University of Sheffield

Thesis submitted for the degree of Doctor of Philosophy to the

Department of Chemistry, University of Sheffield

Kinetic and biophysical characterisation of 5’-flap nucleases in DNA maintenance and repair

Supervisor: Professor Jane Grasby

Reuben Joseph Ouanounou

August 2023

Declaration

Except where specific references have been made to other sources, the work in this thesis is the original work of the author, and has not been submitted, wholly, or in part for any other degree.

Mutagenesis, molecular cloning, and production of the D34N mutant construct was performed by Dr David Finger and Dr Sana Algasaier. Additionally, Mass spectrometry analysis of the FAN1 oligonucleotides was performed by Professor Mark Dickman.

Reuben Joseph Ouanounou

Acknowledgements

Science has always offered me salvation from adversity, feelings of self-doubt and other ill-conduced internal conflicts. Therefore, pursuing a PhD felt only natural. PhD's may be awarded to an individual, but they are earned as a group. Because of this, there are many people I wish to thank for their never-wavering love and support during my time in Sheffield. Firstly, I would like to thank my supervisor Professor Jane Grasby for her support and guidance throughout the last four years, ultimately moulding me to become a better scientist. A name who can't be forgotten is Dr David Finger, my lab mentor and friend. A man with unparalleled knowledge of biochemistry, and the singing voice of an angel. An outside group member is Dr James Wilson, who provided mentoring in shaping my career, and taking on the painstaking task of proofreading my cover letters. For that, I apologise James…

I'm indebted to my friendship group, where shenanigans were never in short supply. People including Jonny, Simon, Neil, Fourat, Jenny, Tom, Anton, Millie, and the unequivocally apologetic Josie, never failed to bring excitement and laughter to my long work hours, even if I was regrettably the butt of the jokes on several occasions. Other special mentions are Freya Cleasby and Courtney Thompson. From day one, Freya has supported me through my highest achievements, and darkest hours. For that, Freya is someone I'll hold in the highest regard, as a scientist and as a person. Words can't begin to describe the bond between Courtney and I. I've shared some of the best moments of my life with Courtney, from picturesque weekend brunches to gruelling workout routines, we are inseparable. Courtney tirelessly finds the best in everyone and everything. Courtney, if you're reading this, ''Who's gonna carry the boats!?''. This section isn't complete without my best friend Mark Gainford, the human personification of emotional resilience. The obstacles we've overcome together have sculpted us to never shy away from failure. Mark, you're going to achieve great things in this world, and I can't wait to celebrate them with you. Lastly out of my friends is Dr Jim Scotson, although I consider him more of a brother than a friend. The conversations we shared over the years, followed by the wisdom I've gained from your invaluable life lessons have shaped me to become the person I am today, someone proud, hardworking, and respectful of others. Thank you, Jim, I'll always be there for you as you have been for me.

Finally, I would like to thank my family for supporting me throughout my entire life. For people who don't know my story, my family have experienced great loss and sacrifice. However, we do not mourn the loss of loved ones, we remember them, so that our own lives may prosper. For anyone reading this, follow your dreams, conquer your fears, but most of all, just be you.

אני מבקש לא משאות קלות יותר,

אלא כתפיים רחבות יותר

''I ask not for lighter burdens, but for broader shoulders.''

– Jewish proverb

If someone says ''I have worked hard, and I have not been successful'', don't believe him.

If someone says ''I have not worked hard and I have been successful'', don't believe him.

If someone says ''I have worked hard, and I have been successful'', believe him!

– The Talmud, Megilla, 6b

Abstract

Genome fidelity is reliant on structure-specific nucleases to make accurate and reproducible incisions into DNA intermediate structures during replication and repair processes. Organisms lacking structure-specific nuclease activity are increasingly susceptible to genetic disease. Flap endonuclease 1 (FEN1) and Fanconi anaemia nuclease 1 (FAN1) are structure specific 5'-nucleases with preference for branched DNA structures, hydrolysing the 5'-flap with metal ion dependency.

FEN1 is a well characterised nuclease with preference for substrates bearing a 1-nucleotide 3'-flap and a varying length 5'-flap, subsequently hydrolysing the phosphodiester bond 1-nucleotide into the 5' duplex of the substrate, referred to as the +1-1 position. Reproducible hydrolysis at the +1-1 position is crucial during Okazaki fragment processing and long-patch base excision repair (LP-NER). In doing so, it creates a ligatable product which is a substrate for DNA ligase. Protein and DNA conformational changes drive FEN1-mediated catalysis and specificity. 3'-flap binding promotes ordering of the helical arch which encompasses the threaded 5'-flap before catalysis. Additionally, DNA conformational changes at the +1-1 position are also required to position the scissile phosphate within attacking distance of the catalytic metal ions, known as active site transfer. A recent report has suggested that the mutant FEN1 D34A is rate-limited under excess enzyme conditions by the DNA (and probably protein) conformational change. These experiments used a +1-1 tandem 2-aminopurine (2-AP) reporter assay, describing a burst phase that precedes the +1-1 hydrolysis, although this remains untested by other research groups. Furthermore, limited information is available surrounding arginine and lysine residues (R239, K244 and R245 in human FEN1) in the HELIX-2-TURN-HELIX (H2TH) DNA binding domain of FEN1 with regards to changes in kinetic rates and conformational change at the +1-1/-1-2 positions. To answer these questions, a 2-AP reporter assay with the less severe D34N mutant was performed but presented no evidence of a pre-cleavage burst phase, providing alternative findings to *Song et al*. Furthermore, kinetic and exciton-coupled circular dichroism (ECCD) assays using 5' fluorescein (FAM) and 2-AP fluorescent labelled substrates respectively showed a modest reduction in multiple/single turnover rates and change in ECCD signal in FEN1 mutants R239A, K244A and R245A, implying these residues do not significantly alter FEN1 catalysis or active site transfer.

FAN1 is a less well-characterised nuclease and is linked to interstrand crosslink (ICL) repair in Fanconi anaemia (FA) by means of the Fanconi anaemia repair pathway (FA-pathway). Although human and bacterial homolog structures are currently available, a considerable number of questions remain unanswered concerning FAN1 mechanism. Reports have demonstrated FAN1's ability to catalyse the reactions of a plethora of DNA substrates, showing preference for double-flapped substrates bearing a 5'-phosphorylated, 1-nucleotide 5'-flap and an 8-nucleotide 3'-flap (DF(p1,8)). Hydrolysis of this substrate involves an N+3 incision model from the 5'-flap terminal phosphate however the pattern alters depending on 5'-flap length. Despite this, in-depth kinetic parameters for any FAN1 substrates have not yet been defined, raising questions as to the true FAN1 mechanism during 5'-flap processing. In this work, human FAN1 production was optimised and nuclease assays were performed on various 5'-phopshorylated substrates. Our nuclease assays show concordance to previous studies of an observed N+3 incision model for the DF(p1,8)3'FAM with multiple products produced. However, only 1 product was produced for short double flapped substrates and substrates lacking a phosphorylated 5'-flap. Finally, the rate of FAN1 catalysis is hindered in the presence of a 5'FAM label compared to 3'FAM labelled substrates, implying the need to explore other methods of product measurement for FAN1-DNA reactions outside dHPLC for kinetic characterisation.

Table of contents

Declaration **1**

Acknowledgements **2**

Abstract **4**

Table of contents **6**

List of abbreviations **9**

Chapter 1 Introduction **11**

1.1 Overview of DNA structure and function 11

1.2 DNA nuclease function and mechanism 13

1.3 Structure-specific nucleases and the Rad2/XPG (FEN) superfamily 14

1.4 Introduction to FEN1 and its importance in genome stability 16

1.4.1 The role of FEN1 in DNA replication 16

1.4.2 The role of FEN1 in DNA repair 20

1.5 FEN1 Structure 22

1.5.1 The hydrophobic wedge 23

1.5.2 The 3'-flap binding site 23

1.5.3 The helical arch and cap 24

1.5.4 The H2TH:K^+ binding site 25

1.5.5 The active site 25

1.6 Proposed models of FEN1-mediated catalysis 27

1.6.1 Track and clamp model 27

1.6.2 Threading model 27

1.6.3 Base unpairing and 5'-flap nucleotide flipping model 28

1.6.4 Loop-wedge allosteric binding model 29

1.7 Conformational changes required for FEN1-mediated catalysis 31

1.7.1 5'- and 3'-duplex bending 31

1.7.2 Recognition of the 3'-flap results in conformational ordering of the helical arch 35

1.7.3 Active site transfer of the scissile phosphate towards the catalytic metals 38

1.8 Introduction to Fanconi anaemia nuclease 1 40

1.8.1 The role of FAN1 in the FA dependent and FA independent pathways 40

1.8.2 Disease association to FAN1 deficiency 44

1.8.3 FAN1 substrate specificity 45

1.8.4 FAN1 structure 48

1.8.4.1 N-terminal helical bundle and the SAP domain 49

1.8.4.2 C-terminal basic binding pocket 51

1.8.4.3 Active site 53

1.8.5 Proposed models of dimeric FAN1 54

1.8.6 Comparison of FEN1 and FAN1 properties 57

1.9 Aims of the project 58

Chapter 2 Materials and Methods **60**

2.1 WT & mutant FEN1 construct cloning and protein expression 60

2.1.1 Site directed mutagenesis & PCR amplification 61

2.1.2 Bacterial cell transformation 63

2.1.3 Plasmid preparation for sequencing 63

2.1.4 Protein expression and purification of WT-FEN1, R239A, K244A & R245A 64
2.2 Kinetic assays of WT-FEN1, R239A, K244A and R245A 67
2.2.1 Kinetic assay substrates 67
2.2.2 Multiple turnover nuclease activity kinetic assay of FEN1 68
2.2.3 Single turnover kinetic experiment of WT/mutant FEN1 and DF(p5,1)5'FAM 70
2.3 Measuring conformational change in FEN1 by 2-aminopurine substitution at the +1-1 position of the hydrolysed duplex. 71
2.3.1 ECCD assay substrates 71
2.3.2 ECCD assay of WT and mutant FEN1 72
2.4 2-AP fluorescence assay of the FEN1 mutant D34N and DF(14,1)$_{+1-1}$PP 73
2.4.1 2-AP fluorescence assay 73
2.5 FAN1-NusA construct cloning and protein expression 74
2.5.1 PCR amplification 74
2.5.2 Cell transformation and plasmid preparation for sequencing 74
2.6 Purification of FAN1-NusA 74
2.6.1 Small-scale purification of FAN1-NusA 75
2.6.2 Large-scale expression and purification of FAN1-NusA 75
2.6 FAN1-SUMO/ FAN1-MBP construct cloning and FAN1-MBP protein expression 75
2.7.1 PCR amplification 75
2.7.2 Cell transformation and plasmid preparation for sequencing 78
2.8 Protein expression and Purification of FAN1-MBP 78
2.8.1 Small-scale purification of FAN1-SUMO and FAN1-MBP 78
2.8.2 Large-scale Purification of FAN-MBP 78
2.8.3 Revised method for the purification of FAN-MBP after differential scanning fluorimetry 80
2.9 Differential scanning fluorimetry screening 81
2.9.1 Screening for optimum enzyme and SYPRO-Orange dye concentration 81
2.9.2 Commercially bought screens and in-house produced screens 82
2.10 FAN1 oligo production and kinetic assays 83
2.10.1 DMT on production of FAN1 oligos 83
2.10.2 FAN1 nuclease assays 85
2.10.2.1 Denaturing UREA PAGE analysis of quenched FAN1 reactions. 87
2.10.2.2 RP-dHPLC analysis of FAN1-SB5,1HT2 reactions 88
Chapter 3 2-aminopurine fluorimetry assay measuring changes in D34N conformational changes and reaction rate 89
3.1 Introduction 89
3.2 Measuring the rate of 2-AP fluorescence change in D34N with DF(14,1)$_{+1-1}$PP 95
3.2.1 D34N and WT 2-AP assay 95
3.3 Summary 100
Chapter 4 Kinetic and biophysical characterisation of WT FEN1 and H2TH mutants 101
4.1 Introduction 101
4.2 Site directed mutagenesis of and construct production of R239A, K244A and R245A 106
4.3 Expression and purification of WT, R239A, K244A & R245A 106
4.4 Kinetic characterisation of WT, R239A, K244A and R245A 108
4.4.1 Characterisation of WT and H2TH mutants under steady-state conditions 108
4.4.2 Characterisation of WT and H2TH mutants under single turnover conditions 112
4.5 Measuring the effects of the H2TH mutants on AST using ECCD and 2AP labelled substrates 115
4.5.1 ECCD experiments with DF(5,1)$_{+1-1}$PP and DF(5,1)$_{-1-2}$PP 115

4.6 Summary 119
Chapter 5 Expression and purification of FAN1 120
5.1 Introduction 120
5.2 PCR amplification and plasmid preparation 124
5.2.1 Production of the FAN1-NusA construct 124
5.2.2 Production of the FAN1-SUMO and FAN1-MBP constructs 125
5.3 FAN1 production 129
5.3.1 Small-scale expression and purification of NusA, SUMO and MBP FAN1 constructs. 129
5.3.2 Large-scale expression and purification of FAN1-NusA 131
5.3.3 Large-scale expression and purification of the FAN1-MBP construct 134
5.4 DSF screening 137
5.4.1 DSF screening to identify buffer and additive conditions. 138
5.5 Revised purification strategy for FAN1-MBP with improved buffer conditions 142
5.5.1 Refined purification of FAN1-MBP 143
5.6 Summary 145
Chapter 6 Preliminary biophysical characterisation of FAN1 146
5.1 Introduction 146
5.2 FAN1 nuclease assays and HPLC method development 149
5.2.1 FAN1 nuclease assay with SB5,1,HT2 149
..... 150
5.2.2 FAN1 nuclease assays with DF(p1,8)3'FAM and DF(p1,8)5'FAM 151
5.2.3 FAN1 nuclease assays with DF(p1,1)3/5'FAM, SF(p0,1)3/5'FAM and SF(p0,8)3/5'FAM substrates 157
5.2.4 Reverse-phase ion pair HPLC gradient screening of 3'FAM PB1 and DF (1,8) 3'FAM samples 160
5.2 Summary 163
Chapter 7 Conclusions and Future work 164
7.1 Kinetic and biophysical studies of FEN1 active site and H2TH mutants 164
7.1.1 Chapter 3 conclusions and Future work 164
7.1.2 Chapter 4 conclusions and future work 165
7.2 Production and preliminary characterisation of FAN1 168
7.2.1 Chapter 5 and chapter 6 conclusions and future work 168
References 171
Chapter 8 Appendices 186
8.1 pET28b WT human FEN1 plasmid 186
8.2 Agarose gel electrophoresis of FEN1 and FAN1 constructs 186
8.3 WT FEN1 and H2TH mutant Michaelis Menton graphs 187
8.4 FEN1 and FAN1 sequencing 187
8.4.1 H2TH FEN1 mutant construct sequencing 187
8.4.2 FAN1-SUMO construct sequencing 200
8.4.3 FAN1-MBP construct sequencing 208
8.5 Quench flow kinetics table 214
8.6 Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (FTMS) of FAN1 oligonucleotides 215

