clearly highlights the three dye coupling classes, i.e. free dye or DNA coupled from either one or two sides. The initial lifetime is generated by a rate that can be further decoupled into two components: photoisomerization-dependent and photoisomerization-independent rates (see Methods section in Supporting Information). The latter is a sum of the radiative ($k_r$) and other non-radiative rates ($k_{nr}$). This photoisomerization-independent rate exhibited a $\sim$1.3-fold difference between the free Cy3 and the class of iCy3 constructs, which accounted for a slight increase of the lifetime of iCy3-coupled DNA (Figure 4.4.1C). On the other hand, the photoisomerization-dependent rate ($k_{iso}$) from the excited trans state, in pure water, showed a $\sim$20-fold difference between the free Cy3 and the class of iCy3 constructs (Figure 4.4.1D). Taken together, these two observations showed that the variance in the initial lifetime between different constructs stemmed mainly from the photoisomerization pathway.
We next investigated the effects of the overall structure of an oligo-dye complex, given a particular sequence, linker, and fluorophore position, on the fluorescence lifetime of the fluorophore. Previously, the term PIFE was mostly associated with an effect induced by the presence of a protein. Here, we extended this nomenclature to reflect any enhancement of fluorescence resulting from a disruption of the overall structure of the DNA-dye complex. We hypothesized that annealing a complementary strand to an oligo would disrupt the structure formed by the oligo-dye complex, just as a protein-binding would. We tested this hypothesis with a subset of the internally-iCy3-labeled oligo library (Figure 4.3D), with initial lifetimes ranging from 0.92 to 2.24 ns. The annealing of their complementary strands induced fluorescence modulations (PIFE or PIFQ) that followed the general principle observed in presence of RPA on such oligos (Figure 4.4C). This demonstrates that PIFE and PIFQ are not necessarily associated with a specific protein, but rather to the molecule that propagates the interaction, in this case the iCy3-labeled oligos. We found the average final fluorescence lifetime (1.39 ns) to be lower than that of the RPA-induced modulation (1.85 ns) (Figure 4.4D), suggesting that the external modulator can exert some influence on the final lifetime. Nevertheless, the anti-correlation between the initial lifetime and the fluorescence change still dictated the direction of the fluorescence modulation.

In the case of environmentally sensitive fluorophores, the notion of a single lifetime is replaced by a broad ‘landscape’ of lifetimes that the fluorophore can access in different environments and configurations. This has been previously characterized using a microarray of short DNA oligos labeled with Cy3 or Cy5.
assayed for their relative intensities (245, 246, 259). However, these studies were limited, as the only variable characterized was the short DNA sequence. Here, we compiled a landscape of Cy3 accessible lifetime values, from a very significant number (N=285) of lifetime measurements (presented in this study), encompassing diverse contexts such as DNA sequence, linker chemistry, buffer conditions, DNA conformation, and protein association (Figure 4.4D). This landscape stretched between two extreme values, that of free Cy3 in water and that of O-328 (18) in RPA buffer. Added to the environmental sensitivity of the dye (Figure 4.4B), this landscape allowed us to deepen our understanding of an environmentally sensitive fluorophore, somewhat similarly to the way we understand an environmentally insensitive dye from the information given by a single lifetime value.
Figure 4.4 Insights into the structural properties of the DNA-dye complexes. (A) Effect of viscosity on fluorescence; bar charts represent the fluorescence lifetimes of different fluorophores (free or DNA-Dye complexes), at increasing concentrations of glycerol (0-100%), in 10% increments. (B) Dependence of the fluorescence lifetime, without glycerol, on the viscosity resistance. The plot shows a linear dependence with indicated slope and y-intercept. The horizontal error bars represent the standard error of the mean. The goodness of the fit $R^2$ value for the linear fit is reported. (C) Fluorescence
lifetimes of 11 internal-iCy3 labeled oligo library with different sequences (grey) and their corresponding dsDNA (red). The dashed line shows the average fluorescence lifetime of dsDNA in this library. The Pearson coefficient of the correlation between the initial fluorescence lifetime (ssDNA) and the fluorescence change upon annealing the complementary oligo (%) are reported with their respective standard errors. (D) Fluorescence lifetime landscape of Cy3 compiling all the Cy3 lifetimes measured in this study (N=290). The horizontal dashed line represents the median value of the landscape. (E) Bar chart indicating the fluorescence lifetime of O-328 and its derivatives, in ssDNA (grey) and dsDNA (black) forms, measured in RPA buffer. (F) Bar chart indicating the fluorescence lifetime of O-328 in ssDNA form, in different individual RPA buffer components.

Finally, since O-328 stood out as having the highest Cy3 fluorescence, even higher than Cy3B, we aimed to further probe the basis of its high fluorescence. We postulated that such high fluorescence could be due to the rigididification of the excited trans state imposed by the overall DNA structural configuration and possibly the additional effect of other pathways beyond photoisomerizaton. To test the rigidification hypothesis, we decided to systematically perturb this structural configuration by trimming down the size of O-328 two nucleotides at a time (1 nt from each side), from 22 nts (O-328) down to 8 nts (Figure 4.4E). We found that the lifetimes of this set of oligos remained largely consistent (~2.75 ns), up to a length of 16 nts, then they dropped down gradually as the oligo length decreased to 8 nts (1.3 ns). Upon annealing the corresponding complementary strands, the formed dsDNA exhibited a constant lifetime (~1.2 ns) across different lengths. Taken together, these results indicate that a specific DNA structure may form within the core of O-328 (central 16 nts). If this assumption holds, then one would expect that some specific factors could stabilize or destabilize the formation of this
structure within O-328. In our previous experiments, we observed the high fluorescence of O-328 using RPA buffer (50 mM HEPES pH=7.5, 5% glycerol, 50 mM KCl, 1 mM MgCl₂). We postulated that the buffer viscosity, ionic strength, pH and/or divalent metals could affect the DNA-Dye structure, and consequently the dye’s fluorescence. Hence, in order to pin down the effects of these variables, we measured the fluorescence lifetime of O-328 in solutions containing each of the buffer components separately, along with the RPA buffer as a positive control, and water as a negative control (Figure 4.4F). Surprisingly, we found that O-328 lost its extraordinarily high fluorescence in water and other solutions except for the one containing 50 mM KCl, which maintained a 2.8 ns lifetime. We speculated that potassium ions probably induced the formation of a secondary structure within O-328.
Figure 4.4.1 Decoupling the rates generating the initial lifetime. (A) Conversion curve from glycerol percentage in water to dynamic viscosity. (B) Dependence of the fluorescence lifetime for the different constructs, shown in Figure 4.4A, on dynamic viscosity. Each curve was fitted to a Michaelis–Menten type hyperbola, as described in the Methods section. $K_{1/2}$ values and the inverse value of $\tau_\infty$ is indicated for each construct, together with their standard error of the mean. (C) Bar chart of the inverse values of the lifetimes in saturating glycerol $\tau_\infty$, for each of the constructs shown in Figure 4.4A. Vertical error bars represent the standard error of the mean. (D) Bar chart of the photoisomerization rates from trans* in pure water of the constructs shown in Figure 4.4A. Vertical error bars represent the standard error of the mean. The equation below c and d describe the rate components, photoisomerization-independent (in Cyan) and – dependent (in blue), of the measured lifetime for any fluorophore in a particular condition.
4.3.4 PIFQ and PIFE can be used at the single molecule level to follow catalysis, DNA conformational changes and protein binding

To explore the applicability of PIFQ to single molecule techniques, we sought to characterize FEN1’s catalytic efficiency, using a single labeled-DF. In our previous work, we used DF-6,1 with iCy3 placed at the tip of the flap and Alexa647 placed at position 12 downstream of the nick junction (141, 176, 252). With this labeling scheme, the substrate had a FRET efficiency of 0.8 in the absence of FEN1. Upon FEN1 binding and DNA bending, the FRET efficiency dropped from 0.8 to 0.48 for few frames, before FEN1 cleaved the iCy3-labeled 5’flap, leading to the instantaneous loss of the 5’flap and consequently FRET (Figure 4.5A) (176, 252). This FRET change was confirmed by the anti-correlated change in donor and acceptor signals, also evident in the time traces (Figure 4.5A and Figure 4.5.1A). The distribution of the dwell times spent in the bent state, before the loss of signal, was fitted to a gamma distribution. This distribution averaged at ~163 ms, and was used to estimate the single turnover catalytic rates (252). The distribution of these dwell times showed a rise and decay indicating a multi-step rate-limiting pathway.

As shown in Figure 2, FEN1 binding to DF-6,1 induced a 40% quenching of the iCy3-labeled 5’flap fluorescence. This PIFQ may be a useful single molecule cleavage assay, since it can reduce the complexity of the smFRET cleavage assay, especially when choosing an optimized FRET pair, detectable FRET change, and acceptors’ stability. Furthermore, the cleavage by FRET reports on all post DNA bending catalytic steps, whereas cleavage by PIFQ may report on
specific catalytic steps only. In the smPIFQ assay (Figure 4.5B), FEN1 binding causes a single step quenching before the signal is lost (Figure 4.5B and Figure 4.5.1B). This quenching step is interpreted as FEN1 recognizing its substrate, and the loss of iCy3 signal marks the instantaneous 5’flap release. Our confidence in attributing the iCy3 loss to a cleavage of the 5’flap, rather than photobleaching, is reinforced by the fact that iCy3 is stable in the absence of FEN1, under similar buffer conditions (252). Surprisingly, the distribution of dwell times before cleavage from smPIFQ showed a single exponential decay (Figure 4.5B), suggesting a single rate-limiting step, in contrast to the rise and decay distribution observed using a smFRET cleavage assay (Figure 4.5A). Furthermore, the average dwell time from smPIFQ (116 ms) was slightly shorter than that from smFRET (163 ms). We propose that the difference in the average dwell times and the distribution patterns of these dwell times, obtained via the two assays, can be interpreted by imagining that smPIFQ and smFRET start reporting at different time points of the FEN1 cleavage reaction. It may be possible that FEN1 induces quenching of Cy3’s fluorescence, only once the flap has been fully threaded and positioned, whereas smFRET starts reporting on the DNA bending, prior to the threading step. These results reveal an important mechanistic information regarding the timing of the 5’flap threading, in relation with DNA bending (252). More importantly, this shows that smPIFQ can be used to question different aspects of this cleavage pathway, hence providing an extra dimension and adding to the information already acquired from smFRET assays (176, 252).
Our next system relied on the observation that the fluorescence of O-328 was enhanced upon addition of potassium ions (Figure 4.4F). We therefore hypothesized that the presence of K\(^+\) changed the overall structure of the Cy3-labeled oligo, perhaps through the formation of a secondary structure of some sort, similar to G-quadruplexes case (260). It would therefore be expected that the addition of RPA to the system, a protein known to melt secondary structures (261), would quench the highly fluorescent O-328/K\(^+\) and attempt to restore the original fluorescent state of O-328. Our hypothesis that the addition of K\(^+\) possibly induces the formation of a secondary structure in O-328 posits one caveat. Perhaps, the addition of K\(^+\) induced the formation of a dimer, rather than a secondary structure in the monomer. To test this hypothesis, we used a single molecule technique. In the assay, we used a longer oligo containing the sequence of O-328 at one end and annealed to a biotinylated oligo at the other end, creating a primer (P)/template (T) junction. This ensured that the fluorescent end of O-328 was free and away from the surface. This design provided a well-spaced surface, where the formation of a dimer was almost impossible. In the absence of K\(^+\), the P/T junction showed a low fluorescence intensity profile, centered around \(3.8 \times 10^4\) a.u. (Figure 4.5C). The addition of 50 mM KCl clearly shifted the fluorescence intensity distribution to a higher fluorescent regime, centered around \(5.3 \times 10^4\) a.u. (40 % increase), similarly to what we observed using bulk lifetime measurements (Figure 4.4F). More importantly, the time traces upon the injection of KCl (Figure 4.5C and Figure 4.5.1C) showed an initial instantaneous transition followed by a stable higher fluorescence state. This indicates that the K\(^+\)-induced fluorescence enhancement
is a single-step process (within our temporal resolution) rather than a progressive process that goes through multiple states to reach the maximum fluorescence state. Moreover, the time traces exhibited one-step photo-bleaching events, further confirming the monomer nature of the P/T junction (Figure 4.5C and Figure 4.5.1C).