List of abbreviations

Abbreviation	definition
DSF	Differential scanning fluorimetry
[E]	Enzyme concentration
[S]	Substrate concentration
+1-1	+1-1 respective bases to the scissile phosphate of FEN1
-1-2	-1-2 respective bases to the scissile phosphate of FEN1
2-AP	2-aminopurine
2-BME	2-β-mecaptoethanol
2XYT	2 × yeast/tryptone media
AIC	Akaike information criterion
ATR	Ataxia-telangiectasia and Rad3-related kinase pathway
BER	Base excision repair
BP	Base pair
BTP	Bis-tris propane
CV	Column volume
DEA	Diethanolamine
DMT	Dimethoxytrityl
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
dsDNA	Double stranded Deoxyribonucleic acid
DSF	Differential scanning fluorimetry
DTT	Dithiothreitol
E	Enzyme
ECCD	Exciton-coupled circular dichroism
EDTA	Ethylenediaminetetraacetic acid
EXO1	Exonuclease 1
FAM	Fluorescein
FAN1	Fanconi anaemia nuclease 1
FB	Folding buffer
FEN	Flap endonuclease
FEN1	Flap endonuclease 1
FPLC	Fast protein liquid chromatography
FRET	Fluorescence resonance energy transfer
GEN1	Gap endonuclease 1
GFP	Green fluorescent protein
H2TH	HELIX-2-TURN-HELIX
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HIC	Hydrophobic interaction chromatography
HIS-TAG	Histidine tag
HMT	Histidine-tag/Maltose binding protein/TEV cleavage site
ICL	Interstrand crosslink
IMAC	Immobilised metal affinity chromatography
IPTG	Isopropyl ß-D-1-thiogalactopyranoside
k_{cat}	Turnover number
k_{cat}/K_M	Catalytic efficiency
$k_{distort}$	Rate of DNA distortion
K_M	Michaelis-Menten constant
$k_{on\text{-}bend}$	Rate of DNA bending

k_{STmax}	Maximal single turnover rate
Lb	Lysogeny broth
LP-NER	Long patch base excision repair
MBP	Maltose binding protein
MeCN	Acetonitrile
MMS	Methyl methanesulfonate
MNNG	N-Methyl-N'-nitro-N-nitrosoguanidine
mRNA	Messenger ribonucleic acid
NL	No FAM label
NusA	N utilisation substance A
OD_{600}	Optical density at 600 nm
PaFAN1	*Pseudomonas aeruginosa* FAN1
PBS	Phosphate buffered saline
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
pET-HMT	pET vector containing a His-tag, maltose binding protein and TEV cleavage site
PIPES	piperazine-N,N'-bis(2-ethanesulfonic acid)
RB	Reaction buffer
RPA	Replication protein A
RP-dHPLC	Reverse-phase High performance liquid chromatography
RPM	Rotations per minute
RRB	Reduced reaction buffer
RT-PCR	Real-time polymerase chain reaction
S	Substrate
SDS-PAGE	Sodium dodecyl-sulfate polyacrylamide gel electrophoresis
smFRET	Single molecule fluorescence resonance energy transfer
SOC	Super optimal culture broth
SP-BER	Short-patch base excision repair
ssDNA	Single-stranded deoxyribonucleic acid
SUMO	small ubiquitin-related modifier
TAE	Tris-acetate-EDTA
TBAB	Tetrabutylammonium bromide
TBE	Tris/Borate/EDTA
TEAA	Triethylammonium acetate
TEMED	Tetramethylethylenediamine
TEV	Tobacco Etch Virus
T_M	Half maximal melting temperature
TRIS	tris(hydroxymethyl)aminomethane
TY	Bacterial growth media
UREA	Carbamide/UREA
VvFAN1	Vibrio vulnificus FAN1
WT	Wild type
XPG	Xeroderma pigmentosum group G
v/[E]	Rate normalised for enzyme concentration

Chapter 1 Introduction

1.1 Overview of DNA structure and function

Deoxyribonucleic acid (DNA) is the molecular storage system of genetic information throughout all cellular life. The absence of DNA would cease all complex life due to its critical role in encoding and storing the genetic information of host organisms. Eukaryotic DNA is efficiently packaged into chromatin and forms macromolecular chromosomes within the nucleus of the cell [1]. The structural properties of DNA contribute substantially to its function in storing genetic information; DNA is formed of individual nucleotides which consist of a phosphodiester backbone; a 2-deoxyribose sugar and one of the four purine or pyrimidine nitrogen bases (adenine, thymine, cytosine, and guanine). However during DNA transcription uracil replaces thymine when messenger RNA (mRNA) is being formed (Figure 1.1A) [2].

The four variations of nucleotide come together to form a DNA double-helix and hydrogen bonds are formed between the complementary bases A-T and G-C. DNA is not linear in directionality, instead, both strands proceed in opposite directions in an antiparallel configuration, resulting in each end of the duplex possessing a 5' phosphate and a 3' hydroxyl functional group in double-stranded, genomic DNA (Figure 1.1B) [3]. The most common conformational form of DNA found in living cells is the right-handed B-DNA structure proposed by the Watson and Crick model, resulting in major and minor grooves observed within the double helix (Figure 1.1C). A-DNA structures are shorter in length and right-handed whereas Z-DNA is left-handed [4]. Although hydrogen bonding occurs between the bases, these bonds are strategically weak to allow routine separation of the two strands during DNA replication, DNA transcription, and various repair cascades. Further stability is afforded to the structure via the twisting and base stacking of the DNA duplex. In addition to this, the hydrophilic phosphodiester backbone protrudes out from the double helix and is subsequently solvated by water molecules [5].

Despite evolutionary efforts, DNA remains a fragile biomolecule, constantly maintained and repaired after repetitive endogenous damage including oxidation, methylation and deamination of the DNA bases. Exogenous DNA damage can occur through inter-strand crosslinks, subsequently forming covalent bonds between the nitrogen bases and thus, inhibiting separation of the two strands during replication [6]–[8].

Therefore, an array of substrate and sequence-specific enzymes are employed to incise, modify, elongate, and finally seal DNA strands to maintain high fidelity of the genome in the form of nucleases, polymerases, and ligases respectively.

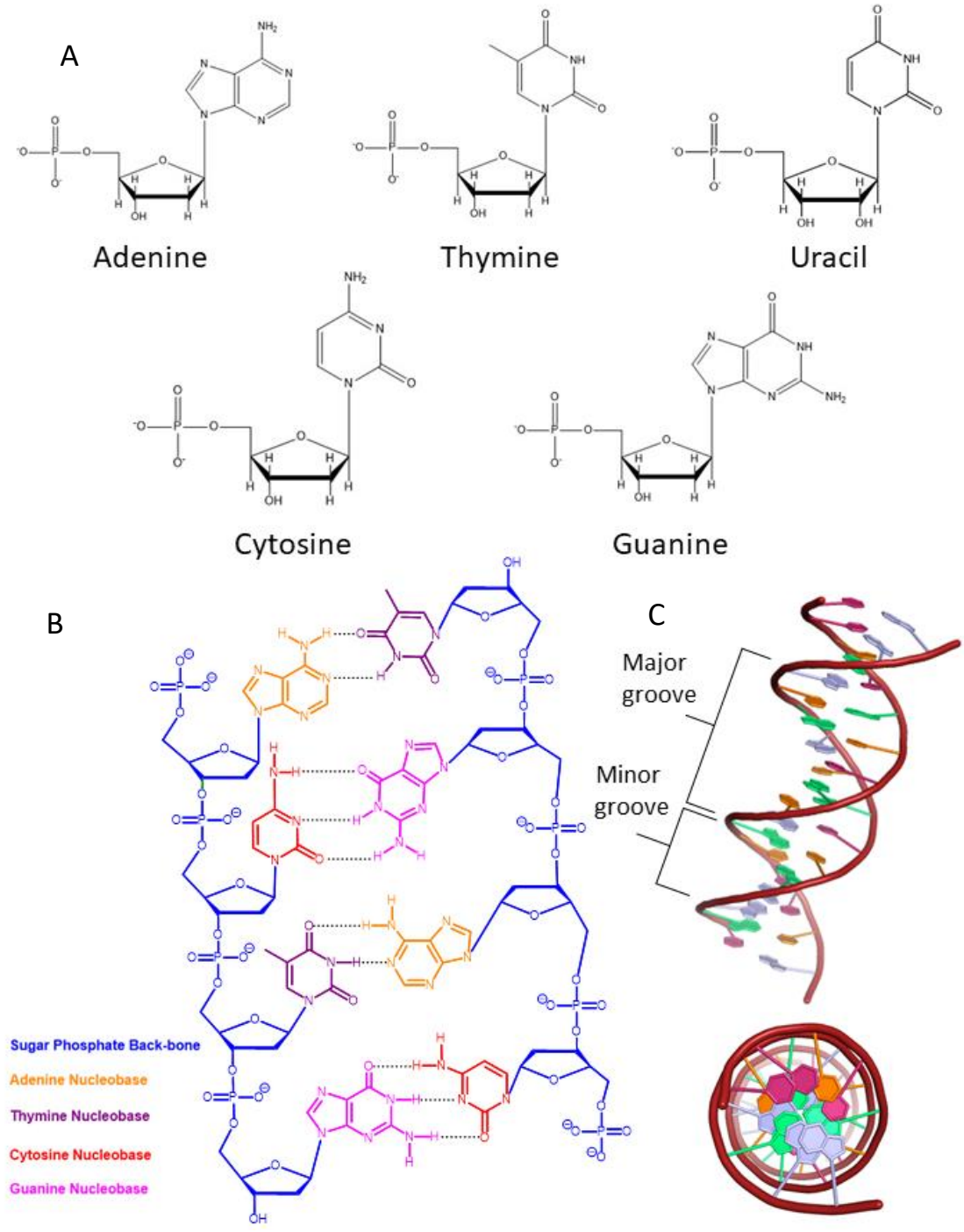


Figure 1.1 Structure of DNA

***A**. Structure of adenine, thymine, uracil, cytosine and guanine nucleotides. **B**. Colour coded structure of two anti-parallel DNA oligonucleotide chains. **C**. Structure of B-DNA illustrating the major and minor grooves and inward-facing bases (PDB code: 5FKW).*

1.2 DNA nuclease function and mechanism

Nucleases are one of the most indispensable components regarding safeguarding, and maintenance of the genome. Nucleases act as 'molecular scissors', hydrolysing the phosphodiester bonds between nucleic acids in the DNA duplex. This is routinely observed in biological processes including DNA replication, DNA repair and DNA proofreading. DNA cleavage can occur within, or at the end of the DNA strands, therefore dividing nucleases into two major categories, endonucleases and exonucleases respectively. Moreover, exonucleases are further divided into the subgroups 5' and 3', primarily depending on which polarity of the DNA strand the nuclease shows increased specificity for catalysis [9]. Nuclease activity is required because the uncatalysed phosphodiester bond in the presence of water is notoriously unreactive, where water acts as the attacking nucleophile. The rate of hydrolysis of phosphodiester bonds at 25 °C at neutral pH was predicted to be $10^{-15}s^{-1}$, resulting in a half-life of approximately 30,000,000 years [10]. Under these parameters, living organisms cannot survive and thus, nucleases are employed to lower the activation energy of the reaction and to physically increase the proximity of the two reactants, resulting in increased reaction rate.

During phosphodiester hydrolysis, activated nucleophiles attack the electrophilic phosphorus resulting in a penta-covalent intermediate or transition state, stabilised by positively charged metal ions/amino acids commonly observed in several nucleases (Figure 1.2). The final stages involve production of a tetrahedral monoester and release of the hydroxyl 3' end [10], [11]. During phosphodiesterase reactions, two divalent metal ions are often incorporated to catalyse incision of the phosphodiester bond in the form of Mg^{2+} or Mn^{2+} commonly denoted as the two-metal ion mechanism. During this process, both metal ions bind the same non-bridging oxygen, however one metal ion binds the nucleophilic hydroxyl group and the other metal ion binds the leaving group oxygen [12], [13].

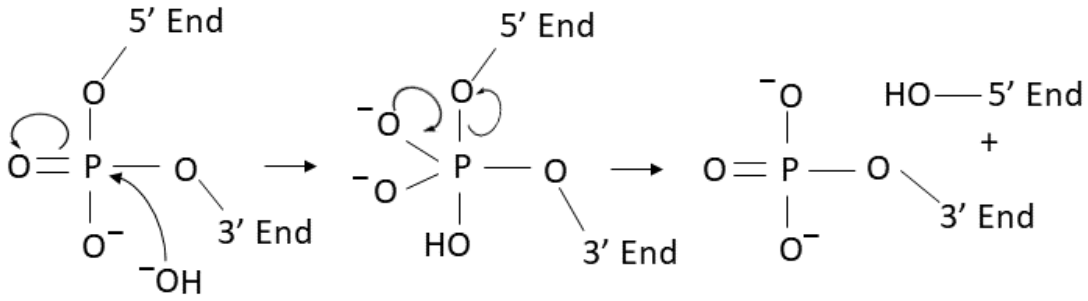


Figure 1.2 Phosphodiester bond hydrolysis of DNA

Phosphodiesterases are employed to facilitate phosphodiester bond hydrolysis and often utilise metal ions to catalyse the reaction. The activated nucleophile (⁻OH) donates electrons to the phosphorus atom (left) resulting in a penta-covalent intermediate (middle). Electron transfer reconfigures the phosphorus-oxygen double bond whilst producing a phosphate monoester and a 3'-hydroxyl product (right).

1.3 Structure-specific nucleases and the Rad2/XPG (FEN) superfamily

Substrate specificity is a major factor in nuclease function throughout genome maintenance and repair. Nucleases can recognise substrates through sequence, or structure. Sequence dependent nucleases such as Bsa1, DPN1, Ssp1 and nucleases in the Crispr/Cas9 system perform incisions dependent on nucleotide sequence [14]. However, genome stability is also reliant on structure-specific nucleases as they recognise intermediate substrate structures during DNA maintenance and repair. A key example of structure specific nucleases are those found in the FEN nuclease superfamily which consists of 6 enzymes, four of which are DNA nucleases comprising of Flap Endonuclease 1 (FEN1), Exonuclease 1 (EXO1), Gap Endonuclease 1 (GEN1) and Xeroderma pigmentosum group G endonuclease (XPG); the remaining enzymes XRN1 and XRN2 are ribonucleases involved in RNA processing. Interestingly, members of the FEN superfamily display starkly different substrate specificities despite high domain conservation. FEN1 binds double flapped substrates and hydrolyses the 5'-flap during Okazaki fragment processing and base excision repair (Figure 1.3A) [15]. EXO1 is a 5-3' exonuclease and functions to resect substrates during the DNA mismatch repair pathway. EXO1 also produces resected ends during homologous recombination of double strand breaks (Figure 1.3B) [16], [17]. GEN1 is predominantly acts in the meiosis phase of the cell division cycle. Functioning as a homodimer, GEN1 hydrolyses Holliday junctions produced during chromosomal crossover (Figure 1.3C) [18]. Finally, XPG collaborates with the ERCC1/XPF nuclease complex to remove bubble DNA structures. XPG performs 5' incision into the duplex as a result of ultraviolet light exposure during nucleotide excision repair (Figure 1.3D) [19].

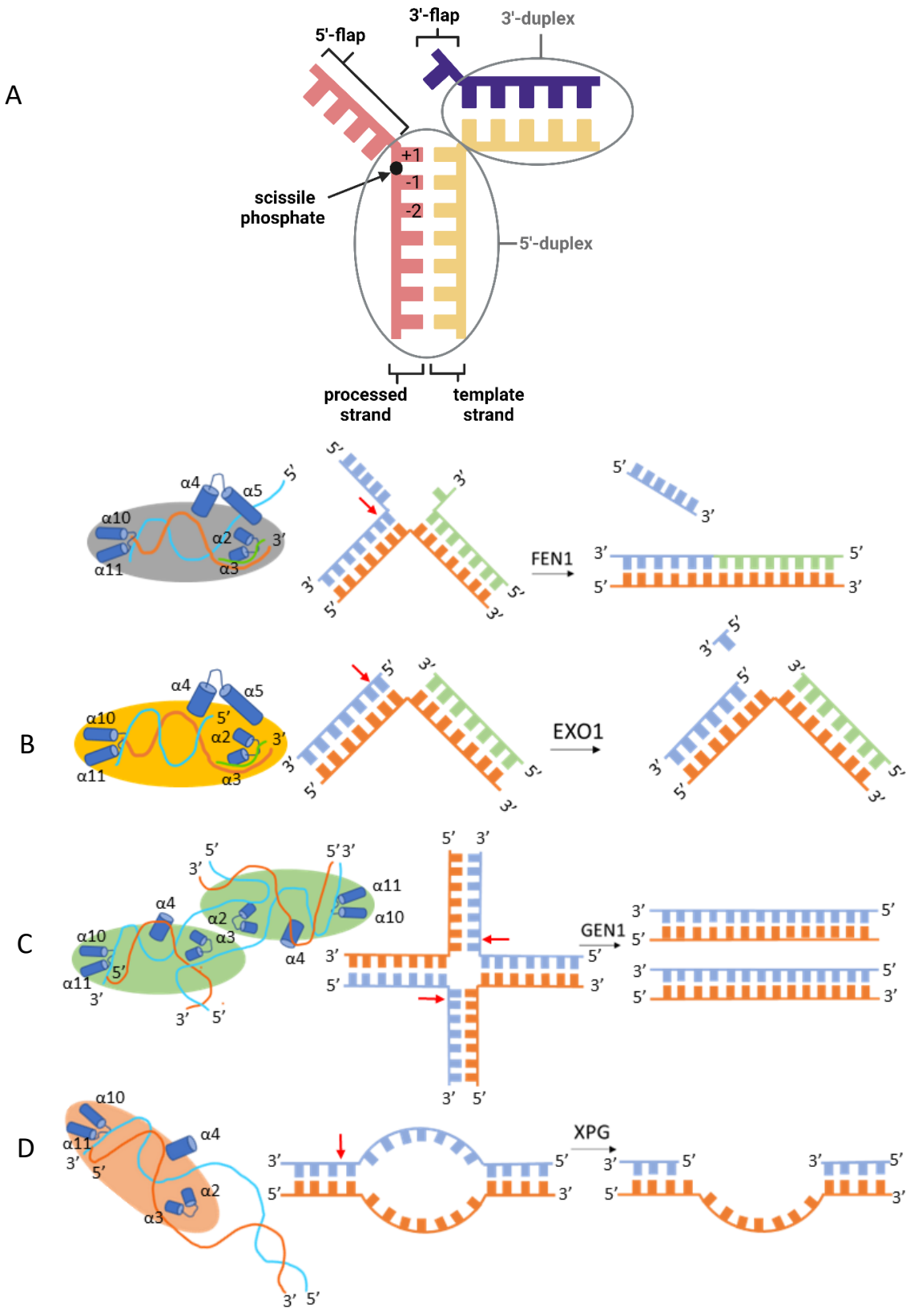


Figure 1.3 FEN superfamily substrate specificity

All schematic representations show the most critical helices required for function (blue cylinders). The processed strand is shown in blue, and the template strand is shown in orange. FEN1 and EXO1 have a third DNA strand shown in green. Sites of phosphodiesterase hydrolysis for each enzyme is shown by a red arrow. ***A****. Schematic representation of the FEN1 substrate naming system and associated substrate.* ***B****. Schematic representation of EXO1 and its associated substrate.* ***C****. Schematic representation of GEN1 and its associated substrate.* ***D****. Schematic representation of XPG and its associated substrate.*

1.4 Introduction to FEN1 and its importance in genome stability

FEN1 is a well characterised nuclease involved in the hydrolysis of 5'-flaps during DNA replication and DNA repair via its structure specific, divalent metal-ion phosphodiesterase activity [20]. FEN1's link to disease in higher eukaryotes has been well documented. Reports have demonstrated that FEN1 abolishment in mice models resulted in embryonic lethality [21]. However, the upregulation of FEN1 has also been reported in lung, breast, prostate, gastric and colorectal cancer, suggesting the over-production of FEN1 in living organisms plays a significant role in tumour proliferation [22]. Additionally, FEN1 gene knockout in DT40 chicken cell lines increased sensitivity to H_2O_2, methylene sulfonate (MMS) and N-methyl-N'-nitro-N-nitrosoguanidine (MNNG). DNA damaging agents are routinely removed from DNA through the base excision repair pathway (BER). Re-introduction of the FEN1 gene resulted in reduced sensitivity towards these agents. Therefore, *in vivo* reports of FEN1 would suggest that in order to uphold genome integrity and disease resistance, it is paramount that basal metabolic levels of FEN1 is maintained in cells [23].

1.4.1 The role of FEN1 in DNA replication

DNA replication is a highly regulated molecular event occurring in S-phase of the cell cycle. It consists of several complex biological components resulting in exactly one copy of the parental genomic DNA, before translocating into the daughter cell. For replication to be undertaken, the compactly folded genomic DNA within the nucleosome is first unravelled to expose the genomic duplex DNA [24]. Replication events begin at sites called 'origins' where the 11 subunit, double hexamer CMG complex (DNA helicase) binds and unwinds the double stranded DNA (dsDNA) in an ATP-dependent manner, producing two structures known as replication forks (Figure 1.4A) [25]. The separated strands (parental strands) act as the new template for which the complementary strands are synthesised. The polymerase-α/primase complex binds the template strands and produces a ~40 RNA/DNA primer before polymerases synthesise the daughter strands (Figure 1.4A). However, due to the antiparallel conformation of DNA [26], one daughter strand is produced continuously (leading strand) and the other is produced discontinuously (lagging strand). Synthesis of the leading and lagging strands are performed by polymerase-ε (pol-ε) and polymerase-δ (pol-δ) respectively. Pol-ε synthesises continuously in the 5'-3' direction whereas this is not attainable for pol-δ, instead, lagging strand synthesis results in ~200 bp long Okazaki fragments (Figure 1.4B).

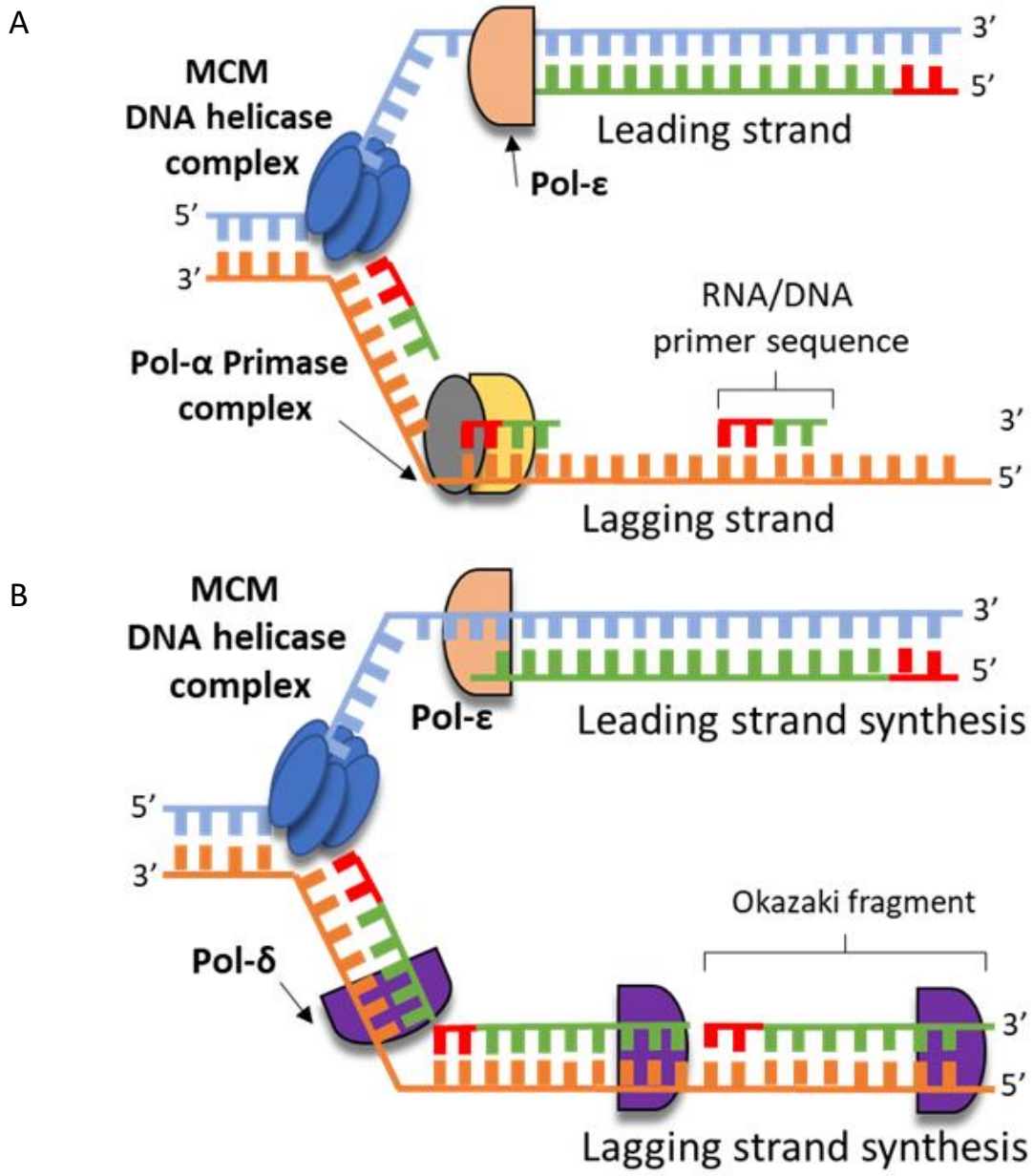


Figure 1.4 Leading and lagging strand synthesis in DNA replication

***A**. Schematic representation of DNA replication before Okazaki fragment processing. DNA helicase (blue ovals) breaks the hydrogen bonds between the template strands (light blue and orange). The pol-α/primase complex produces short RNA-DNA primer sequences shown by red and green ladders. The leading strand is then continuously synthesised by pol-ε (beige). **B**. Schematic representation of lagging strand synthesis whereby pol-δ binds the RNA-DNA primer sequence. The lagging strand is then discontinuously synthesised by pol-δ (purple), ultimately producing Okazaki fragments.*

Lagging strand maturation involves RNA primer sequence removal and annealing of the individual Okazaki fragments (Figure 1.5). This process is achieved enzymatically through the combined efforts of pol-δ and FEN1 [27]. Furthermore, the homotrimer proliferating cell nuclear antigen protein (PCNA) acts as a binding site for pol-δ and FEN1 during lagging strand synthesis and processing [28]. It is suggested that pol-δ displaces the RNA primer sequence of the following Okazaki fragment through its 5'-3' synthesis directionality, resulting in a 1-2-nucleotide 5'-flap that is subsequently hydrolysed by FEN1 [27], [29]. This iterative cycle known as nick translation, would repeat until no RNA primer sequence remains before DNA ligase I seals the DNA-DNA nick region between the two Okazaki fragments [27]. On the other hand, another model has suggested that pol-δ strand displacement results in longer 5'-flaps of >25 nucleotides. The longer 5'-flaps are coated by the stabilising protein RPA and become a viable substrate for the enzyme Dna2. However, Dna2-mediated 5'-flap hydrolysis is not precise and results in 2-6-nucleotide 5'-flaps and therefore requires FEN1 to remove the remaining 5'-flap before nick sealing by DNA ligase I [30]. Interestingly, the latter model regarding pol-δ-FEN1-mediated processing, is widely observed in most mammalian cells, whereas the Dna2-FEN1 model is predominantly found in yeast model systems.