Next, we tested whether RPA could melt this K⁺-induced secondary structure. Using the same P/T junction, and in the presence of KCl, we confirmed the high intensity profile of this substrate at an average of $6.8 \times 10^4$ a.u. (Figure 4.5D). The addition of RPA, even with added K⁺, quenched the Cy3 intensity to a peak centered around $4.9 \times 10^4$ a.u. (28% decrease), indicating that RPA is capable of melting the secondary structure and reversing the K⁺-induced fluorescence enhancement to a lower intensity profile. Time traces (Figure 4.5D and Figure 4.5.1D) showed quenching in a single-step upon RPA binding, as well as single-step photo-bleaching events.

Taken together, these single molecule assays showed how fluorescence modulation, if controlled, could be used to understand diverse biochemical phenomena, at single molecule level. Both single molecule cleavage and the monitoring of KCl-induced structural changes in O-328, as well as its disruption by RPA, are examples that can be studied in a simple way, using a single fluorophore.
Figure 4.5. Single-labeled fluorescence modulation (PIFQ and PIFE) can be used at the single molecule level. (A) smFRET cleavage assay used to follow the catalytic kinetics of FEN1 on DF-6,1 substrate. Top: schematic representation of the change in FRET, from E~0.8 (free DF substrate) to bound/bent substrate upon the addition of FEN1 (E~0.48). This is followed by loss of signal upon the incision of the flap and loss of the donor. Bottom: representative single molecule trace of a cleavage event showing the FRET change and the following loss of signal. The distribution of the time spent in bent
state for N=131 cleavage events was plotted and fitted with gamma distribution. The mean and standard error of the mean are reported. More traces are shown in Figure 4.5.1A. (B) Single molecule cleavage assay (smPIFQ) with a single label (Cy3) placed at the tip of the flap. Top: schematic showing the design of the assay where FEN1 interacts with the Cy3-labeled flap and quenches Cy3’s fluorescence. In the presence of Mg\(^{2+}\), FEN1 incises the flap, leading to its release and hence the loss of Cy3 signal. Bottom: representative single molecule trace monitoring of the fluorescence quenching of Cy3 upon interaction with FEN1 and subsequent signal loss. The time spent in the quenched state is quantified for N=213 events, and its distribution is plotted and fitted to an exponential decay, with mean and standard error of the mean reported. Additional traces are shown in Figure 4.5.1B. (C) Secondary structure formation. Left: schematic showing the experimental setup. Upon the addition of K\(^{+}\), a secondary structure is formed leading to an enhancement of Cy3 fluorescence. Middle: Cy3 fluorescence intensity histograms with (green) K\(^{+}\). The histograms are fitted to Gaussian distributions. Right: representative single molecule trace showing the transitions to the enhanced-fluorescence state, upon formation of the secondary structure, as a result of the binding of K\(^{+}\). Movies were recorded with a 100 ms temporal resolution. Additional traces are shown in Figure 4.5.1C. (D) Melting of the secondary structure. Left: schematic drawing showing the DNA construct used in C with a secondary structure formed in the presence of K\(^{+}\). However, upon addition of RPA, this secondary structure is melted, leading to the quenching of Cy3’s fluorescence. Middle: Intensity histograms of the Cy3 fluorescence, in the absence (green) and presence (blue) of RPA. Histograms are fitted with Gaussian distributions. Right: representative single molecule trace showing the transitions to the quenched-fluorescence state, upon melting of the secondary structure induced by the binding of RPA. Measurements were recorded with a 100 ms temporal resolution. Additional traces are shown in Figure 4.5.1D.
Figure 4.5.1. Examples of single molecule traces supporting Figure 4.5. (A) Representative smFRET traces from cleavage assay of DF-6,1 by FEN1, as described in Figure 4.5A. (B) Representative smPIFQ traces from cleavage assay of DF-6,1 by FEN1, as described in Figure 4.5B. (C) Examples of time traces of the secondary structure formation, within the context of O-328 and upon addition of K⁺ as well as their corresponding Cy3 fluorescence enhancements, as shown in Figure 4.5C. (D) Additional time traces of the melting of the secondary structure by RPA, within O-328 and in the presence of K⁺. Traces show the quenching of Cy3 fluorescence, upon interaction of RPA with O-328, as illustrated in Figure 4.5D.
4.4 Discussion

Understanding biological processes using fluorescence tools, whether at the single molecule level, or at the ensemble-level, may benefit from simpler approaches that do not require many extraneous labels. Increasingly, the modulation of fluorescence upon binding of an interacting-partner or conformational changes within a single biomolecule is being leveraged to gain valuable knowledge of a multitude of biological phenomena and processes, and to elucidate the underlying mechanisms. In the context of protein binding to dye-labeled DNA, PIFE has been widely used in both single molecule- and ensemble-assays. Furthermore, the modulation of the fluorescence intensity, within a single biomolecule, has also been used to understand the conformational changes within single proteins (262).

In this study, we showed, with our FEN1/DF system, that an opposite effect of PIFE exists. In this system, we witnessed both a fluorescence enhancement upon substrate formation in the absence of any protein, as well as a subsequent FEN1-mediated fluorescence quenching (Figure 2C and D). Annealing the oligos to make the DF substrate had a dramatic influence on the fluorescence of Cy3, as quantified by both the fluorescence lifetime and quantum yield. Experiments with the Cy3B-labeled flap substrates strongly suggest that the fluorescence modulation is mediated by photoisomerization (Figure 2E). The anti-correlated fluorescence enhancement upon formation of the substrate and the subsequent quenching upon FEN1 binding clearly suggest that quenching itself is due to the modulation of the inherently high fluorescent state of Cy3 in the DF substrate. We
examined multiple parameters associated with Cy3 labeling (for example, linking chemistry, linker length and Cy3 neighboring DNA sequence) as well as different fluorophores, to investigate which of these parameters might affect the observed quenching process (Figure 2E-H). Among the tested fluorophores, Cy3 and Cy5 displayed a significant quenching upon FEN1 binding.

Our results suggest that the initial fluorescence state is a key factor that determines whether a PIFE or a PIFQ effect will occur. Indeed, when using a library of iCy3-labeled oligos that exhibit various initial fluorescence states, we observed that both PIFE and PIFQ effects could be generated upon RPA binding, depending on the initial fluorescence of the Cy3-DNA (Figure 4.3B-D). We also encountered few Cy3-labeled oligos that practically showed no effect. Most importantly, for both PIFQ and PIFE, the change in fluorescence and the initial fluorescence were found highly anti-correlated (Figure 4.3B-D). Taken altogether, the data suggest that RPA binding to the ssDNA does not actively modulate the Cy3 fluorescence to a similar outcome (i.e. fluorescence increase, decrease or no change), but rather the outcome itself is dependent on the initial fluorescence.

The results presented in this study drew our attention to the use of PIFE (or PIFQ) as a protein-binding assay, in particular for associating of PIFE (or PIFQ) with a positive binding of the protein, at a specific site. Special caution should be exercised when interpreting the fluorescence modulations. For instance, ssDNA oligos with lengths that support the binding of RPA, but with sequences that have an initial fluorescence around the average value of the final Cy3-labeled DNA-RPA complex, did not show any significant change upon addition of excess RPA (Figure
Hence, without a proper preliminary assessment, one could falsely attribute the lack of change in fluorescence to a poor (or even non-existent) affinity of RPA to these DNA sequences. In such scenarios, one could possibly change the sequence to generate a fluorescence modulation, or to amplify the existing one for a more favorable modulation. We must also stress that caution should be exercised with the interpretation of results in certain situations that require an unchanged quantum yield of either donor or an acceptor, or both, as it is the case when distinguishing simultaneous FRET changes and PIFE or PIFQ. This can be achieved by changing the sequence or the linking chemistry to show minimal or no effect whatsoever.

A systematic comparison of the commonly used fluorophores in order to determine their relative predisposition and degree of fluorescence change upon protein binding, has not, to our knowledge, been reported. For this reason, we used RPA-ssDNA system to establish such parameters for commonly used fluorophores (Figure 4.3F). Except for Cy3B, which was found insensitive to protein binding (as expected), all other fluorophores displayed a modulation of the fluorescence, upon binding of RPA. The dominant PIFE outcome might be a result of our experimental design in which only one particular sequence was used, probably a sequence with a low initial fluorescence lifetime. Interestingly, Alexa555 showed an extreme change in fluorescence, for all tested positions, consistently exhibiting a PIFE effect of more than 100%. This suggests that Alexa555 may be an ideal environmentally sensitive fluorophore, and hence be quite powerful for PIFE studies (Figure 4.3F).
Since the preceding RPA experiments implicated the initial state of a system in the modulation of fluorescence, we further investigated this initial state in the Cy3-DNA system without protein. Our results showed that when Cy3 is more rigidly bound to the DNA, its lifetime displays less sensitivity to changes in the environmental conditions, thus exhibiting a less dynamic change in its fluorescence (Figure 4.4A). They also showed that when Cy3 becomes more rigidly bound to the DNA, its initial lifetime gets longer (at 0% glycerol). Indeed, the relationship between $1/K_{1/2}$ (a measure of viscosity resistance and therefore of a fluorophore’s rigidification) and the initial lifetime, under various Cy3 conditions (bound and free), showed a linear dependence (Figure 4.4B). The slope of this linear relationship can then be used as a quantitative parameter to evaluate any fluorophore’s propensity to change its fluorescence, under the change of environmental factors.