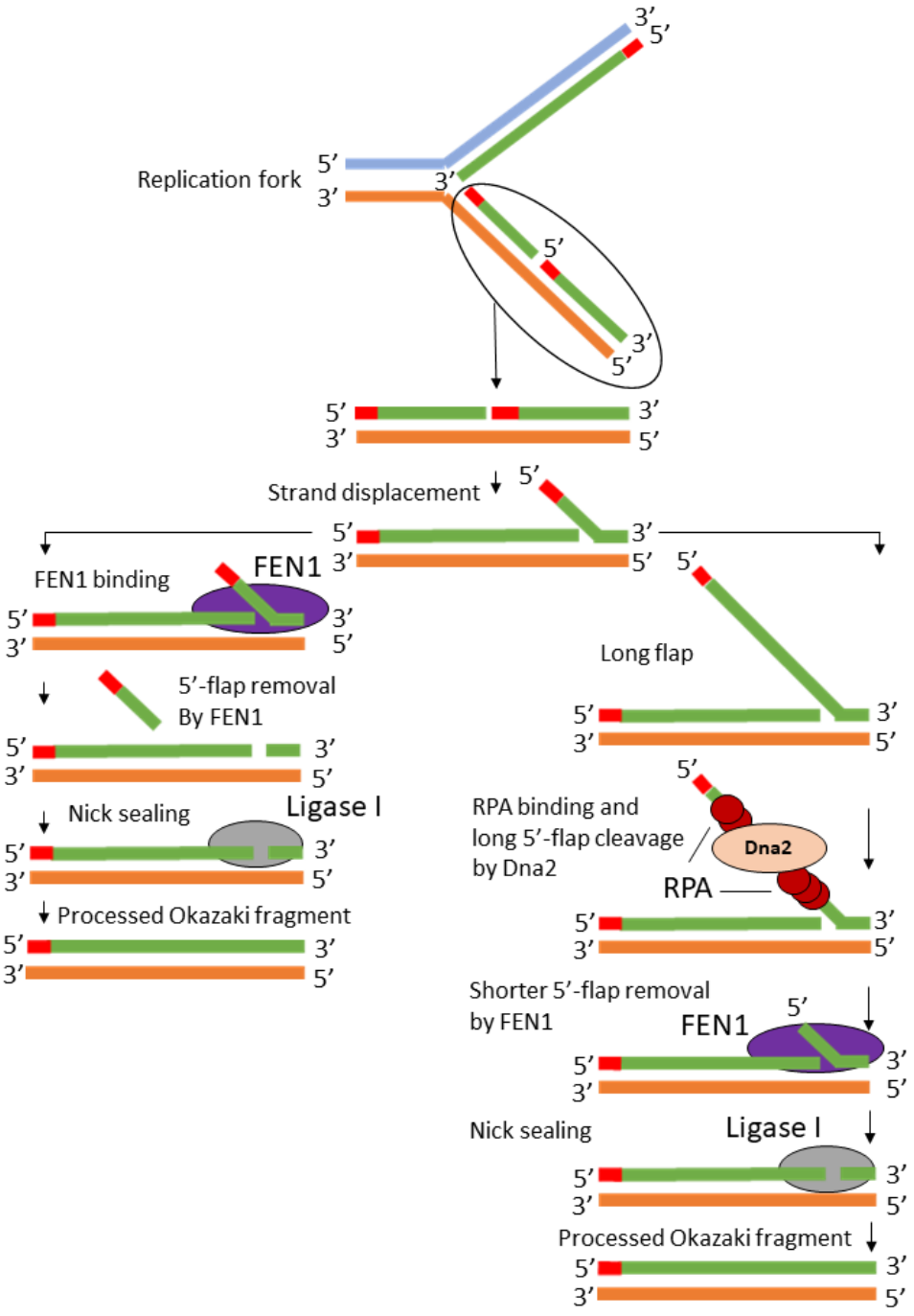


Figure 1.5 Proposed Okazaki fragment processing models by FEN1 and Dna2

Schematic representations of two proposed models for Okazaki fragment processing during DNA replication. One model is proposed for shorter 5′-flaps produced by strand displacement synthesis. During this process FEN1 (purple oval) hydrolyses the shorter 5′-flap before DNA ligase (grey oval) seals the nick (left descending cascade). If longer 5′-flaps are produced, RPA stabilises the longer 5′-flap and Dna2 (beige oval) non-specifically hydrolyses the 5′-flap. This produces a shorter 5′-flap which is subsequently processed by FEN1 before nick sealing by DNA ligase. The template strands of the replication fork are shown in blue and orange; the pol-α/primase primer sequence shown in red; synthesised Okazaki fragment sequence by pol-δ is shown in green.

1.4.2 The role of FEN1 in DNA repair

Genomic DNA must be scrupulously maintained to ensure cell survival and proliferation in all life. Despite this, DNA remains under constant assault from exogenous, and endogenous DNA-damage agents [31], [32]. Due to the varying types of DNA damage the genome is exposed to, cells have evolved a plethora of DNA repair pathways to ensure damaged DNA sites are identified, removed, and complementary DNA is restored so the transfer of genetic information from the parental cell to the daughter cell is not interrupted [33]. In the absence of DNA repair pathways, replication fork stalling and genomic degradation would occur, ultimately resulting in an increased susceptibility to multiple diseases including cancer [34]. The base excision repair (BER) pathway allows for the removal of damaged DNA from the double helix (Figure 1.6). For example, oxidation, deamination and alkylation are all types of damage rectified by the BER pathway. BER is further sub-divided into short-patch BER (SP-BER) or long-patch BER (LP-BER), depending on the number of bases required to be resynthesised [35].

In SP-BER, a DNA glycosylase is recruited to the DNA damage site, inverting (flipping) the damaged base away from the double helix before removal of the damaged base entirely through N-glycosidic bond cleavage. This results in an apurinic site where the 2-deoxyribose sugar is still attached to the phosphodiester backbone. The AP endonuclease, APE1, is recruited to remove the 2-deoxyribose sugar from the phosphodiester backbone [35]–[37]. Polymerase-β (pol-β) is employed and binds a free nucleotide to the 1-nucleotide gap in the DNA chain. Finally, DNA ligase seals the nick, resulting in a fully repaired duplex (Figure 1.6) [38]. In LP-BER, DNA glycosylase and APE1 are still recruited to remove the damaged base and apurinic site respectively. However, strand resynthesis by pol-δ/ε leads to longer 2-11 nucleotide patches. This longer patch, or 5’-flap, is then hydrolysed by FEN1 before ligation by DNA ligase (Figure 1.6) [39].

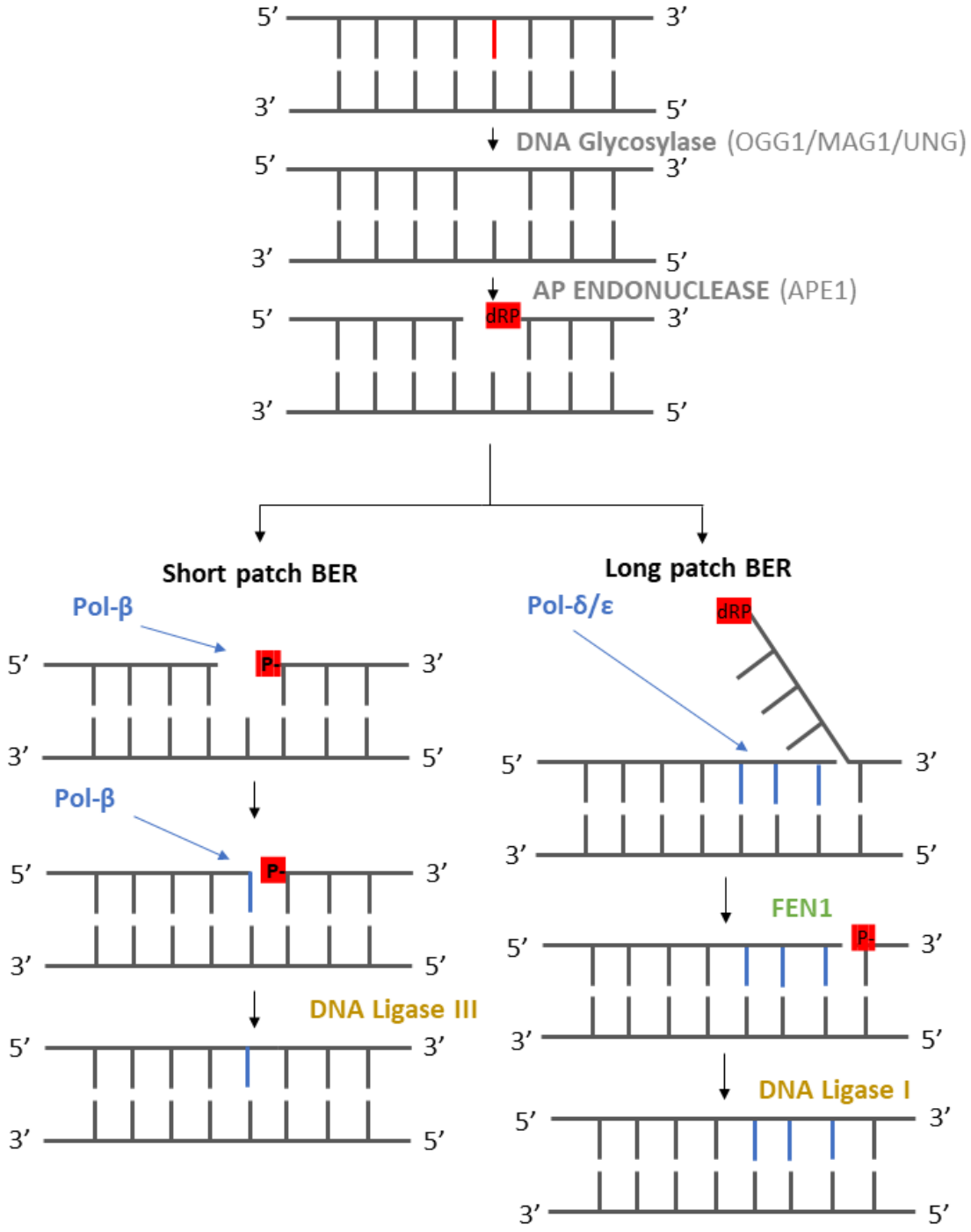


Figure 1.6 SP-BER and LP-BER

Schematic representation of SP-BER and LP-BER. Both pathways begin with the recruitment of DNA glycosylases to excise the damaged nitrogen base before recruitment of APE1 endonuclease to catalyse the cleavage of the phosphodiester backbone at the AP site. Pol-β is recruited to synthesise short patches before ligation with DNA ligase III. Longer patches produced by pol-δ/ε produce a 5′-flap intermediate which is hydrolysed by FEN1 before ligation with DNA ligase I.

1.5 FEN1 Structure

X-ray crystallographic studies have allowed greater insight into the structural properties of FEN1 and its family members in the presence, and absence of DNA [20], [40]–[42]. The structure of FEN1 resembles a left-handed boxing glove, consisting of five major domains (Figure 1.7A). Firstly, a DNA recognition site with a hydrophobic residue rich wedge was found within multiple crystal structures, formed from α2- α3, denoted as the hydrophobic wedge. Secondly, a 3'-flap binding site was found at the top of the glove, interacting with the 3'-flap and 3'-duplex . Thirdly, a HELIX-2-TURN-HELIX (H2TH) DNA binding domain (α10- α11) which contributes to the wrist region. Fourthly, a helical gateway region comprising of α4, α5 and the C-terminus of α2 which permits ssDNA entry towards the active site and finally, a conserved di-metal ion active site [20].

FEN1 consists of a 332 amino acid nuclease core domain and 47 amino acid C-terminal tail. The C-terminal forms two β-strands and one α-helical conformation to form a βA-αA-βB motif which acts as the main interaction site with PCNA. The nuclease core domain of FEN1 is composed of α and β secondary structures forming a central groove via seven stranded, twisted β sheet plus two helical regions where one helical region comprises α15-11. The other helical region consists of α1- α4, α12 and α13, all of which are positioned on either side of the β sheet, resulting in groove configuration. In addition, a 46-residue component known as the helical gateway (β3, α4 and α5) projects outwards from the main polypeptide core surface is shown to possess a saturated presence of hydrophobic and charged residues, suggesting its role of weaving single-stranded 5'-flaps into the active site [41], [43]. Within the active site, the two metal ions make direct contact with the scissile phosphate backbone on the processed strand and are directly coordinated by four carboxylate residues D86, E160, D179 and D181 that produce inner sphere contacts with both metal ions. Moreover, Glu160 operates as a bridging residue for the two divalent metal ions. Three more acidified residues accompanied by Y234 produce outer sphere contacts with the metal ions via water molecules within the structure [44].

1.5.1 The hydrophobic wedge

The hydrophobic wedge of FEN1 functions to induce bending of the 5’ and 3’ duplexes that make up part of the DNA junction. This is facilitated by the α2-α3 loop region and β6-β7 (β-pin) (Figure 1.7B) [45]. The α2 helix stacks against the template strand on the 5’-duplex via the hydrophobic residues A45, I44 and F42. Additionally, the α2-α3 loop region enforces bending by stacking against the first dsDNA base pair of the 3’-duplex. Finally, the β-pin is used to secure the 3’-duplex in place through positively charged residues [45].

1.5.2 The 3’-flap binding site

Double flapped substrates containing a 1-nucleotide 3’-flap display significantly increased FEN1 nuclease activity [46]. Because of this, it was suggested that archaeal and mammalian FEN1 possess a conserved region of the FEN1-DNA binding interface specifically evolved for 3’-flap recognition and binding (denoted as the 3’-flap binding site) [47]. Structural data suggests several FEN1-DNA contacts to the 3’-flap extrahelical nucleotide, originating from α3, α14 and α15 in the form of L53, T61, F316 and R320 (Figure 1.7C) [45], [47]. Moreover, R47 has also been claimed to participate in inter-domain cooperation between the 3’-flap binding site and the helical gateway, suggesting its importance in 3’-flap recognition. Furthermore, beyond the 3’-flap binding site are multiple electronegative residues, which are claimed to block the 3’-flap from progressing beyond the 3’-flap binding site during FEN1-mediated catalysis [20]. Despite the number of contacts reported at the 3’-flap binding site, functional data has reported varying changes in human FEN1 activity with respect to individual mutations. Insignificant changes in nuclease activity were reported for M65A, F316 and R320 [48]. A 50-fold reduction in activity was demonstrated in T61A, suggesting the loss of the hydrogen bond to the 3’-hydroxyl of the 3’-flap significantly affects reaction [48]. Moreover, R47A resulted in a 300-fold reduction in activity, implying its importance in functional crossover between the helical gateway and the 3’-flap binding site. However, the most important contact is L53, with the mutation L53A bearing a 1000-fold reduction in activity when compared to WT and therefore, plays a crucial role in 3’-flap recognition, helical arch ordering and scissile phosphate orientation [48].

1.5.3 The helical arch and cap

The FEN1 helical gateway and cap is a highly conserved FEN superfamily motif which functions to regulate the passage of ssDNA into the active site. The helical gateway of FEN1 is located directly above the active site, consisting of α4 and the C-terminus of α2. Furthermore, the helical arch is also observed in this motif and comprises of α4 and the α5 helical cap. (Figure 1.7D) [45]. Unlike other FENs, FEN1 and EXO1's helical arch possesses positively residues on helices 4 and 5. Despite this, mutations in the helical cap observe modest reductions in human FEN1 function, with a modest 1.5-fold reduction in catalytic efficiency by the mutation R129A, a residue that contacts the template strand at the ss/ds junction [20]. However recent data has demonstrated that the single turnover rate (k_{ST}) is significantly affected by R129A, bearing a 10-fold reduction in rate compared to WT (unpublished communication, Mark Thompson).

Below the helical cap is α4. This section of the helical arch contains the positively charged and hydrophobic residues K93, R100, R104 and Y40. However, alanine substitutions of these residues impact FEN1 function to varying degrees of significance. K93 and R100 contact the 5'-monoester of the 5'-flap and observed the largest reduction in nuclease activity by >400-fold in their respective mutants K93A and R100A. Y40 aids in 5'-flap positioning and stacks against the +1-1 base in both substrate and product structures, the Y40A mutant subsequently resulted in a 20-fold reduction in nuclease activity. Finally, the mutation R104A modestly affected catalytic activity by 3-fold, implying its less important role during reaction [20], [40]. Although, like R129A, the single turnover rate was reduced by 40-fold compared to the WT (unpublished communication, Mark Thompson). The helical arch of FEN1 remains disordered in the absence of substrate. However, in the presence of substrate, the disordered helical arch produces a wide enough aperture to allow 5'-flap threading to occur. Upon 3'-flap recognition, the helical arch transitions to the ordered state, reducing the aperture to approximately 13-15 Å. In doing this, the helical arch permits passage of single-stranded (ssDNA) whilst rejecting dsDNA [45], [48].

1.5.4 The H2TH:K^+ binding site

The H2TH and K^+ binding site is a routinely observed DNA binding domain in a variety of proteins including nucleases, glycosylases and polymerases [49]. The FEN1 H2TH:K^+ binding site is situated at the wrist region of FEN1 (α10-α11). Reports suggest the FEN1 H2TH:K^+ domain is responsible for initial FEN:DNA binding of the template strand on the 5'-duplex (Figure 1.7E) [45]. The K^+ ion is co-ordinated by the backbone carbonyl oxygens of I238 and I241 and the side chain hydroxyl of S237. Additionally, the H2TH domain contains basic arginine and lysine residues at positions, 239, 244, 245 and 267. It is proposed that the combination of the K^+ binding site and positively charged residues may facilitate anchoring of the substrate to FEN1 before 5'-flap threading and 3'-flap binding. Interestingly, the basic residues at position 239 and 245 are conserved in human FEN1 and EXO1 however are not observed in XPG or GEN1, possibly alluding to their role in positioning double flapped and nicked substrates [45], [50]. However, detailed kinetic and biophysical analysis of these residues has not yet been undertaken.

1.5.5 The active site

The FEN1 active site facilitates 5'-flap hydrolysis through the co-ordination of divalent Mg^{2+} ions and a single water molecule. The active site consists of seven highly conserved acidic residues, also observed in other FENs (Figure 1.7F) [45]. The active site houses 2 divalent cations, one of which is co-ordinated by D181 and D233, and the other is co-ordinated by D34 and D86, all of which display severely reduced nuclease activity when any of the carboxylates are substituted for alanine or asparagine [20], [51]. Structures of the FEN1 active site have shown the water molecule to position approximately 3 Å away from the scissile phosphate during its activation by the catalytic metals. It is also suggested that water activation is facilitated by the main chain amino group of Gly2 as the methionine residue is removed post translationally from the N-terminus [40]. However, basic residues from the helical arch also contribute towards the active site of FEN1. For example, R100 and K93 make electrostatic interactions with the scissile phosphate to offer further stability during catalysis whereas the hydroxyl of Y40 forms a hydrogen bond to the +1 phosphate, stacking against the +1 nucleobase, ensuring efficient transfer of the +1 base throughout the active site [40], [51].

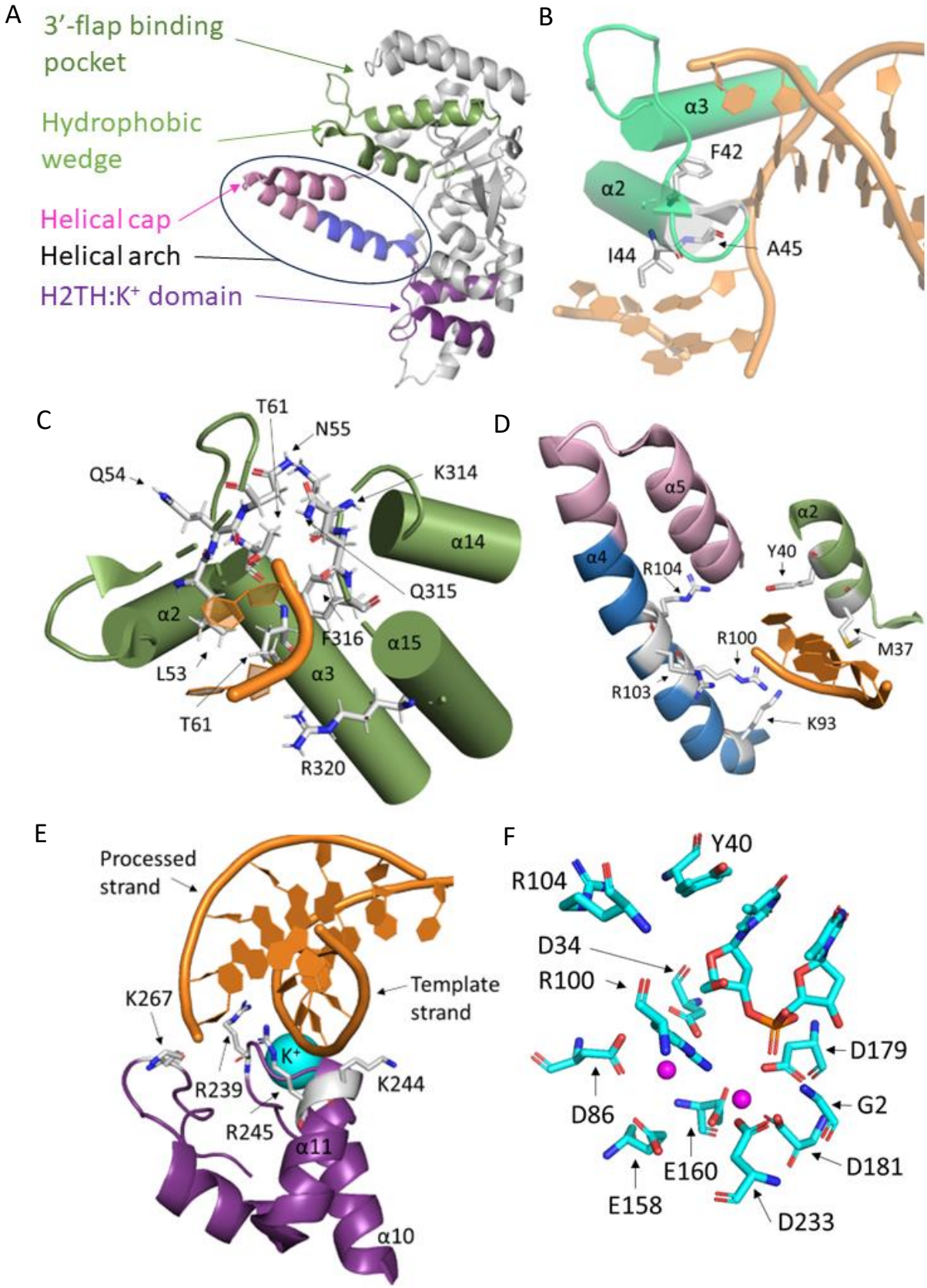

A
3'-flap binding pocket
Hydrophobic wedge
Helical cap
Helical arch
H2TH:K+ domain
B
α3
F42
α2
I44
A45
C
T61
N55
Q54
K314
α14
Q315
α2
F316
α15
L53
α3
T61
R320
D
α5
α2
α4
R104
Y40
R100
M37
R103
K93
E
Processed strand
Template strand
K267
K+
K244
R239
R245
α11
α10
F
Y40
R104
D34
R100
D179
D86
G2
E160
D181
E158
D233

Figure 1.7 Structural domains of FEN1

***A**. Structure of FEN1 showing the important α-helices required for catalysis. **B**. magnified image of the α2-α3 hydrophobic wedge (green) with the substrate shown in orange. **C**. Magnified structural image of the 3'-flap binding site consisting of α2, α3, α14 and α15 with the 10 amino acid residues that make up the binding site shown as sticks. Carbon atoms for the individual amino acids are shown in grey; nitrogen's shown in blue and oxygens shown in red. **D**. Helical arch of FEN1 with α4 shown in blue; the helical cap shown in pink and the hydrophobic wedge helix α2 shown in green. **E**. The H2TH:K^+binding site of FEN1 (purple). **F**. Active site of FEN1 with the conserved carboxylates accompanied by the two metal ions (pink spheres). PDB code: 3Q8K.*

1.6 Proposed models of FEN1-mediated catalysis

FEN1's quintessential role in maintaining genome fidelity has provoked research groups to elucidate the series of events that occur before FEN1 phosphodiesterase activity is achieved. By uncovering the true model of FEN1-mediated processing, more robust and effective drugs therapies can be developed to combat diseases whereby FEN1 is up, or down regulated.

1.6.1 Track and clamp model

Earlier reports postulated that FEN1 binds the end of the 5'-flap and proceeds to track down the ssDNA before clamping down around the scissile phosphate at the ss/dsDNA junction. This was postulated through crystallographic structures of T5 5' exonuclease, a homolog of FEN1 demonstrating that the aperture of the helical arch was too small accommodate dsDNA however could accommodate ssDNA with platinated adducts and other branched structures [52]–[54]. These studies initially offered an attractive model which showed consistency to FEN1's structural features regarding it's disordered helical arch, which could accommodate modified ssDNA accompanied by its subsequent ordering around the scissile phosphate, whilst at the same time excluding the 5'-duplex (Figure 1.8A) [55].

1.6.2 Threading model

Conversely to the track and clamp model, other groups postulated that human FEN1 preferably binds the base of the 5'-flap independent of flap length. FEN1 manoeuvres the 5'-flap through the helical arch in a threading motion (Figure 1.8B) [55]. This was demonstrated by cleavage assays and binding assays where the 5'-flap of the substrate was blocked with streptavidin which could no longer pass through the helical arch. Nuclease assays revealed unblocked substrates were fully consumed at the lowest FEN1 concentrations whereas pre-blocked substrates showed the lowest level of product formation independent of FEN1

concentration. Additionally, binding affinity assays demonstrated higher affinity binding between FEN1 and longer 5'-flaps however, blocked, and short 5'-flaps significantly reduced the dissociation rate of the FEN1-substrate complex by approximately 10-fold compared to unblocked long 5'-flaps with a dissociation rate of 10 minutes [55].

Interestingly, reports have demonstrated FEN1's ability to endonucleolytically process 5'-gapped flap substrates [46]. These substrates contain a single-stranded region followed by dsDNA on the 5'-flap. This caused conflict within the literature as dsDNA cannot pass through an ordered FEN1 helical arch, raising questions as to whether the track and clamp model or threading model is utilised in FEN1-mediated catalysis. However, later reports offered confirming evidence for the threading model. *Patel et al* demonstrated that 5'-gapped flap substrates annealed with a 5'-streptavidin tetramer before the addition of FEN1 was too large to pass through an ordered or disordered helical arch, resulting in reaction inhibition. In contrast, Inhibition was not observed when the 5'-streptavidin tetramer was added to the preformed enzyme-substrate complex, implying the streptavidin was not inhibiting the reaction and that the processed strand must have already passed through the helical arch (described as trapped). Additionally, competitor assays demonstrated that the trapped 5'-gapped flap with a 5'-streptavidin tetramer was not out competed by other 5'-gapped flap substrates, indicating the complex was trapped in a permanent state by streptavidin [56]. As the gapped flap cannot pass through the structured arch, these findings also suggest that the FEN1 helical arch is disordered when the gapped flap substrate passes through and that gapped flapped DNA acts like single-stranded 5'-flaps. In doing so, the duplex region of the flap can pass through the helical arch before scissile phosphate catalysis is undertaken. This model outcompetes the tracking model because FEN1 binding at the base of the flap negates issues of binding non-optimal substrates containing complicated secondary structures on the 5'-end, which may otherwise slow the catalytic process.

1.6.3 Base unpairing and 5'-flap nucleotide flipping model

The findings shown in section 1.6.2 led to the consensus that 5'-flap threading is crucial for the removal of varying 5'-flaps in human FEN1 and to this day, is still widely agreed upon [55]. Nonetheless, a new model was proposed when structural data suggested 5'-flap base flipping and double nucleotide unpairing in order to achieve catalytic competency (Figure 1.8C) [20]. These findings suggest that nucleotides on the 5'-flap

beyond the scissile phosphate become inverted to protect the 5'-flap from non-specific incision of phosphodiester backbone, offering shielding it from the catalytic metals [20]. Furthermore, it was suggested that unpairing of the +1-1 bases ensures that the scissile phosphate is transferred towards the catalytic metals with Y40 stacking against the +1 base on the substrate complex before being stacked against the -1 base in the product complex [20], [57]. However, recent structural data has opposed this model after demonstrating pairing of both the +1 and -1 base pairs on the template strand in a D86 mutant human FEN1-substrate complex, and that the scissile phosphate was already in proximity of the catalytic metal ion active without the need for unpairing [40]. Nevertheless, it was found that the DNA is substantially distorted (twisted) to reach the active site.

1.6.4 Loop-wedge allosteric binding model

New evidence has suggested that allosteric binding of the 3'-flap to the loop wedge motif on α2-α3 drives FEN1 conformational ordering and manoeuvring of the scissile phosphate towards the active site to achieve FEN1 catalytic competency (Figure 1.10 & Figure 1.11) [48]. Similar to previous models presented in section 1.6, Initial stages of the reaction begin through DNA bending at the ss/dsDNA junction and the H2TH:K^+ domain of FEN1 acts as an anchoring point for the template strand on the 5'-duplex [15], [45]. Following 5'-flap threading through the helical arch, recognition of the 3'-flap induces conformational changes in both the 3'-flap binding site, helical arch and the +1-1 bases on the 5'-duplex in allosteric fashion, all regulated by L53 at the 3'-flap binding site [48]. This allosteric binding model stands out from previous theories because it demonstrates why FEN1 kinetically favours double flap substrates containing a 1-nucleotide 3'-flap and why this loop-wedge of α2-α3 motif is highly conserved across multiple domains of life [46], [48]. Considering the scientific communities' current consensus that this is the most accurate mode of FEN1 function, greater understanding of how conformational changes in FEN1 drive catalysis are required. Although discussed briefly in this section, section 1.7 undergoes a more enriched understanding as to how these conformational changes occur, why they occur, and the evidence supporting this proposed model.