The fluorescence lifetimes of Cy3 generated under all the different conditions being studied were compiled and yielded the fluorescence landscape shown in Figure 4.4D. This comprehensive landscape provides a valuable insight into Cy3 fluorescence lifetime bounds. Free Cy3 and O-328 derivative occupy the extremities of this curve. These upper and lower bounds exhibit ~15-fold difference in lifetime with a median of 1.35 ns for the distribution of all Cy3 lifetimes. Together, the viscosity dependence (Figure 4.4B) and the fluorescence landscape (Figure 4.4D) demonstrate the ease of modulating Cy3’s fluorescence and the range of this modulation, upon changing the environment. We believe that such a better understanding of the Cy3 photophysics may be applicable to other
carbocyanine dyes. The importance of the characterization of the Cy3 fluorescence parameters and their modulation can be indispensable for the design of fluorescence-based experiments. For instance, one would choose a condition that yields a fluorescence state at the bottom of the lifetime landscape, if the preferred outcome is to generate a PIFE effect. The same logic would be applied if PIFQ is sought, for which one would prefer a condition where Cy3 fluorescence lies in the upper region of the landscape. Likewise, conditions yielding fluorescence lifetimes around the median of the landscape could be favored, if no photophysical effect upon protein binding is required, for example, when the fluorophore is to be used for accurate FRET measurements to monitor conformational changes. Sequences of desired initial lifetime could be used from the library presented in the current work or from other high-throughput sources (245, 246, 259). To further optimize such systems, we propose, in future experiments, to identify patterns in labeled DNA sequences and to correlate them with their corresponding structure and consequently their lifetime.

iCy3 labeled ssDNA exhibited a variance in their initial fluorescence lifetimes but were brought to similar values by the presence of RPA (Figure 4.3B-D). Likewise, annealing the corresponding complementary oligo to the same set of internally-iCy3-labeled oligos mimicked the presence of RPA to a large degree (Figure 4.4D and Figure 4.3D). Thus, we concluded that the initial fluorescence in these iCy3-oligos most likely stemmed from their overall structure. This overall structure has the potential to be disrupted by an external modulator (RPA or complementary oligo). Nonetheless, the difference in the average final lifetimes in
the presence of RPA or the complementary oligo indicates a potential role for the external modulator in the final outcome (PIFE, PIFQ or no change). It is also worth noting that the iCy3-dsDNA system at an average fluorescence lifetime of 1.39 ns lies in the median region (1.35 ns) of the Cy3 lifetime landscape (Figure 4.4C and D). Hence, we advise extreme caution when such dsDNA constructs are to be investigated with protein binding assays.

O-328 revealed an extremely interesting aspect in that it exceeded Cy3B in its lifetime (Figure 4.3D). We hypothesized that such state was generated by the rigidification of the trans state, possibly upon the formation of a secondary structure. Perturbing this secondary structure by shortening the oligo from the extremities resulted in a gradual decrease in lifetime of Cy3, presumably due to the destabilization of the secondary structure. However, disrupting this structure by annealing the corresponding complementary oligo to the different O-328 derivatives displayed a striking consistently low lifetime, across all lengths (22 to 8 bps) (Figure 4.4E). It then seems logical that the high fluorescence in this oligo was indeed due to the formation of a secondary structure, and that melting it decreased the fluorescence almost to the median value of Cy3’s lifetime landscape (Figure 4.4D and E). Probing some factors that could result in the formation of this secondary structure revealed that the fluorescence of Cy3-labeled-O-328 was significantly sensitive to the presence of K⁺ ions reminiscent of the G-quadruplex structure (Figure 4.4F). We expect that this sequence can be integrated into other sequences to generate PIFQ.
Using multiple single molecule experiments, we showed how our understanding of the fluorescence modulations phenomena could be applied to rationally design PIFE or PIFQ experiments to study catalytic kinetics, conformational changes and protein binding at the single molecule level (Figure 4.5). As a proof of concept, our smPIFQ cleavage assay with a singly labeled DF/FEN1 system showed that we can visualize FEN1’s cleavage reaction, in real time, and that we can access FEN1 catalytic kinetics in a similar fashion to smFRET assay (176, 252). However, with our smPIFQ cleavage assay, we were able to track the cleavage at various steps different from those observed with smFRET, making PIFQ and smFRET complimentary assays, rather than redundant. On the other hand, single molecule assays using a singly labeled P/T junction, within the context of O-328 sequence, allowed us to monitor the formation of the K⁺-induced secondary structure, and its melting by RPA, in real time. Thus, we demonstrated how PIFQ, similar to PIFE, could be used to study conformational changes as well as protein binding.

Using a combination of time-resolved fluorescence lifetime, steady state fluorescence measurements, and single molecule assays, we have shown that Cy3, and by reasonable extrapolation all the other environmentally sensitive dyes, had the unique property of context-dependent fluorescence identity, and that this property could be significantly leveraged by careful considerations of the fluorescence parameters, in the initial state of the system being studied. If the initial fluorescence state of the Cy3-conjugated system can be manipulated by changing the parameters affecting Cy3’s fluorescence (DNA sequence or surrounding amino
acid residues, position of fluorophore, fluorophore linker chemistry, and buffer conditions including viscosity and ionic strength) to occupy either extremity of Cy3 fluorescence landscape, fluorescence enhancement or quenching can be achieved. Here, we demonstrated that the change in fluorescence did not require the presence of an external molecule, as internal conformational changes in systems such as ssDNA to dsDNA conversion were sufficient to significantly and effectively change the fluorescence signal.

4.5 Materials and Methods

4.5.1 Proteins expression and purification
Human FEN1 was expressed and purified, as described previously (252). For human RPA, the plasmid expressing all 3 subunits (pET11d-tRPA) was a generous gift of Prof. Marc S. Wold, and the protein was expressed and purified as described (217).

4.5.2 DNA oligos and substrates
DNA oligos were custom-synthesized with their respective modifications by Integrated DNA technologies (IDT) or Sigma-Aldrich. Oligos with phosphoramidite coupled Cy3 (iCy3) or Cy5 (iCy5) were directly purchased from IDT or Sigma. However, oligos used for Cy3B, Alexa647, Alexa555, Cy3N and DyLight633 labeling were ordered from IDT harboring site-specific amine-modified thymine. All
DNA oligos were HPLC purified and the list of these oligos is shown in Figure 4.6. The monofunctional NHS dyes were purchased from GE healthcare. For the labeling reactions, the dyes were dissolved in DMSO to a final concentration of 20 mM while the modified oligos were dissolved in DNase free H₂O to a final concentration of 0.2 mM. The reactions proceeded by mixing the oligos at a final concentration of 0.05 mM with ~40-fold molar excess of the different dyes in freshly prepared labeling buffer containing 50 mM Na₂[B₄O₅(OH)₄] (pH 8.5) followed by 6 hr incubation at room temperature in the dark with gentle mixing. Ethanol precipitation was then used to precipitate the labeled oligos out of the excess dyes. Further purification of the labeled oligos was performed using denaturing polyacrylamide gel electrophoresis (PAGE). The labeling efficiency was calculated by comparing the absorbance of DNA at 260 nm and that of the respective dye at its absorption maximum wavelength. Oligos were labeled with >90% labeling efficiency.

Double Flap (DF) substrates were constructed by annealing the three strands, template: 5’ flap: 3’ flap in 2: 1: 4 molar ratios, in TE-100 buffer (50 mM Tris–HCl pH 8.0, 1 mM ethylenediaminetetraacetic acid (EDTA) pH 8.0, 100 mM NaCl). For annealing, a thermocycler PCR machine was used where the oligos mixture was heated at 95°C for 5 min followed by step cooling at a rate of 1°C per 1 min down to 25°C. Similarly, dsDNA and primer/template (P/T) junction substrates were constructed by mixing labeled: unlabeled oligos with a ratio of 1:3 in TE-100 buffer and annealed using the same thermocycling method. Substrates purity was quantified over non-denaturing PAGE and was assessed to be >90%.
Figure 4.6. List of all the DNA oligos and substrates used in this study.

4.5.3 Steady state fluorescence

Steady state fluorescence measurements for FEN1/DF system were conducted at room temperature using Fluoromax-4 (HORIBA JOBIN YVON). The fluorescence intensities of DF substrates (DF-2,1 to DF-18,1) harboring iCy3 at the tip of their 5' flaps were measured in FEN1 reaction buffer (50 mM HEPES-KOH pH 7.5, 100 mM KCl, 5% (v/v) glycerol, 10 mM CaCl₂, 0.1 mg/mL bovine serum albumin (BSA), and 1 mM Dithiothreitol (DTT)). For measurements in the presence of FEN1, a...
saturating concentration of FEN1 was added to the mixture of labeled substrates in FEN1 reaction buffer and allowed to reach equilibrium state after 3 mins incubation at room temperature. In both cases, iCy3 was excited at 535 nm ($\lambda_{\text{max-ex}}$) 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 FEN1 reaction buffer. For each measurement, the steady state fluorescence intensity was evaluated by integrating the corresponding emission spectrum with ± 5 nm (slit width) bounds around 565 nm ($\lambda_{\text{max-em}}$). The steady state (SS) fluorescence change upon FEN1 addition to the DF substrates was calculated as a percentage of the difference between the initial fluorescence intensity ($I_i$) of DF substrates and the final fluorescence intensity ($I_f$) after the addition of FEN1 to the initial fluorescence intensity (SS Fluorescence Change = $100 \times (I_i - I_f) / I_i$).

For oligo/RPA system, steady state fluorescence measurements of iCy3-labeled oligos were performed at room temperature using a microplate spectrofluorometer (TECAN infinite M1000) in RPA reaction buffer (50 mM HEPES-KOH pH=7.5, 50 mM KCl, 5% glycerol and 1 mM MgCl$_2$) in the absence and presence of RPA at saturating concentration. The samples were excited at 535 nm ($\lambda_{\text{max-ex of Cy3}}$) and emission was collected at 565 nm ($\lambda_{\text{max-em of Cy3}}$) with 5 nm slit width for both excitation and emission and an integration time of 0.1 s. The measurements were corrected by subtracting the emission of RPA reaction
buffer used as a blank. The fluorescence change (%) upon addition of RPA was evaluated as described above.

4.5.4 Absorbance measurements

The absorbance of iCy3-labeled DF substrates in the absence and presence of FEN1 was measured at room temperature using Thermo-Scientific™ Evolution™ Spectrophotometer. These measurements were taken within FEN1 reaction buffer in 10 mm quartz cuvettes. The bandpass was set at 1 nm and concentrations of samples were kept diluted such that the absorbance measurements were below 0.1. This measure was taken into consideration to minimize the reabsorption effect. iCy3 absorption was quantified by integrating with ± 1 nm bounds around the absorption maximum (535 nm). The instrumental baseline was recorded before each measurement with a blank sample containing FEN1 reaction buffer. For each sample, the absorbance measurements of DF substrates in the absence of FEN1 \((A_i)\) were recorded, then FEN1 was added at a saturating concentration to the same cuvette and incubated for 3 mins at room temperature. Following that, the absorbance \((A_f)\) of labeled DF-substrates was recorded in the presence of FEN1. Dilution due to the volume change upon FEN1 addition was accounted for when performing further calculations. Similar to the SS fluorescence change, the absorbance change upon FEN1 addition to the DF substrates was calculated as follows: Absorbance Change = 100 * \((A_i - A_f) / A_i\).
4.5.5 Quantum Yield measurements

The quantum yield measurements were performed following established protocols (263) based on emission and absorption measurements. For each iCy3-labeled DF substrate, three quantum yields were measured: that of iCy3-labeled 5' flap oligo and that of the DF substrate in the absence and presence of FEN1 all within FEN1 reaction buffer. For each quantum yield, the absorbance of the sample was measured (as described above), then the emission spectrum of the same sample was collected (as described above). Samples were diluted and these measurements were reiterated 3 times keeping the absorbance in the range of 0-0.1 where the absorbance and emission correlate linearly. Cy3B, dissolved in same buffer, was used as a reference dye and same procedure was applied. Hence, the quantum yield of iCy3 in each case was calculated according to this equation:

\[
Q_{\text{Cy3}} = Q_{\text{Cy3B}} \frac{E_{\text{Cy3}}(1 - 10^{-A_{\text{Cy3B}}})n_{\text{Cy3}}^2}{E_{\text{Cy3B}}(1 - 10^{-A_{\text{Cy3}}})n_{\text{Cy3B}}^2}
\]

where \( Q \) is the quantum yield, \( E \) is the integrated emission across the whole spectrum (550-700 nm), \( A \) is the absorbance at 535 nm, \( n \) is the refractive index of the medium. The values of the refractive indices are same for both the reference and sample. The quantum yield of Cy3B (0.67) was used as reported (264).

4.5.6 Theoretical non-radiative lifetime loss calculations

For the determination of the non-radiative loss of Cy3- and Cy3B-coupled DNA, the natural fluorescence lifetime (radiative lifetime, in the absence of non-radiative
competing pathways) was calculated based on the widely used Strickler-Berg equation (265). For a transition from a lower energy \( l \) to a higher energy level \( u \), induced by light absorption, the radiative lifetime of the excited state \( (\tau_{\text{rad}}) \) is given by:

\[
\frac{1}{\tau_{\text{rad}}} = 2.88 \times 10^{-9} n^2 \langle \tilde{\nu}_f^{-3} \rangle_{\text{Av}}^{-1} (g_l/g_u) \int \epsilon(\tilde{\nu}) d\ln(\tilde{\nu})
\]

where \( n \) is the refractive index of the buffer, \( \tilde{\nu} \) is the frequency of the transition expressed in cm\(^{-1}\), \( \epsilon(\tilde{\nu}) \) is the molar extinction coefficient spectra, and \( g_l \) and \( g_u \) are the degeneracies of the lower and upper states, respectively, with \( (g_l/g_u) = 1 \) for fluorescent transition. The integration and the averaging of the third negative power of the frequency of the transition are performed over the whole absorption spectra. This formula was initially used for atomic species, but it estimates well the radiative fluorescence lifetime for rigid fluorophores. In the presence of non-radiative competing pathways, the measured fluorescence lifetime will always be shorter than this theoretically estimated value, in correlation with the quantum yield. Therefore, the ratio between the measured fluorescence lifetime and theoretical radiative fluorescence lifetime serves as an estimate of the non-radiative competing pathways, including trans->cis loss (for reviews, consult (250)). In our experiments, the molar extinction coefficient spectra were determined by normalizing the absorption spectra over several dilutions and the refractive index was approximated by the refractive index of water. Thus, the non-radiative lifetime loss percentage was approximated by \( \Delta \tau = (\tau_{\text{rad}}-\tau_{\text{exp}})/\tau_{\text{rad}} \times 100 \).

For free Cy3 and Cy3B, the radiative fluorescence lifetime was estimated by
dividing the published fluorescence lifetimes by the quantum yields of these two dyes (239).

### 4.5.7 Time-resolved fluorescence lifetime measurements

Time-resolved fluorescence lifetime measurements were carried out using QuantaMaster 800 spectrofluorometer (Photon Technology International Inc.) equipped with a Fianium supercontinuum fiber laser source (Fianium, Southampton, U.K.) operating at 20 MHz repetition rate. Arrival time of each photon was measured with a Becker-Hickl SPC-130 time-correlated single photon counting module (Becker-Hickl GmbH, Berlin, Germany). Measurements were collected under magic angle (54.7°) conditions and photons were counted using time to amplitude converter (TAC). To reduce the collection of scattered light, a longpass filter (550 nm) was placed at the emission side. In all measurements, 10,000 counts were acquired. The instrument response function (IRF) was estimated using a Ludox colloidal silica suspension dissolved in water.

Measurements were recorded at room temperature in FEN1 reaction buffer for DF substrates-related samples, in RPA buffer for RPA-related samples, or other buffers as indicated in the corresponding sections. Samples containing green fluorophores (iCy3, Cy3N, Alexa555, and Cy3B) were excited at 532 nm and emission was collected at 568 nm with 5 nm slit width for both excitation and emission. On the other hand, samples containing red fluorophores (Cy5 and Alexa647) were excited at 632 nm and emission was detected at 650 nm with similar slit widths (5 nm) for both excitation and emission. For DyLight633 the
excitation/emission wavelengths were 640/655 with 5 nm slit widths. The fluorophores lifetime decays were then obtained using FluoFit software package (PicoQuant) applying the IRF and fit to one or two-exponential decay. The one-exponential decay fit was used for free fluorophores while the two-exponential decay fit was utilized for all other samples. The best fit was chosen based on reduced chi-square and randomness of the residuals.

The reported lifetimes are the amplitude-averaged lifetimes. The fluorescence change based on the lifetime measurements was calculated as the percentage of the difference between the final lifetimes upon protein addition to DF substrate (in case of FEN1) or oligo (in case of RPA) and the initial lifetimes of these DNA constructs as compared to the initial lifetimes, keeping in mind the sign of the change. The Pearson correlation coefficient reports the correlation between this percentage of fluorescence lifetime change and the initial lifetime. This coefficient of the correlation was calculated using bootstrap statistics (266).

Experiments probing the effect of viscosity on the fluorescence lifetimes of fluorophores were performed according to the methods described above. Each studied fluorophore, free or attached to DNA, was dissolved in increasing concentrations of glycerol (0-100% v/v in increments of 10%) diluted in DNAse free water. These increasing concentrations of glycerol were used to determine the dynamic viscosity of each sample according to (257, 258), keeping in mind the measurements were done at standard temperature and pressure. Fluorescence lifetimes of the different samples were then plotted against the calculated dynamic viscosity and the curves were fit with a Michaelis–Menten type hyperbola as
described below. The inverse of $K_{1/2}$ (a measure of a fluorophore’s rigidity in a particular context) was plotted against the fluorescence intensity in the absence of glycerol and the data points were fitted to a linear dependency.

4.5.8 Determination of the hyperbolic viscosity dependence of fluorescence lifetime and decoupling of rates

In general, for a fluorophore that can photoisomerize from an excited state that is capable of fluorescent de-excitation, the measured fluorescence lifetime $\tau$ of this state is given by:

$$\frac{1}{\tau} = k_r + k_{nr} + k_{iso}$$

where $k_r$ is the rate of the fluorescent radiative de-excitation, $k_{nr}$ is the rate resulting from all the non-radiative pathways except photoisomerization and $k_{iso}$ is the photoisomerization rate.

In the simplest model, the photoisomerization rate from any given state has been shown to depend on the viscosity of the environment through the dependence (267, 268):

$$k_{iso}(\eta) = D \eta^{-a} \exp(-E_0/RT)$$

where $\eta$ is the dynamic viscosity, $D$ is parameter associated with the rotational freedom of the fluorophore, $E_0$ is the height of the energy barrier that has to be crossed for photoisomerization, $a$ is a constant between 0 and 1, $R$ is the gas constant and $T$ is the absolute temperature.
For simplicity, the value of the parameter $a$ can be considered 1. Moreover, this parameter is not of interest for us as we do not intend to characterize the exact mechanism of photoisomerization. With this simplification, the measured fluorescence lifetime $\tau$ as a function of viscosity becomes:

$$\frac{1}{\tau(\eta)} = k_r + k_{nr} + \frac{D \exp(-E_0/RT)}{\eta}$$

Calculating the limit at saturating viscosity gives:

$$\frac{1}{\tau_\infty} = \lim_{\eta \to \infty} \frac{1}{\tau(\eta)} = k_r + k_{nr}$$

which yields the fluorescence lifetime at saturating viscosity, given that photoisomerization is completely inhibited.

The rate of isomerization at any viscosity can, therefore, be easily obtained by subtraction:

$$k_{iso}(\eta) = \frac{1}{\tau(\eta)} - \frac{1}{\tau_\infty}$$

We are particularly interested in the rate of photoisomerization from trans* in pure water, for which we determine the lifetime of a particular construct in pure water and in saturating glycerol.

Rearranging the terms in the first equation gives:

$$\tau(\eta) = \frac{1}{\frac{1}{\tau_\infty} + \frac{D \exp(-E_0/RT)}{\eta} + \frac{\tau_\infty \cdot \eta}{\eta + \tau_\infty \cdot \exp(-E_0/RT)}}$$
Making the notation $K_{1/2} = \tau_\infty \exp\left( -\frac{E_0}{RT} \right)$, the above equation becomes:

$$\tau(\eta) = \frac{\tau_\infty \cdot \eta}{K_{1/2} + \eta}$$

This equation describes a Michaelis-Menten type hyperbola with the asymptotic limit $\tau_\infty$. The parameter $K_{1/2}$ represents the dynamic viscosity at which half of the maximum lifetime is achieved. At constant temperature, and since $\tau_\infty$ is an intrinsic constant of the fluorophore, $K_{1/2}$ depends on the rotational freedom of the fluorophore expressed by $D$ and the height of the photoisomerization energy barrier $E_0$.

### 4.5.9 Single molecule fluorescence measurements

Single molecule measurements followed similar protocols as described earlier for smFRET measurements (141, 176, 252). Briefly, the measurements were all performed at room temperature in a custom airtight microfluidic flow cell with a glass coverslip that was 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.). DNA substrates (100-200 pM) were immobilized onto the surface using biotin-NeutrAvidin interaction. Prior to the DNA immobilization, the surface was incubated with 0.2 mg/ml NeutrAvidin for 10–15 min followed by excessive washing with reaction buffer to remove excess NeutrAvidin and block any extra unspecific binding sites. To enhance the fluorophores' photostability and reduce photo-blinking, our imaging buffer included
a mixture of reaction buffer, an oxygen scavenging solution as described (159) and 2 mM Trolox (Sigma-Aldrich). All single molecule experiments were performed using a custom-built TIRF-FRET setup (147). Several movies of each condition were recorded on different fields of view in two-color alternating excitation (2c-ALEX) (161) mode and/or continuous excitation mode. The time resolution for the different experiments is mentioned in their respective figure legends. Data extraction using twotone software (160) followed the protocols described previously (176, 252).