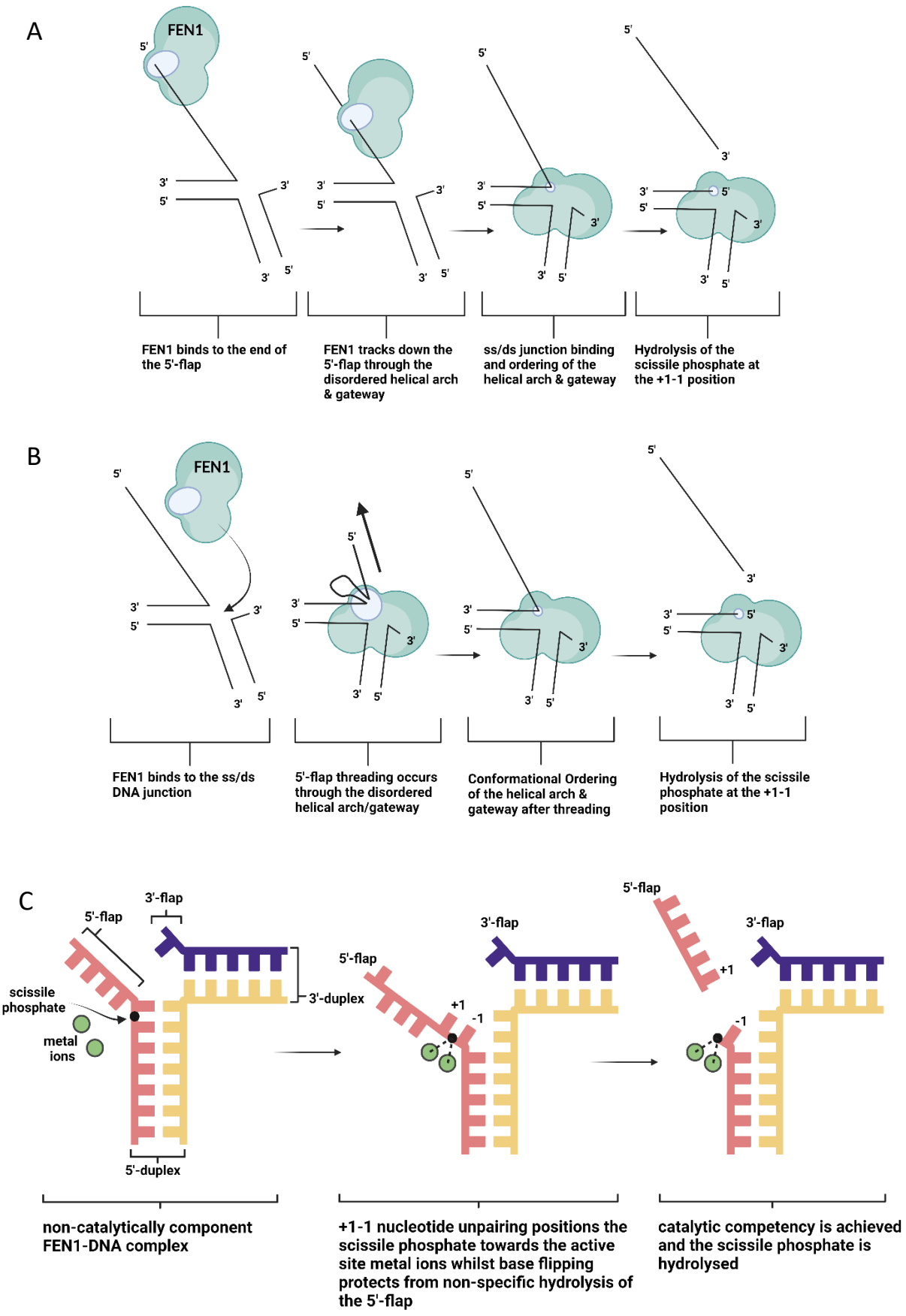


Figure 1.8 Proposed models of FEN1-mediated catalysis

Historically proposed models of FEN1-mediated catalysis. This figure was created using Biorender.com. ***A****. Track and clamp model.* ***B****. Threading model.* ***C****. Nucleotide unpairing and base flipping model.*

1.7 Conformational changes required for FEN1-mediated catalysis

FEN1 is an important component of Okazaki fragment processing and resolves 5'-flaps in the latter stages of LP-BER [58], [59]. FEN1's association to these pathways is attributed to its ability to facilitate precise and reproducible incision of the scissile phosphate on the 5'-duplex. For accurate 5'-flap hydrolysis to occur, FEN1 has evolved to ensure the correct substrate is recognised, orientated, and checked before hydrolytic incision of the phosphodiester backbone is facilitated. The primary strategy used in FEN1-mediated catalysis is the utilisation of conformational changes in both the protein and substrate. Moreover, conformational changes are not isolated to one specific event during the reaction, but instead observed throughout multiple stages of FEN1-DNA processing in the form of DNA 5' and 3'-duplex bending, ordering of the helical arch post 5'-flap threading following 3'-flap recognition and finally, active site transfer of the +1-1 scissile phosphate. Therefore, to truly elucidate the FEN1-DNA mechanism, a deeper understanding of how conformational changes contribute towards FEN1 function is ultimately required.

1.7.1 5'- and 3'-duplex bending

Conformational changes in the FEN1-DNA reaction are first encountered following binding. FEN1 forms contacts with the 2 sections of dsDNA within the substrate in which the template strand of the 5'-duplex is contacted by the H2TH:K^+ domain and the 3'-flap and 3'-duplex is contacted by the 3'-flap binding site [45], [57]. By proxy, a 100 ° bend between the 5' and 3'-duplexes is observed and subsequently stabilised by the conserved hydrophobic wedge domain on α2 (denoted as DNA bending) [60], [61].

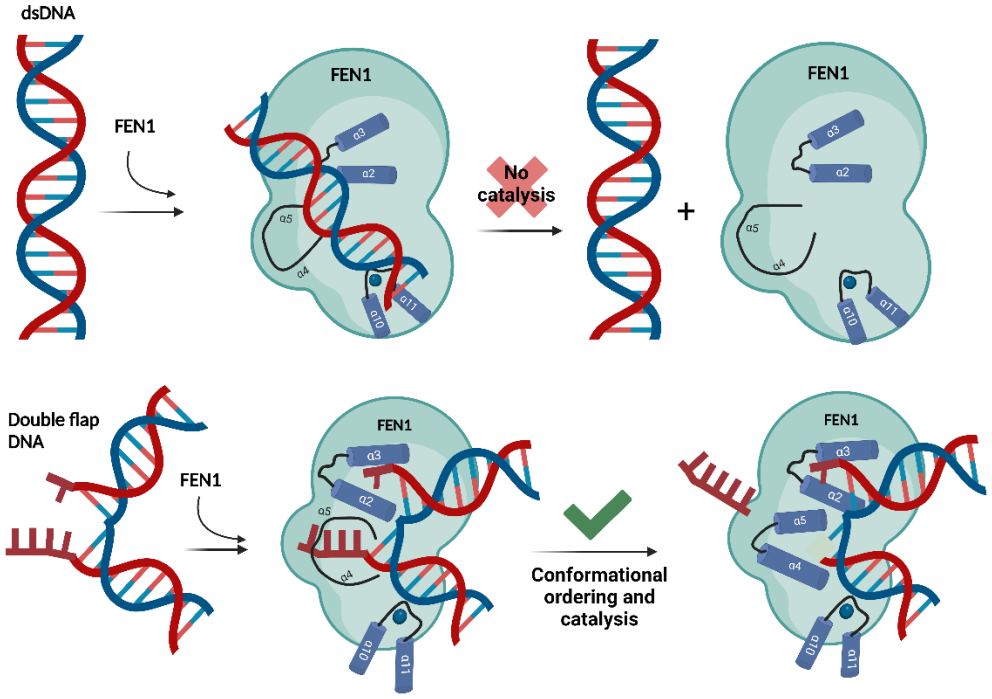
Forster resonance energy transfer (FRET) has been useful in uncovering the conformational dynamics involved in the FEN1-DNA reaction mechanism. Briefly, FRET involves the non-radiative transfer of energy between an excited donor fluorophore and to another acceptor fluorophore by long-range dipole-dipole coupling. FRET acts as a molecular ruler that can measure the molecular proximity and has shown to be highly efficient if the donor and acceptor fluorophores are positioned within 1-10 nm. Upon interaction between the donor and acceptor dyes, a FRET signal is produced and the FRET efficiency is described as the inverse 6^{th} power of the intermolecular separation and can be used to quantify the energy transfer between fluorophores [62]–[65]. FRET studies suggest that DNA bending provides a scaffold whereby key features of

the substrate can be recognised including the 3'-flap, 5'-flap and ss/ds DNA junction. Due to FEN1 being unable to bend continuous dsDNA compared to junction-containing DNA, the process of bending has been suggested to be an initial checkpoint stage to reject non-compatible substrates to FEN1 [60], [61] (Figure 1.9A). Studies have suggested that FEN1 sculpts the DNA into a bent conformation compared to capturing the substrate in the already bent conformer [60]. However, unpublished NMR data (unpublished communication, *Finger et al*) has provided evidence that that DNA substrates contain different conformational states in the absence of protein, implying the substrate is 'captured' by FEN1, instead of 'sculpted' into the bent conformer by the H2TH 3'-duplex binding site, 3'-flap binding site and hydrophobic wedge domains (Figure 1.9B).

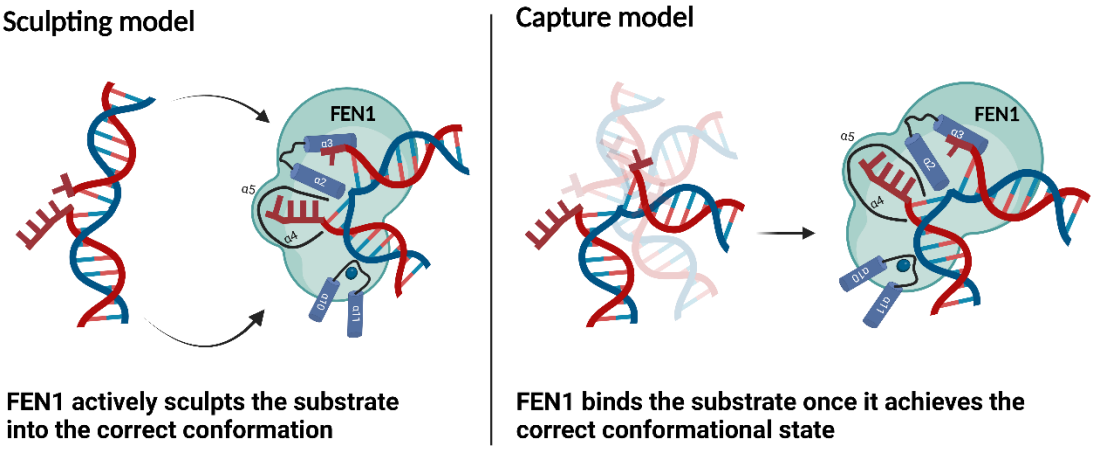
Single-molecule FRET (smFRET) experiments of FEN1 double flap substrate reactions under catalytic conditions showed that non-optimal substrates were hydrolysed by FEN1, but only after multiple binding, releasing and rebinding events occurred before the catalytically competent state was achieved stochastically. On the other hand, optimal FEN1 substrates such as double flaps possessed one specific DNA bending event prior to 3'-flap-induced conformational ordering, leading to faster 5'-flap hydrolysis. These findings show that FEN1 can catalyse non-optimal substrates, however these events do not significantly impact the overall fidelity of the cell due to the preference for more optimal double flapped substrates (Figure 1.9C) [66]. Furthermore, this model is consistent with kinetic data demonstrating an 8-1000-fold reduction in the k_{STmax} for suboptimal substrates [46], (unpublished communication, Mark Thompson). Therefore, it is suggested that the FEN1 reaction rate is affected by FEN1's ability to melt the 3'-duplex, resulting in the formation of a 3'-flap required to facilitate catalysis [46], [48]. SmFRET studies have also offered greater understanding as to which structural and molecular components of FEN1 are required for DNA bending to be undertaken. [67], [68]. Although metal ions are required to facilitate catalysis and threading of the 5'-flap in the form of Mg^{2+} or non-catalytic Ca^{2+} [56], the FRET efficiency only marginally increased upon FEN1-DNA binding in the presence of Ca^{2+} ions, this implies that metal ions are not required to achieve the bent DNA conformation measured when the donor and acceptor fluorophores were attached to the 5' and 3'-duplexes [69]. Moreover, no significant change in FRET signal was observed with active site mutants K93A, R100A, K93/R100A, L130P and D181A in the presence of Ca^{2+} ions, implying these active site residues are non-critical

in achieving WT-like DNA bending. Additionally, the 5’-flap is also not required to achieve global substrate bending [69].

A


dsDNA
FEN1
FEN1
No catalysis
FEN1
+
Double flap DNA
FEN1
FEN1
Conformational ordering and catalysis
FEN1

B


Sculpting model
FEN1
FEN1 actively sculpts the substrate into the correct conformation
Capture model
FEN1
FEN1 binds the substrate once it achieves the correct conformational state

C

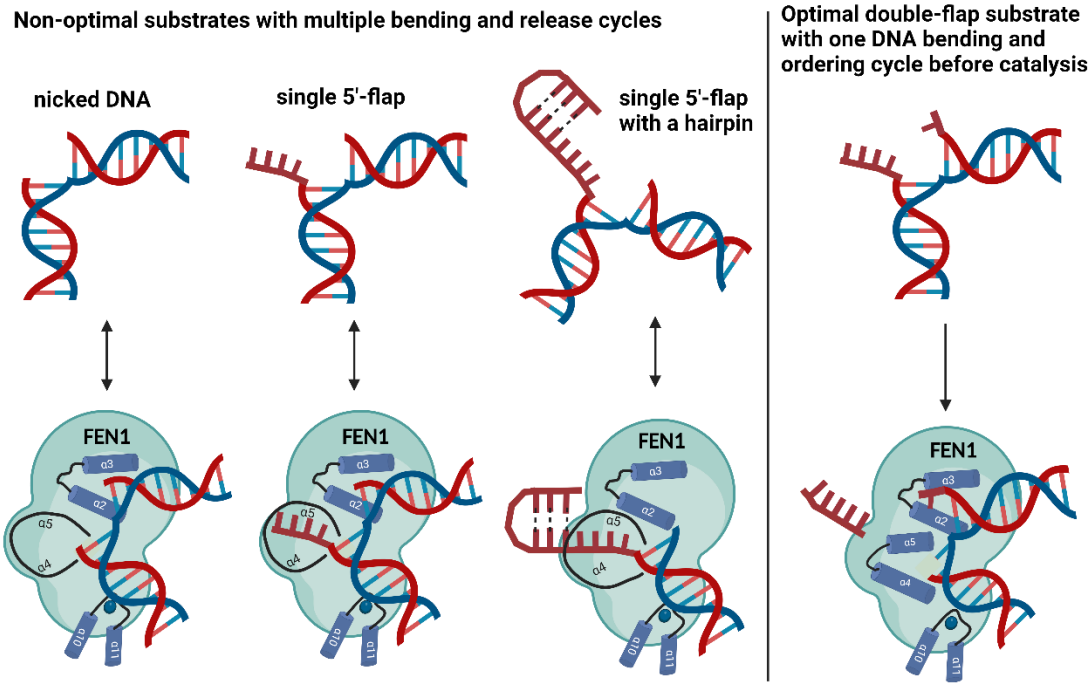

Non-optimal substrates with multiple bending and release cycles
nicked DNA
single 5'-flap
single 5'-flap with a hairpin
FEN1
FEN1
FEN1
Optimal double-flap substrate with one DNA bending and ordering cycle before catalysis
FEN1