4.5.9.1 Single molecule FEN1 nuclease assays

FEN1 nuclease assays of DF-substrates were performed using double-labeled substrates for smFRET measurements as described previously (176, 252) or using single-labeled DF substrates to observe smPIFQ. In both experiments, 250 nM FEN1 was used to cleave the substrates in reaction buffer (50 mM HEPES-KOH pH 7.5, 100 mM KCl, 5% (v/v) glycerol, 10 mM MgCl₂, 0.1 mg/mL bovine serum BSA, and 1 mM DTT). Recording using continuous excitation of green laser was initiated prior to the arrival of FEN1 to the flow cell. The movies were recorded under continuous flow of protein at a time resolution of 50 ms.

For smFRET experiments, a cleavage event was identified by the preceding decrease of FRET signal from E~0.8 to E~0.48 signifying the bending step, an essential step in FEN1-substrate recognition, followed by the loss of signal after the incision of the 5’-flap and the loss of the donor. This decrease in FRET signal is clearly distinguishable and the donor and acceptor signals anti-correlate. In this case, the time spent in the low-FRET state (bent state) ($\tau_{\text{FRET}}$) was calculated by
manually counting the frames. The distribution of $T_{\text{FRET}}$ was plotted and fit to a gamma distribution using Matlab dfittool and the mean and standard error of the mean are reported. The average $T_{\text{FRET}}$ was then used to estimate FEN1 catalytic rate ($k_{STO}$), as we reported earlier (176, 252).

Similarly, for smPIFQ experiment, a cleavage event was identified by the preceding quenching of iCy3 signal signifying FEN1 engaging the flap substrate and followed by the loss of iCy3 signal after the incision of the flap. This quenched-state step is clearly distinguishable as it only occurs when followed by the complete loss of signal. Likewise, the cleavage event and loss of signal is distinguishable from iCy3 photobleaching as it is always preceded by a quenched-state step. The time spent in the quenched-state ($T_{\text{quenching}}$) was calculated by manually counting the number of frames. The distribution of $T_{\text{quenching}}$ was plotted and fit to a single exponential decay using Matlab dfittool and the mean and standard error of the mean are reported. This average $T_{\text{quenching}}$ was used to infer key information regarding FEN1 kinetic rates and catalytic cycle.

4.5.9.2 Single molecule monitoring of secondary structure formation and melting

For this set of experiments, the DNA substrate used is a primer-template (P/T) junction composed of a long (82 nt) iCy3-labeled oligo containing O-328 (22) sequence at its 5’end and annealed to a biotinylated complimentary short (22 nt) oligo at its 3’end. This substrate is immobilized to the surface through biotin-NeutrAvidin interaction with dsDNA region near the surface and the ssDNA region containing O-328 (22) extending further away from the surface. For monitoring the
secondary structure formation, first, the DNA substrate (100-200 pM) was immobilized to the surface in RPA reaction buffer excluding KCl. Three movies of different fields of view were recorded at equilibrium using continuous excitation of green laser. Second, RPA reaction buffer containing 50 mM KCl was injected into the flow cell. Prior to the arrival of the KCl-containing buffer to the flow cell, recording was started under continuous flow of buffer. Finally, three movies of different fields of view were recorded after equilibrium with the exchanged buffer was reached. The movies taken at equilibrium before and after the injection of 50 mM KCl were used to construct the distributions of iCy3 intensity in the two conditions. These distributions were fit with Gaussian peaks using OriginPro and the center of these peaks are reported. The movie recorded under flow was used to monitor the change of Cy3 fluorescence, in real time, as shown in the time traces. Similarly, to observe the melting of the secondary structure, the same P/T substrate was immobilized to the surface in RPA reaction buffer containing 50 mM KCl. Three movies were recorded, at equilibrium, before the injection of 100 nM RPA in the RPA reaction buffer. A movie was recorded, starting prior to RPA arrival to the flow cell and under continuous flow. At last, three movies were recorded after the final equilibrium with RPA was reached. These movies were used to construct the iCy3 intensity histograms and time traces, in a similar fashion to those described for the formation of the secondary structure.

Authors:
Fahad Rashid#, Vlad-Stefan Raducanu#, Manal S. Zaher#, Muhammad Tehseen, Satoshi Habuchi and Samir M. Hamdan.

# These authors contributed equally to the work.

Contributions:
F.R. and S.M.H. conceived the project. F.R., V.S.R., M.S.Z. and S.M.H. designed, the experiments. F.R., V.S.R., and M.S.Z. performed the experiments and analyzed the data. M.T. labeled and purified DNA oligos and prepared DNA substrates. F.R., V.S.R., M.S.Z., S.H. and S.M.H. discussed the results and wrote the manuscript.
Chapter 5

5. Discussion

5.1 Conformational dynamics and substrate recognition

Enzymes are critical for sustaining cellular life, as they allow biochemical reactions to happen in biologically relevant timescales. The role of conformational dynamics of enzymes and its correlation to catalysis have been extensively studied using all available experimental and computational methods (269). Conformational dynamics of enzymes have been postulated to play a key role in the recognition of the substrates itself. The classic “lock and key” model, of substrate recognition by proteins, attributed to Emil Fischer has been recognized to be inadequate, considering the wealth of information gleaned from various biophysical and biochemical methods since then (270). For the biochemical reactions which involve DNA as a substrate, one further dimension of complexity presents itself in form of potential inherent and induced substrate conformational dynamics and perturbations (271). The mechanisms and bases of these conformational changes in proteins and substrates during macromolecular interactions have been subject of intense debate. Two proposed mechanisms have been put forth, induced-fit model and conformational selection model, to explain the molecular interaction mechanism (108). Both these models are based on the idea that the biomolecule
can explore free energy landscape, which is manifested by distinct and multiple conformational states in equilibrium. The conformational selection model describes a scenario in which the biomolecule exists in a series of dynamically equilibrating states, one of which would be corresponding to the catalytically active state. The interacting biomolecule binds to the population that presents itself in the form of that active state. On the other hand, the induced-fit model proposes that the ligand binding shifts the conformational plasticity to the catalytically active state conformer. Both of these models need not to be mutually exclusive, as the overall recognition can be the product of different degrees of both models in different arenas of binding (108, 153, 154, 272).

In case of DNA processing enzymes, the different degrees to which each of these models play a role in DNA conformational changes and protein conformational changes are still subject to intense debate. Furthermore, whether these conformational changes in substrate and protein are coupled or whether these are independent processes has still to be resolved. The importance of resolving this question lies in the fact that the answer to this question has the potential to resolve the decades old question about the foundation of enzymatic fidelity (153, 203, 273). Additionally, if it is deciphered which recognition mechanism occurs during either of these pathways, synthetic chemicals can be designed that can lock the DNA in particular conformation to aid or inhibit the enzymes that act on these DNA intermediates.
5.2 FEN1 actively sculpts its cognate substrate

FEN1, like other members of the 5’ nuclease superfamily, provides a model system to understand the conformational changes in the substrate and the enzyme, since the DNA must be bound in a particular conformer and the enzyme must undergo through conformational transition to the ordered form (102, 136, 203, 274-276). The outcome of these conformational changes is the assembly of an active site that licenses catalysis in the case of cognate substrate. An interesting aspect of substrate recognition in 5’ nucleases is that they are sequence independent and are forced to rely on the physical attributes of the substrates rather than a direct physical readout of the DNA sequence (20). How this indirect recognition method influences the overall substrate-enzyme interaction as compared to the sequence-dependent direct readout has yet to be deciphered (136, 277). 5’ nucleases are not alone in using this indirect method of substrate recognition but it is common among other DNA processing enzymes. For example, enzymes tasked with finding damaged DNA in DNA repair, such as XPC in NER, rely on an indirect readout mechanism via “physical sensing” mechanism such as base flipping which has been studied extensively (278). The obvious question here is whether these damages present themselves as a deformity that is captured by the appropriate repair enzyme or the enzyme actively interrogates/sculpts the DNA as part of the recognition mechanism (108, 153, 272, 273, 279).

In the case of FEN1, we were able to deconvolute various transient and stable interactions that contribute to its substrate recognition mechanism. We
found that FEN1 subjects its substrate to a series of vetting processes using multiple checks by testing its various physical attributes (Chapter 2). The consequence of a successful vetting process is an ordered active site that promotes the transition state and catalysis. The penalty of failing any aspect of the vetting process is increasing the probability of being rejected by FEN1 without catalysis (Chapter 2). The first step in this vetting process is the superfamily-unifying DNA bending intermediary step. Threading the 5’ flap into the cap-helical gateway, active pulling of the 3’ flap into a pocket on the protein surface and the subsequent structuring of cap-helical gateway, all serve to increase the probability of forming a catalytically competent active site that cleaves the cognate substrate with high efficiency and extreme precision, while decreasing the probability of forming a catalytically-competent active site in non-cognate substrates to prevent its promiscuous cleavage (Chapter 2).

We unraveled that FEN1 actively bends the DNA by the direct visualization of the DNA bending process at the single molecule. Using single molecule FRET, with time resolution ranging from sub millisecond to hundreds of milliseconds, we showed that double flap (DF) substrate is essentially unbent and does not transition to a bent state without the presence of FEN1. Furthermore, we found that the rate of binding/bending to be $10^8$ M$^{-1}$s$^{-1}$, which is comparable to the expected diffusion limited bi-molecular association rates for the small DNA oligonucleotide substrate and 40 kDa enzyme (Chapter 2). We also showed that nick DNA adopt a single conformer and that it is actively bent by EXO1 in diffusion-limited association rates. Similar results of diffusion-limited association rates of
DNA binding/bending were calculated for IHF-DNA complex, where the authors concluded that DNA binding/bending proceeds in a stepwise mechanism and that DNA is actively being bent by IHF \((111)\). Other DNA bending enzymes including \textit{taq} MutS binding to single base mismatch has been shown to follow significantly lower association rates of \(-10^6 \text{ M}^{-1}\text{s}^{-1}\), suggesting that the initial interrogation and DNA bending are two decoupled processes in this overall recognition mechanism\((280)\). The active bending of DF substrate was further supported by molecular dynamics simulations that showed that significantly higher energy than that available at room temperature is required to transition the DF from extended to a \(90^0\)-bent state (Chapter 2).

Earlier work suggested that two nucleotide unpairing is a key step for substrate specificity \((29)\). However growing evidence from both structural as well biophysical work suggests that unpairing might not hold much significance in FEN1 substrate selection \((281)\). Rather, both EXO1 time resolved crystal structures and FEN1 threaded complex structure have shown that the scissile phosphate rolls over to the active site for cleavage to take place \((203, 281)\). In support of this, recent biophysical study using 2-AminoPurine (2-AP) have shown no significant perturbation in the two nucleotide regions on the 5’ side of the junction \((282)\).
5.3 5’ Flap threading as a part of an overall FEN1 vetting process

Continuing with further experiments, we sought to understand how FEN1 achieves its extraordinary precision. Our single molecule data showed that threading is indispensable for catalysis, not only because it provides a pathway to access the active site but also the binding of FEN1 to unthreadable flap substrate is inherently transient. This transient interaction precludes any real possibility of aberrant cleavage due to the clamping like condition (Chapter 2). Recent crystal structure of the threaded complex of FEN1, from our collaborator’s lab, provided a strong support for the importance of 5’ flap threading in accessing the active site and also a compelling elucidation of how FEN1 prohibits any inadvertent cleavage during the threading process. This remarkable feat was achieved by inversion of the phosphate backbone away from the active site metal ions during threading, which is accomplished by electrostatic steering, thus precluding inadvertent cleavage during threading process itself (281). This phosphate inversion is achieved by key basic residues (R103, R104, R129 and K132) that pilot the inverted flap through the gateway region and towards the active site. Mutations in any of these key residues were shown to have drastic effect on FEN1 nuclease activity and increase in trinucleotide repeat expansion (281). Time resolved crystal structure of EXO1 also showed a 5’ flap threaded complex with 2-nt 5’ flap. The authors here support a “fly-casting” model for flap threading which would limit the entropy associated with the DNA, thus proceeding to transition state along the way. The complex undergoes a series of interlocking transitions in the mobile arch region as it goes
from open, closed to clamped states (203). Our smFRET experiments also suggest that unthreaded FEN1 is a bit shy of reaching a fully bent state. This could be due to FEN1-flap substrate complex having significant degree of conformational flexibility, which hinders the complex from reaching a more clinched state. This clinched state is necessary to reach a final transition state (Chapter 2). Remarkably, the very recent structure of FEN1 mutants captured a pre-threaded complex of FEN1 and showed that the DNA was indeed not fully bent and most importantly positioned away from the gateway (283). This structure suggests that 5’ flap threading is required to roll the DNA closer to the active site and strengthen the interaction of FEN1 with the junction.

5.4 Coupling of substrate and enzyme conformational changes

Coupling enzymatic arch motions to the conformational changes in the DNA can provide the specificity by restricting the active site assembly for only cognate substrate. This mechanism of allosteric transition has been used to explain diverse molecular mechanisms that cannot be explained by classical ideas that restrict active site control to the catalytic process (284). In FEN1, it was hypothesized that cradling of a 1-nt 3’ flap into a binding pocket structures the cap helical gateway, thus equipping it for catalysis (28). Two main arguments support this hypothesis. First, biochemically it was shown that having this unpaired 1-nt 3’ flap augments the rate of forward reaction (44) and secondly, the 3’ flap binding pocket is disordered in the apo-enzyme and is ordered upon substrate binding (60, 285).
However, since FEN1 still cleaves single flap (SF) substrates bearing only a 5’ flap, it was unclear if the ordering of the protein absolutely requires this unpaired 3’ nucleotide. Our experiments show that FEN1 actively pulls this 1-nt 3’ flap, even in a SF substrate and in the equilibrated DF substrate (Chapter 2). This allosteric signal is further translated to the cap-helical gateway via R47 residue and other that link the two disordered segments of the protein (Chapter 2). This mechanism of “control at distance” of the active site is exquisite as it allows FEN1 to sense the 3’ flap and couple the ordering of the 3’ flap binding pocket with the downstream interaction of FEN1 with the DNA junction and the assembly of the active site. This allosteric coupling of the two disordered sites was further supported by a recent study that showed how different wedge residues including R47, L53, and T61 form a signal cascade that transduces the signal from 3’ flap ordering to cap-helical gateway ordering followed by assembly of the active site (286).

5.5 Avoiding off-target cleavage by probabilistic active site assembly

The other important aspect of any nuclease is its promiscuity. FEN1 is known to cleave several substrates that bear physical resemblance to the cognate DF substrate albeit with minor alterations compared to the cognate substrate (28). Binding efficiency based differences cannot fully explain the catalytic discrimination shown by FEN1 for these substrates. More importantly, the in-vivo concentration of FEN1 in the nucleus, where it is confined to a small nuclear
volume, would make its binding a non-limiting factor for its nuclease activity. To understand where this substrate discrimination stems from, we performed single molecule cleavage experiments, where we compared the FEN1 catalytic behavior on cognate and non-cognate substrates. We found that while FEN1 cleaves the cognate substrate in its first encounter, the cleavage of the non-cognate substrate was more probabilistic in nature, where the enzyme had to rely on multiple attempts to cleave the non-cognate substrate (Chapter 2). This method of avoiding off target catalysis by not only increasing the off-rate of FEN1 on the non-cognate substrate but also minimizing the chances of catalysis in these interactions is a potent way to increase the fidelity of the catalysis. A similar method of increasing the dissociation rate of off target substrates is used by the famous Cas9 nuclease in the case of mismatches in the proximal to protospacer-adjacent motif (287).

By comparing the dwell times of the missed cleavage events in non-cognate substrates and the successful cleavage event, we concluded that FEN1 suppresses the assembly of the active site by locking the protein and the DNA in noncatalytic conformers. This is in contrast to a model in which the protein and the DNA continuously sample conformations including the catalytically competent one, but the probability of finding the catalytic-competent one becomes largely reduced in non-cognate substrates. We proposed that this locking mechanism is mediated by the 3’ flap-induced protein ordering acting as a terminal step (Chapter 2). However, a recent structure of a pre-threaded 5’ flap complex shows that the 3’ flap engages its binding pocket and that the pocket is ordered while the cap-helical gateway is not (283). This is in line with our single molecule finding that showed
FEN1 can pull out the 3’ flap even when the 5’ flap threading is blocked. The ordering of the 3’ flap binding pocket but not the cap-helical gateway in the pre-threaded complex suggests that the signal between ordering the 3’ flap binding and the cap helical gateway could not be transduced in the mutant used to obtain the pre-threaded complex. Interestingly, in this complex, R47 engages the DNA, which is positioned away from the active site and is not fully bent (Chapter 2). This suggests a potential mechanism in which FEN1 marks its binding at the junction by first engaging its 3’ end in the 3’ flap binding pocket. R47 would act as a sensor that engages the DNA in the pre-threading state and switches its interactions to a region that lines up the interface between the 3’ flap binding pocket and the cap-helical gateway. This switching is triggered by 5’ flap threading, which helps FEN1 to pull the DNA closer to the active site and position the junction for its subsequent verification for catalysis. Therefore, the signal of ordering the 3’ flap binding pocket will be transferred to the cap-helical gateway only when the 5’ flap is threaded and the DNA moves closer to the active site. Using R47 as a switch may provide an elegant mechanism that will transfer the signal of ordering the 3’ flap binding pocket to the cap-helical gateway only when FEN1 engages the junction and threads the 5’ flap.
5.6 Backup pathway for FEN1 mediated Okazaki fragment maturation

Enzyme perfection is not the goal of evolution, for it if were the case, evolution would itself be impeded. Thus, enzymatic imperfection is the road to evolution. The consequences of these imperfections result in sometimes better traits than the original ones. For FEN1, we saw such imperfection presenting itself in form of missing to cleave the DF substrate as the flap length increases as shown in Chapter 3. The Okazaki fragment maturation process is usually carried out by the action of idling of Pol δ that generates the 5' flap, cleavage of the 5' flap by FEN1, and sealing the nick by DNA ligase 1, all these activities are coordinated by the sliding clamp PCNA. Defects in Okazaki fragment maturation have detrimental effects on genome stability (180). Nearly a quarter of each Okazaki fragment is synthesized by Pol α-primase complex, which is required to be completely removed since it contains initiator RNA primer and low fidelity DNA tracts. Additionally, unligated nicks can cause serious problems like double strand break. Moreover, if any flaps are generated but not properly processed, this could lead to duplications especially in repeats regions (208, 209, 274). These defects might interfere with DNA replication progression and in severe cases might signal cell death. Overall, they could contribute to mutator phenotypes and may act as hotspots for initiation and/or progression of cancer (180). In fact, deletion mutations of several of the primary enzymes in the Okazaki fragment maturation process
have been linked to cancer predisposition as well as other neurodegenerative diseases (25, 74, 210).

Yeast genetic studies using deletion mutation of Rad27 (homolog of FEN1) have shown that the flap could be as long as 100 nts (186). Furthermore, the existence of longer flaps has been recently supported by an electron microscopy study in vivo where the researchers observed flaps in wild type yeast cells with an average length of ~51 nts (184).

With the existence of these longer flaps supported, the next logical question is why would the cell opt for another pathway to deal with Okazaki fragment maturation? Despite the higher fidelity that the long-flap pathway would offer by removing a longer patch including not only the initiator RNA but also the DNA primer incorporated by the lower fidelity Pol α, this higher fidelity would come at a cost. The involvement of RPA and Dna2 (known as the long-flap processing pathway) might induce a delay to the processing of the Okazaki fragments on the lagging strand. But our current understanding is that this long flap pathway is only a backup system in rare cases. It has been hypothesized that the long flap pathway is deliberately induced in regions of active genes to insure their replication with high fidelity (181).
5.7 FEN1 binding and cleavage of long flap substrates

We set out to understand why and how this long flap pathway system works. We first started by questioning how would FEN1 escape the short flap to grow into a longer flap in the first place, given the tight coupling and the proposed highly efficient hand-off mechanism between FEN1 and Pol δ (145, 150). We questioned whether FEN1’s binding and/or substrate recognition could be the culprit. We found that substrate recognition, in particular bending efficiency, is not affected by flap length (Chapter 3). FEN1 binds and recognizes the longer flaps up to 60nt-long as efficient as it does on shorter flaps for all practical reasons (Chapter 3). In fact, this is supported by FEN1 crystal structure with flap substrates, where most interactions between FEN1 and its substrate are through the duplex DNA region, not the flap (51, 274).

If the flap length is not a challenge for FEN1 to properly recognize its substrate, we next asked if the longer flaps could affect its catalytic efficiency. Employing our unique smFRET cleavage assays based on internal-labeling scheme, combined with stop flow bulk cleavage assays, we examined this possibility using DFs with varying 5' flap-length. We showed that 5' flap does not influence the single turnover kinetics of FEN1. Since our internal label single molecule cleavage assay reports on bending, chemistry, flap release and the DNA unbending portion of the product release, we concluded that the flap length has no significant effect on any of these steps once the DF-substrate has been properly recognized (Chapter 3).
Yet, this assay gave us an insight to understand how FEN1 might escape its flap substrate. Examining the missed cleavage opportunities across the different flap-length substrates, we witnessed that as the flap length grows longer, FEN1 tends to miss the cleavage more (Chapter 3). Therefore, FEN1 binds and bends its substrate with equal opportunities regardless of the flap length, but seems to be challenged in assembling a catalytically competent active DNA-protein complex. The average lifetime spent in the bent conformer in these missed cleavage events falls in the same range as the single turnover kinetics for cleavage (Chapter 3). But with increasing flap length, FEN1 seems to have difficulty to position the substrate while threading the 5' flap into the cap-helical gateway and thus misses the cleavage (Chapter 3).