Figure 1.9 DNA bending in FEN1-mediated catalysis

***A**. FEN1's ability to bend junction-containing DNA but not duplex DNA. **B**. The proposed sculpting and capturing models as to how FEN1 initially binds to junction-containing double-flapped substrates. **C**. FEN1's binding and bending ability of multiple substrates containing a ss/ds junction. Non-optimal substrates lacking a 1-nucleotide 3'-flap have multiple binding and bending events (double-facing arrow) whereas the optimal substrate proceeds to catalyse the substrate after one binding and bending event.*

1.7.2 Recognition of the 3'-flap results in conformational ordering of the helical arch

Previous reports have demonstrated FEN1's ability to hydrolyse a multitude of substrates including double flapped substrates, single 5' and 3'-flaps, nicked substrates containing neither 5'- or 3'-flaps and 5'-hairpins [45], [46]. Despite FEN1's accommodation of multiple substrates, kinetic studies have shown that the greatest reduction in nuclease activity was observed in the absence of a 3'-flap [46], [70]. Reports have also suggested that the 3'-flap plays an important role in progressing the reaction through an identified 3'-flap binding site on the loop region of α2-α3 [42]. More specifically, it is proposed that that 3'-flap recognition induces a conformational disorder-to-order transition of the α4-α5 helical arch after 5'-flap threading, enclosing the helical arch and exposing the 5'-flap to the active site core residues (Figure 1.10A) [20]. Functional coupling of the 3'-flap binding site and the helical arch was demonstrated through alanine mutations of T61, R47, and L53 located on the α2-α3 region, all of which exhibited diminished 5'-flap threading ability which is required to achieve a catalytically competent active site. This implies that as the native environment of the 3'-flap binding site is worsened, a decrease in 5'-flap orientation and stabilisation is observed [48].

3'-flap binding mutants are not only detrimental to 5'-flap threading, but they also have an impact on further DNA conformational changes required to position the scissile phosphate during active site transfer [69]. Mutations such as L53A also perturb conformational distortion (or twisting) of the +1-1/-1-2 base pairs in the 5'-duplex of the double flapped substrate measured through the reporting of conformational changes of neighbouring 2-aminopurine (2-AP) nucleobases [71]. The L53A mutant was unable to enact active site transfer-required DNA conformational changes as shown by the WT, further suggesting that the 3'-flap binding site contributes towards efficient threading and conformational ordering (Figure 1.10B) [48]. The

structural and biophysical data also shows consistency with published NMR data where in the presence of substrate, both the helical arch and the loop region of α2-α3 show a change in conformational states [72].

Despite the literature strongly suggesting that FEN1-DNA processing is driven through conformational changes of the protein and substrate, it provoked the question; why FEN1 evolved a 3'-flap binding site many angstroms away from the helical arch and active site? It is suggested that long-range 3'-flap-induced conformational changes can allow FEN1 to hydrolyse 5'-flaps that possess secondary structures or RNA of varying lengths and thus, offering greater substrate accommodation flexibility compared to an alternative FEN1 structure whereby the 3'- and 5'-flaps were recognised within the nuclease core [66]. Moreover, why in the absence of substrate, is the helical arch and 3'-flap binding sites initially disordered? One report has suggested that the disordered helical arch radius may allow for easier 5'-flap threading accompanied by the accommodation of secondary structures if the 5'-flap is extended [72].

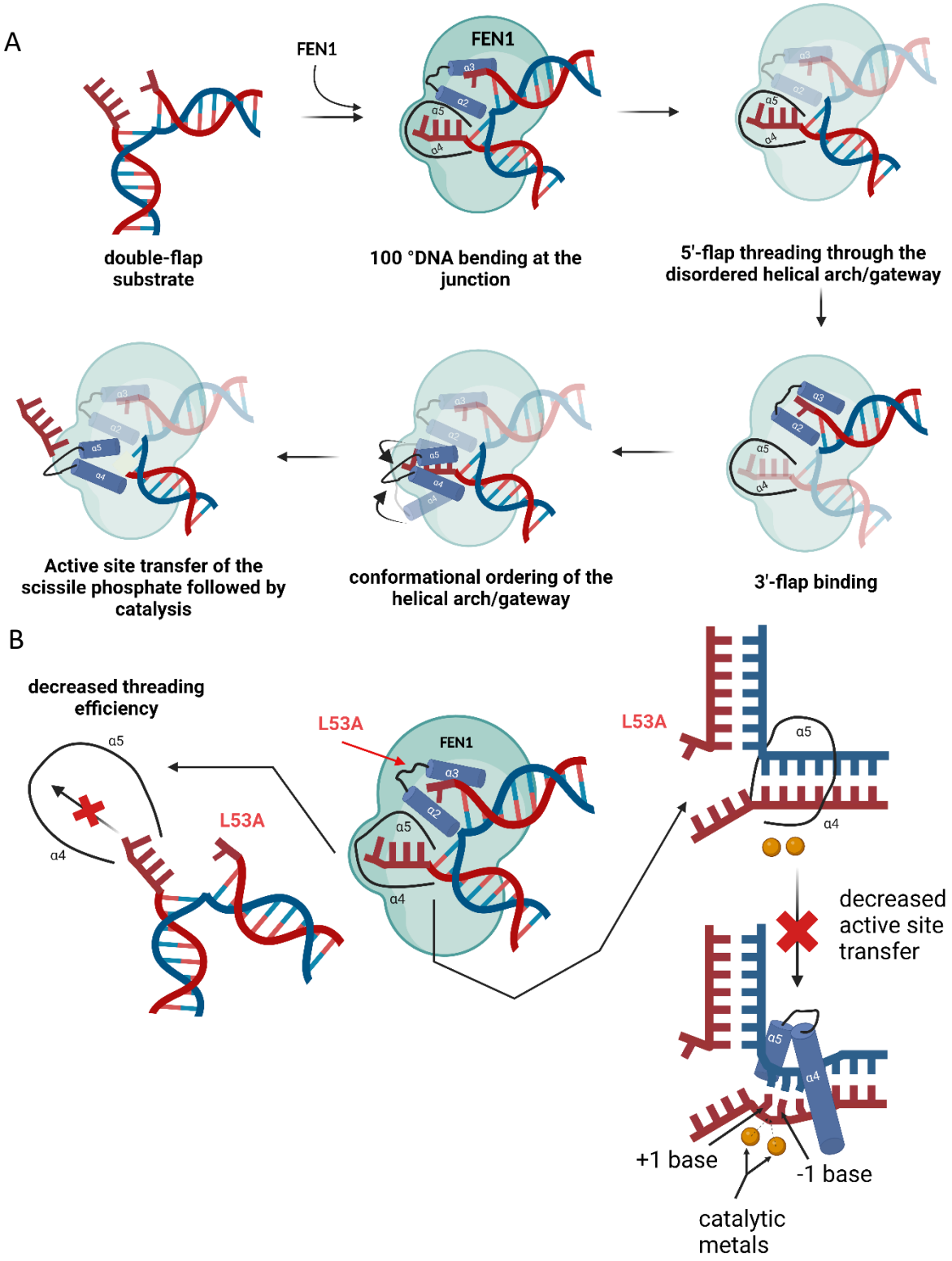


Figure 1.10 Helical arch/gateway ordering upon 3'-flap recognition

***A**. Allosteric binding model of FEN1-mediated catalysis and the series of conformational change events in the 3'-flap binding pocket, helical arch/gateway and the +1-1 bases either side of the scissile phosphate before catalysis. **B**. the 3'-flap binding pocket mutation L53A accompanied by its detrimental effects on 5'-flap threading and active site transfer to the active site. This figure was created using Biorender.com*

1.7.3 Active site transfer of the scissile phosphate towards the catalytic metals

The final area of interest regarding FEN1-substrate conformational changes resides at the active site core of the protein, although, DNA bending and 3'-flap induced ordering of the helical arch are required throughout the reaction process, those molecular events alone do not result in FEN1's inherent substrate specificity [55], [73], [74]. It is reported that active site residues are required to facilitate active site transfer of the scissile phosphodiester located at the +1-1 base positions of the 5'-duplex towards the active site metal ions, only then may precise hydrolytic incision of the DNA backbone be undertaken [69]. However, change in substrate structure and mutations to conserved FEN1 active site/ helical cap residues effect optimal active site transfer to varying degrees of severity utilising ECCD experiments described in section 1.7.2.

ECCD data demonstrated that active site mutations Y40A and R100A resulted in a lesser decrease in ECCD signal compared to WT. In addition, mismatch base modifications at the +1 position result in near no change in ECCD signal -1-2 labelled double flapped substrates. ECCD data of single 3'-flap substrates demonstrated varying degrees of DNA conformational change with respect to the +1-1 or -1-2 2-AP labelled positions. The ECCD signal of K93A mostly resembled that of the WT whereas R100A and R100A/K93A also led to reduced signal but not to the same extent as the WT with +1-1. Furthermore, Y40A, D181A and the helical secondary structure disordering mutant L130P did not significantly reduce the ECCD signal with +1-1 labelled single 3'-flaps. Moreover, single 3'-flap substrates with -1-2 2-APs demonstrated changes in ECCD signal in Y40A, K93A, R100A and D181A like the WT. However, removal of the 5'-terminal phosphate on the processed strand resulted in increased ECCD signal in both L130P and D181A [69]. Said differently, the data suggests conserved active site residues play minimal roles in DNA conformational changes at the -1-2 nucleobase positions and that these DNA conformational changes are primarily driven by helical cap secondary structure ordering and the metal-coordinating ion D181 [56]. These two components come together to 'roll', 'twist' or 'distort' the 5'-duplex independent of a 5'-terminal phosphate. However, at the +1-1 position of the 5'-duplex, active site residues including Y40, R100, R104 (unpublished communication, Mark Thompson) and D181 facilitate optimal active site transfer of the scissile phosphate towards the active site and do so in the presence of a 5'-temrinal phosphate [69].

Crystallographic structures have demonstrated stacking interactions between Y40 to the +1 or -1 nucleobases and fluorescence data also showed rapid +1-1 2-AP signal quenching perceived to be pairing-unpairing events between the two nucleobases and Y40 [20], [69], [75]. The active site residue R100 also forms contacts with the monoester product therefore it is also suggested that R100 may form contacts, and subsequently induce conformational changes at the +1-1-nucleobase base position. Finally, because D181 is a metal-coordinating ion in the active site, its substitution to alanine may negatively affect positioning of the metals required to induce conformational changes (Figure 1.11). Interestingly, reported findings signify that these active site residues do not induce DNA conformational changes of double flap substrates containing 2-AP bases at the -1 and -2 positions [20], [69]. Collectively, these findings accompanied by the fact that Y40A, R100A and D181A reduce nuclease activity by factors of 100, 400 and 800 respectively strongly infer their necessity towards efficient phosphodiesterase activity in FEN1 [20].

Ultimately, the information presented in section 1.7 has presented how FEN1's inherently evolved binding sites and use of protein/substrate conformational changes provides efficient 5'-flap processing. Without FEN1's ability to recognise optimal substrates, this has knock-on effects on how the genome is repaired and maintained. For example, imprecise incision of the 5'-flap at the +1-1 phosphodiester would then require DNA polymerases to be redeployed to the junction to replace the removed nucleotides. Additionally, FEN1's ability to recognise, and favour substrates with a 1-nucleotide 3'-flap produces the perfect ligatable substrate for DNA ligase and thus, increasing the rate of latter DNA replication processes [75]–[77].

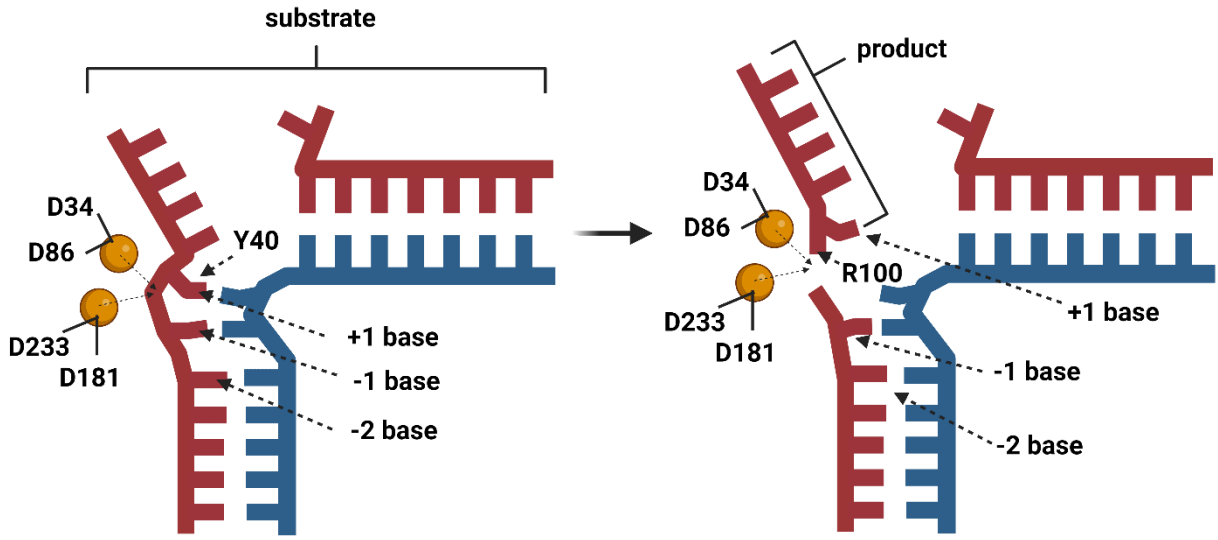


Figure 1.11 Conformational changes required for active site transfer

Active site residue roles in active site transfer before catalysis. Y40 stacks against the +1 nucleobase (left) and R100 interacting with the 5'-flap product (red). The catalytic metals (orange) are co-ordinated by four carboxylates in the active site. This figure was created using Biorender.com

1.8 Introduction to Fanconi anaemia nuclease 1

Genome fidelity relies on multiple DNA maintenance and repair pathway's abilities to recruit DNA nucleases to the site where phosphodiesterase activity is required. Therein, several nucleases have been characterised to elucidate their wider role in DNA repair and how defects in these proteins result in disease [78], [79]. However, the functional significance of several other DNA repair nucleases remains enigmatic. One example of a less well-characterised nuclease is FAN1, a seemingly versatile structure-specific 5'-nuclease involved in the protection against interstrand crosslink (ICL) damaged DNA. The remainder of this chapter will review the current literature surrounding FAN1's involvement in DNA repair, disease, substrate specificity and comparison to FEN1 shall be undertaken before outlining the research aims and objectives of this thesis.

1.8.1 The role of FAN1 in the FA dependent and FA independent pathways

The FA-pathway is the primary route of repair for ICL's during DNA replication events and can be separated into four main events, recruitment of the FA core complex (FANCA, B, C, E, F, G, L and M), formation and ubiquitination of the FANCD2-FANCI (ID2 complex), recruitment of downstream nucleases to unhook/remove the ICL and finally repair of the DNA duplex via homologous recombination (Figure 1.12) [80]. At the site of ICL damage, the MCM 2-7 helicase complex stalls approximately 20-40 nucleotides from

the lesion as a result of steric hindrance and is ejected from the two stalled replication forks [81]. Initiation of the FA-pathway begins through binding of the FANCM-FAAP24-MHF1-MHF2 hetereo-tetrameric complex to the lesion. This sub complex accompanied by RPA acts as a docking site for the newly formed FA core complex and activates the ataxia-telangiectasia and Rad3-related kinase pathway (Figure 1.12) (ATR) [82]. From this, numerous phosphorylation events occur on both the FA core complex (FANCM, A, G & E) and the nuclease recruiter complex FANCD2-FANCI (ID2 complex) [83]. The ID2-complex is mono-ubiquitinated by FANCL, reasons being that many scaffold proteins and nucleases involved in unhooking and removal of ICL's from the DNA damage site possess Ubiquitin binding zinc finger domains (UBZ). Therefore, the ubiquitinated ID2-complex is important in recruiting downstream nucleases to resolve the lesion (Figure 1.12) [83], [84]. These downstream effectors include the scaffold protein SLX4 and various endonucleases such as XPF-ERCC1, MUS81-EME1, SNM1A and FAN1 which work collaboratively to cleave the 3' and 5' sides of the ICL; the unhooked ICL is subsequently bypassed by translesion synthesis polymerases (pol-ζ and pol-ν) resulting in full repair of one DNA duplex (Figure 1.12) [85]–[87]. Consequently, this causes a double-strand break on the partnering duplex and through strand invasion of the TLS repaired duplex, strand invasion occurs to the previously repaired duplex and the secondary duplex is repaired by homologous recombination (Figure 1.12) [81], [84].

FAN1 is one of several nucleases recruited during ICL repair. It was originally assumed that FAN1 functioned exclusively within the FA-pathway due to its association to ICL's through its UBZ domain binding to the mono-ubiquitinated ID2 complex, supported by findings displaying FAN1 deficient cells increased sensitivity to ICL damage, therefore suggesting the interaction with ID2-complex is critical for FAN1 recruitment to crosslinked DNA [88]. However, recent *in vivo* findings have provided alternative evidence where in fact FAN1 can incise ICL's independently of the FA-pathway. Firstly, UBZ-mutated human FAN1 was able to fully rescue ICL repair activities in KIN patient cells. Similar findings were also observed in mouse embryonic fibroblasts (MEF) cells where UBZ-defective human FAN1 rescued ICL repair activities whereas mutations to the nuclease and DNA binding domains of FAN1 perturbed activity [89], [90]. FAN1-mediated ICL repair was also observed in mouse and human cells deficient of FANCD2, collectively implying that FAN1 can migrate and incise ICL-containing

substrates independent of the UBZ domain or ID2-complex provided the FAN1 DNA binding and nuclease domains are unaltered [90].

Reports have suggested that FAN1 plays alternative roles in DNA repair independent of direct ICL unhooking through replication fork cleavage. Studies have shown two distinct binding sites on the N-terminus of FAN1 known as the MIP box (S126, Y128 and F129) and the MIM-box (L155 and L159) which affords binding to the mismatch repair protein MLH1 [91]. Although this interaction is largely dispensable by traditional mismatch repair, defective FAN1-MLH1 complexes promote hypersensitivity to DNA crosslinking agents and CAG/CTG repeat structures observed in Huntington's disease (HD) [91], [92]. Lastly, it is claimed that FAN1 may participate in downstream ICL repair through coupling to homologous recombination. *In vivo* assays using HEK293 cell lines presented a near complete loss of homologous recombination repair activity in RAD51-deficient cell lines in response to ICL damage, a protein that promotes homologous ssDNA invasion of the sister chromatid, a process required to initiate repair of double strand breaks during homologous recombination repair [88], [93]. However, FAN1-deficient HEK293 presented a 30 % decrease in repair efficiency, implying FAN1 is involved in homologous recombination repair although its exact role in the pathway is still currently undefined [88].

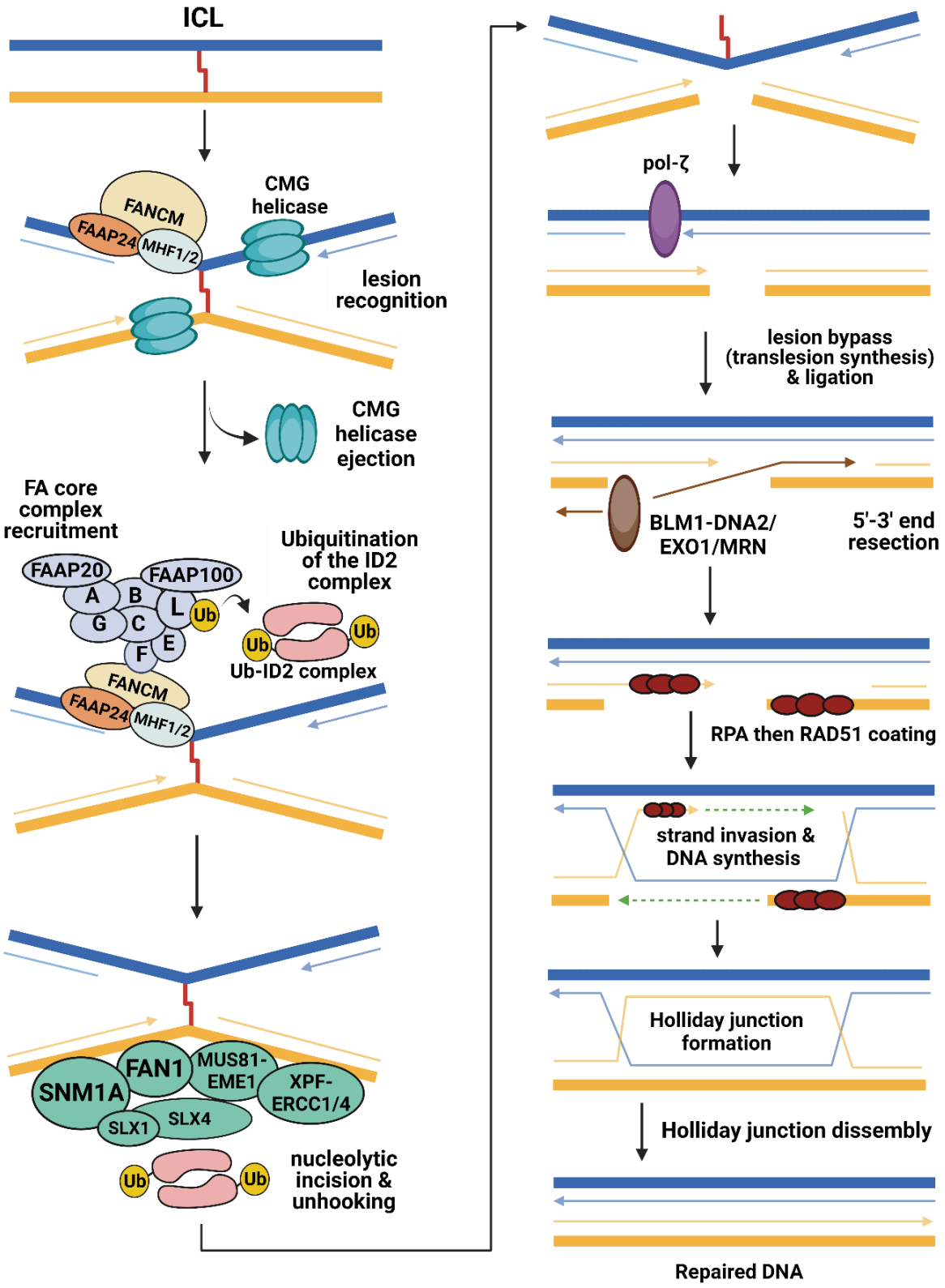


Figure 1.12 FA-dependent pathway in ICL repair

Schematic illustration of FA-dependent ICL repair. Dual replication fork stalling on either side of the ICL promotes lesion recognition by the FANCM complex and disengaging of the CMG DNA helicase. The FA core complex is recruited where FANCL ubiquitinates the ID2 complex. The Ub-ID2 complex recruits multiple 5'- and 3'-nucleases to incise and unhook the ICL. Translesion synthesis incorporates pol-ζ to repair one daughter strand (blue arrow). Homologous recombination begins by 5'-end resections to produce 3'-overhangs for strand invasion of the homologous parental strand, which is coated with RAD51, producing a D-loop. Daughter strand synthesis is performed (Yellow arrow) before dissociation of the D-loop, resulting in fully repaired and replicated DNA. This figure was created using Biorender.com.

1.8.2 Disease association to FAN1 deficiency

Deficiencies in FAN1 expression levels , and nuclease activity has been linked to multiple diseases including karyomegalic interstitial nephritis (KIN), Huntington's disease, autism and schizophrenia with FAN1 defective variants highly linked with KIN and Huntington's disease [94]. KIN is a rare genetic kidney disorder where patients present a slow progressing kidney disease resulting in complete renal failure by early adulthood [95]. A report demonstrated FAN1 deficiency in 8 out of 12 KIN patient cells and is in accord with mouse models whereby nuclease domain mutations of FAN1 result in mild KIN symptoms. Considering the high expression levels of FAN1 in kidney neuronal, and reproductive tissues, it is implied that FAN1 function aids to combat KIN disease. [89], [96]. FAN1 expression levels have also been associated with Huntington's disease, a neurodegenerative disorder caused by CAG repeat expansions in the Huntington gene HTT. Genome-wide and transcriptome association studies have proposed that increased FAN1 expression decreases the rate of Huntington's disease progression by stabilising and resolving CAG repeat expansions in the HTT gene on exon 1 [97]. Similarly, other reports have claimed that FAN1-deficient mouse models result in decreased protection against CGG repeat expansions within the gene Fmr1, a gene related to Fragile X-related disorders [98]. Finally, exome and genome sequencing studies have suggested FAN1 missense mutations to be associated with both autism and schizophrenia and was further justified by a homozygous knockout mouse model of chromosome 15q.13.3 where the FAN1 gene resides, resulting in phenotypes concordant with conditions including autism, schizophrenia and epilepsy [94], [99].

1.8.3 FAN1 substrate specificity

As previously discussed, FEN1 possess a well-documented substrate. In contrast, many substrates have been postulated to be suitable for FAN1, with preference for double flapped substrates as the literature is evolving (Figure 1.13). Therein, an exploration of FAN1's nuclease activity against varying substrates would be attractive towards it functional and mechanistic characterisation. Since its initial discovery in 2010, reports have demonstrated FAN1's ability to process dsDNA, bubbles/D-loops, nicks, ICL's, and 5'-flap secondary structures [88], [94], [100]–[102]. FAN1's nuclease activity was originally demonstrated using FAN1-plasmid DNA reactions. This resulted in open circular and linearised products before the substrate was consumed entirely, implying FAN1 possesses both endonuclease, and exonuclease activity. FAN1 also exhibits minor 5'-exonuclease activity on DNA bubbles and 3'-flap-containing D-loops whereas endonuclease activity is primarily observed in D-loops containing a 5'-flap, suggesting FAN1's potential involvement in DNA replication/transcription and homologous recombination respectively [88]. FAN1 also performs 5'-3' exonuclease activity on nicked and gapped substrates, resulting in 3' overhangs, these overhangs are routinely found in double strand break repair, further alluding to FAN1's potential participation in pathways such as homologous recombination where D-loops are formed by the invading 3'-overhang [100], [103]. On the other hand, the FAN1 nuclease domain contains a helical insertion within the nuclease domain between β2 and β3, subsequently changing the FAN-DNA nuclease domain interface, resulting in limited FAN1 activity towards Holliday junctions, another DNA intermediate observed in homologous recombination repair [101].

Multiple reports have identified FAN1 to preferentially catalyse branched DNA structures including splayed arms and replication forks although it is suggested that FAN1 favours 5'-endonuclease activity on 5'-flaps *in vitro* [88], [90], [100], [101]. Furthermore, FAN1's 5' endonuclease activity is conserved amongst human, mouse and bacterial homologs, with incisions of approximately 4-5 nucleotides into the 5'-duplex of the 5'-flap strands. After the initial incision, all three homologs proceed to incise the 5'-duplex. Following this, bacterial FAN1 traverses to the opposite strand of the 5'-duplex and performs 5' endonuclease incisions of the 5'-duplex, which has not been reported in mammalian FAN1 proteins [101].

The nuclease activity of FAN1 is influenced by the number of flaps in the substrate; the length of the flaps, and the presence of the 5'- terminal phosphate on the processed strand [102]. For example, FAN1 produced higher affinity binding to double flap substrates of shorter 5'-flaps (bearing a 5'-terminal phosphate) and longer 3'-flaps compared to nicks, gaps, or single flapped substrates, potentially suggesting longer 3'-flaps contribute to optimal FAN1-DNA binding. Additionally, FAN1 readily binds gapped substrates however increasing the ssDNA gap length to 35 or more nucleotides substantially reduces binding. Moreover, these substrates are bound with significantly lower affinity in the absence of a 5'-terminal phosphate, implying the 5'-phosphate's integral role in human FAN1-mediated catalysis [102]. Nuclease activity assays produced concordant results with the binding data; substrates bearing longer 5'-flaps diminished activity independent of ionic strength [102]. Interestingly, FAN1 produces varying product lengths depending on the secondary structure of the substrate. Double flap substrates with a 1 or 2 nucleotide 5'-flap and 8-nucleotide 3'-flaps result in an incision pattern of N+3 whereas single 5'-flaps and nicked substrates operate under N+4 and N+2 registers on respectively. FAN1's nuclease activity has also been documented on ICL-containing branched DNA structures. For example, mammalian and bacterial FAN1 homologs display robust nuclease activity against branched DNA structures with ICLs situated in the 5'-duplex. Both homologs make incisions on the 5' and 3' of the ICL lesion and the efficiency of the FAN1's ability to unhook the ICL increases the closer the lesion is the ss/dsDNA junction or 5'-flap terminal phosphate [102], [104].