5.8 FEN1 missing cleavage prompts RPA-mediated second nuclease pathway

When FEN1 misses its cleavage, even once, it gives RPA, a strong ssDNA binder, the opportunity to bind the longer flaps and sequester those flaps from FEN1’s cleavage awaiting Dna2 involvement. On the longer flaps, RPA binding does not inhibit FEN1 accessing the substrate, but it inhibits the proper substrate recognition as we showed for DF-29,1 case (Chapter 3). In this regard, FEN1 and RPA competition to the longer flap is what determines which pathway, short- versus long-flap pathway, is involved. RPA binding to the long-flap substrates can then be viewed as the switch between the two pathways. Further characterizations of RPA
and FEN1 interplay on the short- versus long-flap substrates showed no effect on the short-flap substrates but a strong inhibition on the long-flap substrates. This strong inhibition was confirmed by bulk cleavage assays, FEN1 bending efficiency, and the inhibition of cleavage in smFRET cleavage assay (data not shown as cleavage events were scarce) (Chapter 3). On the short-flap DF-substrate, RPA exhibited no effect on FEN1 cleavage efficiency neither in bulk nor in single molecule assays (Chapter 3). It also had no effect on bending efficiency of the substrate or the product (Chapter 3). Furthermore, using smPIFE assay we uncovered a novel post cleavage binding step to unbent nick DNA product that may serve to relay the product to the appropriate downstream enzyme. Our results suggest a potential substrate handoff-mediated mechanism that coordinates the various steps involved in the maturation of Okazaki fragments. The ability of FEN1 to bind and bend the RPA-bound long-flap substrate might provide a mechanism that allows for an immediate substrate handoff to FEN1 after Dna2 minimizes the length of the 5’ flap and displaces RPA. The newly established two-step release of FEN1 from the nick DNA is also likely to be biologically relevant. The binding of FEN1 to the extended nick DNA product might allow FEN1 to hold on to the product for its handoff to the subsequent nick ligation step.

To sum up, in this study, we have shed some light on the kinetics of substrate recognition by FEN1, and how it might escape its flap substrate to grow longer. As the flap gets even longer, the probability of FEN1 escaping the substrate increases giving RPA the chance to get involved. Further work pinning down what specifically triggers FEN1 to dissociate from its substrate on the longer flap is still
a question to be addressed. Moreover, Dna2 involvement in the process, although biochemically addressed (185, 188), is not understood well. Our lab is currently in progress of understanding the involvement of Dna2 using single molecule imaging.

In future studies, the need to understand the structural basis of missed cleavage can provide much needed information about the FEN1 catalytic mechanism. Furthermore, current work in our lab is focused on using nuclear magnetic resonance (NMR) to understand the basis of disorder to order transition in FEN1 and to correlate the various time and structural dynamics in FEN1 with catalysis. Additionally, FEN1 is known to undergo several post-translational modifications that modulate its activity. However, the mechanism of modulation is unknown. We expect many of these post translational modifications effect to work by similar mechanism of decreasing the probability of cleavage. Our lab is interested in understanding how these post-translational modifications influence FEN1 cleavage kinetics and its interactions within the short- and long-flap pathways. Finally, extending this work to other structure-specific nucleases would decipher if the principles that control and guide FEN1 enzyme are universal to other structure-specific nucleases or each enzyme work by its unique molecular mechanism.
5.9 Use of single molecule fluorescence to answer biological questions

During my PhD thesis work, I stumbled upon a photophysics phenomenon where FEN1 quenches the Cy3 fluorescence that is placed at the tip of the 5' flap. Characterizing this quenching behavior has led to a comprehensive study that sheds light on the initial DNA-dye complex and determines the mode of fluorescence modulation in cyanine dyes to result in either enhancement or quenching. This part of the discussion is dedicated to this study.

The use of fluorescence in answering the biological question arises from their inherent sensitivity and specificity of detection (114). Fluorescence tools have been utilized to study biological questions both in ensemble assays and in single molecule techniques (288). Bulk assays have been used for many decades to understand basic biochemical properties of enzymes. However, bulk phase assays are blind to transient and fleeting interactions which are fundamental to all biochemical processes (129). Single molecule experiments obviate this issue by studying the dynamicity of individually resolved molecules and preclude any mass averaging (119, 129). smFRET in particular has improved our understanding of dynamicity in biological processes and has been critical to study biological processes at the single molecule in detail (289). However, the advantage of smFRET is lost when more than one dimensional vectorial distances are required to study the phenomena. The use of 3-color FRET to get more than one distance simultaneously is non-trivial due to the complexity of the assay and the limited
number of available stable spectrally different dyes (290). Furthermore, site-specific fluorescent labeling at three sites in a biomolecule is challenging. Understanding biological processes using fluorescence tools, whether at the single molecule level, or at the ensemble-level, may benefit from simpler approaches that do not require many extraneous labels.

5.10 Understanding fluorescence modulation of cyanine dyes

Increasingly, the modulation of fluorescence upon the binding of an interacting partner or conformational change within a single biomolecule is being leveraged to gain valuable understanding into the mechanisms of such processes (122, 291). In context of protein binding to dye-labeled DNA, the increase of fluorescence, termed protein-induced fluorescence enhancement, has been used extensively in both single molecule and ensemble assays (122). Furthermore, the modulation of fluorescence intensity within a single biomolecule has also been used to understand the conformational changes within single proteins such as GPCRS (262) where Cy3 showed two distinct fluorescence intensities depending on the conformation of the protein. The change in the protein conformation is expected to change the physical environment around the dye and consequently, its fluorescence parameters such as intensity change as well.

This phenomenal property of Cy3, and similar cyanine dyes, where fluorescence parameters are a function of the physical environment of the
fluorophore, has proven to be advantageous in the design of many experiments. Cy3 fluorescence parameters are mainly a product of cis-trans photoisomerization in the excited state. Within these bounds, one can imagine two different scenarios; either (1) Cy3 fluorescence is mainly the product of a rate competition between photoisomerization upon excitation and fluorescent de-excitation from the trans excited state, which would determine the fluorescence parameters of Cy3 in a particular context, or (2) Cy3 bound to a biomolecule will have a permanent interaction with neighboring residues (DNA bases and amino acids), which would lead to a more “rigidified” excited state. Both these scenarios do not need to be mutually exclusive and Cy3 fluorescence in any particular context (DNA sequence and protein residue interactions) can be a product of both the rate competition and rigidified excited state to a different extent. Within this paradigm, PIFE can be explained by these two hypotheses. In other words, the interaction of Cy3 with certain residues of amino acids can potentially freeze the Cy3 molecule in such a way that favors the trans excited state. Equally possible, the protein binding in the vicinity of Cy3 can change the local viscosity, thus leading to the alteration of the rate of photoisomerization from the trans excited state. Whereas the rigidified hypothesis allows for the existence of an inverse phenomenon where fluorescence modulation can occur in both directions, the rate competition hypothesis argues against the existence of such phenomenon and restricts the fluorescence modulation to increase only.
5.11 Characterization of protein-induced fluorescence modulation

Here, we showed with our FEN1/DF system that an opposite effect of PIFE i.e. protein-induced fluorescence quenching (PIFQ) mechanism does indeed exist. In this system, we witnessed both fluorescence enhancement upon substrate formation in the absence of any protein as well as subsequent FEN1-mediated fluorescence quenching (Chapter 4). Annealing of the oligos to make the substrate had dramatic influence on the fluorescence yield of Cy3 as quantified by both fluorescence lifetime and quantum yield. The experiments with Cy3B-labeled flap substrates strongly suggest that the fluorescence modulation is indeed mediated by the excited state cis-trans rotation of Cy3 (Chapter 4). The anti-correlated fluorescence enhancement upon substrate formation and subsequent quenching upon FEN1 binding clearly points out that quenching itself is due to the modulation of the inherently hyper-fluorescent state of Cy3 in the DF system. The anti-correlated Cy3 enhancement (upon substrate formation) and quenching (upon FEN1 binding) both depend on the linker length. These two directions of fluorescence modulations within the same system led us to hypothesize that both quenching and enhancement of fluorescence can be more generalized and controllable phenomena than previously perceived.
5.12 PIFE and PIFQ both depend on initial fluorescence state of Cy3-DNA complex

Within this thought process, we conjectured that the state of the initial fluorescence might be the key to generating either PIFE or PIFQ effects. Indeed, using a library of 5’ Cy3 labeled ssDNA oligos that exhibited varied initial fluorescence, we observed that both PIFE and PIFQ effects can be generated depending on the initial fluorescence of the Cy3-DNA (Chapter 4). We also witnessed few Cy3-labeled oligos that practically showed no effect. This is surprising in that it suggests that the initial fluorescence is determinant of PIFE and PIFQ effects, since the protein’s presence would increase the local viscosity in their case as well, leading to PIFE only. Indeed, the fluorescence change (PIFQ, PIFE) and the initial fluorescence are highly correlated (Chapter 4). Similarly, the 3’- and internal-labeled Cy3 ssDNA oligo libraries showed overall similar behaviors. In these cases, as well, the fluorescence modulation seems to be well correlated with the initial fluorescence of the system (Chapter 4). Taken altogether, the data suggests that RPA does not actively bring the Cy3 fluorescence in these ssDNA oligos to a similar outcome (fluorescence increase, decrease or no change), but rather the outcome itself is dependent on the initial fluorescence. Furthermore, the final state of Cy3 is conserved as depicted by the box plot of various positions of labeling. This raises the question: what is driving the final state in the presence of the protein to that level? The answer to such a question could rely on both the cis-trans
rotation around polymethine bond in the excited state of Cy3 as well as its modulation through interacting with nearby protein residues.

These results draw the attention to the use of PIFE (or PIFQ) as protein-binding assay especially concerning the association of PIFE (or PIFQ) with positive binding of the protein at a certain site. Special caution should be exercised with the interpretation of these fluorescent modulations. For instance, ssDNA oligos with lengths that support the binding of RPA but with sequences that had initial fluorescence around the average of the final Cy3-labeled DNA-RPA complex, did not show any significant change upon the addition of excess RPA. Hence, without a proper preliminary assessment, one could falsely attribute the lack of fluorescence change to poor (or even non-existent) affinity of RPA to such DNA sequences. In those scenarios, one could possibly change the sequence to generate a fluorescent modulation or amplify the existing one for appreciable change. Here, we must also stress that caution should be exercised with the interpretation in certain situations that would require unchanged quantum yield of either donor or acceptor or both as is the case when distinguishing simultaneous FRET change and PIFE/PIFQ. This can be achieved by changing the sequence (or linking chemistry, etc.) to show minimal or no effect whatsoever.
5.13 Characterization of initial fluorescence state of Cy3-DNA complex

Since the preceding RPA experiments implicated the initial state of a system in the modulation of fluorescence, we further investigated this initial state in the Cy3-DNA system without protein. Cy3 lifetime, in all cases whether free or DNA-conjugated in various configurations, has been shown to depend on viscosity change (233, 247, 249). Cy3B with its rigidified trans state is not expected to display any viscosity dependence for its fluorescence. Indeed, our fluorescence lifetime measurements of Cy3B do show stable fluorescence lifetime upon increasing the concentrations of glycerol up to 60% (Chapter 4); however, it begins to show some quenching effect upon further increase in the glycerol concentration. This can be due to some photophysical artifacts such as absorbance shielding. Two conclusions can be derived from these experiments. The first and more obvious outcome is that when Cy3 is more rigidly bound to the DNA, its lifetime displays less sensitivity to changes in the environmental conditions, thus exhibiting a less dynamic change in its fluorescence. The second and less obvious outcome is that the more rigidly bound Cy3 in a system is, the higher its initial lifetime (at 0% glycerol) is. This indicates that the rate of change of fluorescence upon viscosity increase is highly dependent on the initial lifetime. Furthermore, the cis-trans photoisomerization rate equilibrium does not change the bounds of fluorescence (Chapter 4), hence leading to the hypothesis that the fluorescence can be at least partially dependent on the rigidified excited state. Fluorescence would then depend on the
interconnected trans excited state photoisomerization rate and rigidification. Indeed, the relationship between $1/k_{1/2}$ (a measure of viscosity resistance and therefore of a fluorophore’s rigidification) and the initial lifetime for various Cy3 conditions (bound and free) showed a linear dependence which was rather unexpected (Chapter 4). The slope of such linear relationship can then be used as a quantitative parameter to evaluate any fluorophore’s propensity to change its fluorescence under the stress of environmental factors irrespective of the conjugation system or in its free form.

The fluorescent lifetimes of Cy3 generated under all the different conditions in this study were compiled and yielded the fluorescence landscape as shown in Figure 4.4.4 (Chapter 4). This comprehensive landscape provides a valuable insight into Cy3 fluorescent lifetime bounds that can be occupied by Cy3 under a diverse set of conditions. Free Cy3 and O-328 derivative (18 nt long) occupy the extremities of this curve. These upper and lower bounds exhibit ~15-fold difference in lifetime with a median of 1.35 ns for the distribution of all Cy3 lifetimes. Taken together, the viscosity dependence and the fluorescence landscape (Chapter 4), a crucial understanding of the photophysics of environmentally sensitive dyes can be gained, especially regarding the ease of modulating Cy3 fluorescence and the range of this modulation. We believe that this increased understanding of Cy3 photophysics can be easily transferred to other carbocyanine dyes.

The importance of such characterization of Cy3 fluorescence parameters and their modulation can be indispensable for the design of fluorescence-based experiments, whether in ensemble or at single molecule level. It also opens the
door to a new direction and dimension that, to the best of our knowledge, was never reported before. This study outlines the guidelines and describes the parameters that researchers can manipulate to achieve the desired fluorescence effect that serves their purpose. For instance, one would choose a condition that yields a fluorescence at the bottom of the lifetime landscape if the preferred outcome is to generate a PIFE effect. The same applies if PIFQ is sought, then one would prefer a condition where Cy3 fluorescence lies in the upper region of the landscape. Likewise, conditions yielding fluorescence lifetimes around the center of the landscape could be favored if the absence of any photophysical effect upon protein binding is required; for example, when the fluorophore is to be used for accurate FRET measurements to monitor conformational changes. Sequences of desired initial lifetime can be used from the library presented in the current work or from other sources such as high throughput assays. (245, 246, 259)).

O-328 showed extremely interesting aspect in that it rivalled or surpassed Cy3B in its lifetime (Chapter 4). We hypothesized that such state is generated by the rigidification of trans state possibly upon the formation of some secondary structure. Perturbing this secondary structure by shortening the oligo from the extremities resulted in gradual decrease in lifetime of Cy3 presumably due to the destabilization of the secondary structure. However, disrupting this structure by annealing the corresponding complementary oligo to the different O-328 derivatives displayed a striking consistent low lifetime across all lengths (22 to 8 bps) (Chapter 4). This further lead credence that in order for Cy3 to access such high fluorescence states, it has to have rigidified excited state in some scenarios
in addition to the inhibited trans excited state photoisomerization. It follows then that the high fluorescence in this oligo is indeed due to some secondary structure formation, and melting it decreases fluorescence to the median value of Cy3’s lifetime landscape (Chapter 4). Probing some factors that could result in the formation of this secondary structure demonstrated that the fluorescence of Cy3-labeled-O-328 is significantly sensitive to the presence of K+ ions reminiscent of G-Quadruplex structure. Remarkably, O-328 (18 nt) derivative showed the highest lifetime of any other Cy3-oligo which is 10% higher than Cy3B lifetime. We expect that this sequence can be integrated into other sequences to generate PIFQ.

5.14 Adaptation of fluorescence modulation at single molecule level

Using multiple single molecule experiments we showed how our understanding of the fluorescence modulations phenomena can be applied to rationally design PIFE/PIFQ experiments to study catalytic kinetics, conformational changes and protein binding at the single molecule level (Chapter 4). As a future direction of optimizing such a system we propose identifying patterns in labeled DNA sequences and correlating them with their corresponding structure and therefore lifetime. De novo in silico design and prediction of labeled DNA constructs lifetime can minimize at least the trial-and-error random process needed for obtaining a suitable initial state.
5.15 Concluding remarks

This thesis presents strong arguments for the case of FEN1 actively bending the DNA to a desired conformer. Further, we show that FEN1 also actively sculpts the other parts of DNA (3' flap and 5' flap) to its desired state. We also hypothesize that this sculpting process would allosterically order the protein, thus assembling active site that catalyzes the phosphodiester cleavage of DF substrate. Work in our lab to study this allosteric coupling of FEN1 protein ordering and substrate sculpting is continuing using NMR spectroscopy, which we expect will further decipher the basis and consequences of FEN1 ordering. We also show a novel way of substrate selection and discrimination by FEN1. Using a series of checks to vet the substrate, FEN1 cleaves the substrate if it passes all the requirements of cognate substrate but ejects it if it fails any of these checks. We also showed the basis and consequences of FEN1 missing the cleavage of its cognate substrate as the 5' flap length increases. This miss-cleavage provides the basis and origin of back up nucleases pathway of Dna2, if FEN1 fails to cleave the DF substrate. Lastly, we report the observation and characterization of a novel PIFE-related phenomenon, referred to as PIFQ, which is a missing key of the protein-induced modulation puzzle. This surprising observation led us to further investigate its mechanism and understand that the initial fluorescence state of the DNA-Dye complex plays a fundamental role in setting up the stage for the subsequent modulation by protein binding. Within this paradigm, we propose that enhancement and quenching of fluorescence upon protein binding are simply two different faces
of the same process. Furthermore, we achieved fluorescence modulation (PIFE/PIFQ) “on demand” through comprehensive characterization of interactions between dyes and DNA, dyes and protein, and DNA and protein. Based on this foundation, we demonstrated that our “on demand” PIFE/PIFQ could be used as a new single-molecule tool to characterize DNA-protein interactions. Our observations and correlations eliminate the current inconvenient arbitrary nature of fluorescence modulation experimental design. We anticipate that the combination of PIFE/PIFQ with smFRET will be key future single molecule assays in our lab in studying the mechanism of DNA binding proteins, including 5’ nucleases.
BIBLIOGRAPHY


36. Karanja KK & Livingston DM (2009) C-terminal flap endonuclease (rad27) mutations: lethal interactions with a DNA ligase I mutation (cdc9-p) and


185. Gloor JW, Balakrishnan L, Campbell JL, & Bambara RA (2012) Biochemical analyses indicate that binding and cleavage specificities define the ordered


the National Academy of Sciences of the United States of America 113(9):E1170-1179.


270. Lemieux RU & Spohr U (1994) How Emil Fischer was led to the lock and key concept for enzyme specificity. *Advances in carbohydrate chemistry and biochemistry* 50:1-20.


274. Tsutakawa SE, et al. (2017) Phosphate steering by Flap Endonuclease 1 promotes 5'-flap specificity and incision to prevent genome instability. 8:15855.


Statement of Author’s Contributions

Chapter 2


Authors:


# These authors contributed equally to the work.

Contributions:

**F.R.** and M.A.S. established the single-molecule TIRF experiments. F.R. designed and performed the single-molecule TIRF experiments and purified proteins. P.D.H., H.P., F.R. and S.H. designed the confocal FRET experiments. P.D.H. performed the confocal FRET experiments. M.A.S. performed the time-resolved bulk FRET experiments. M.Z. performed the EXO1 experiments, bulk cleavage assays and supported F.R. in optimizing and analyzing the single molecule TIRF experiments and protein purification. L.I.J. performed the SPR experiments. C.L. and I.I. performed MD simulations. F.R, S.E.T., J.A.T, S.H. and S.M.H. designed the study. F.R. and S.M.H. supervised the study and wrote the manuscript. F.R, S.E.T., J.A.T, S.H. and S.M.H edited, revised and proofread the manuscript. All authors analyzed the data, discussed the results and commented on the manuscript.
Chapter 3


Authors:

Manal S. Zaher, Fahad Rashid#, Bo Song#, Luay I. Joudeh, Mohamed A. Sobhy, Muhammad Tehseen, Manju M. Hingorani and Samir M Hamdan.

# These authors contributed equally to the work.

Contributions:

M.S.Z., F.R. and S.M.H. conceived and designed the experiments. M.S.Z and F.R. established single molecule assays. M.S.Z performed single molecule experiments and bulk cleavage assays in the presence of RPA. F.R. purified human FEN1. B.S., M.S.Z., M.M.H. and S.M.H. designed ensemble kinetic assays. B.S. performed ensemble kinetic assays. L.I.J, M.A.S, M.T. helped with some experiments. M.S.Z. and S.M.H. wrote the manuscript. M.S.Z., M.M.H. and S.M.H. edited, revised and proofread the manuscript.
Chapter 4


Authors:

Fahad Rashid#, Vlad-Stefan Raducanu#, Manal S. Zaher#, Muhammad Tehseen, Satoshi Habuchi, and Samir M. Hamdan.

# These authors contributed equally to the work.

Contributions:

F.R. and S.M.H. conceived the project. F.R., V.S.R., M.S.Z. and S.M.H. designed, the experiments. F.R., V.S.R., and M.S.Z. performed the experiments and analyzed the data. M.T. labeled and purified DNA oligos and prepared DNA substrates. F.R., V.S.R., M.S.Z., S.H. and S.M.H. discussed the results and wrote the manuscript.
Accomplishments

List of publications: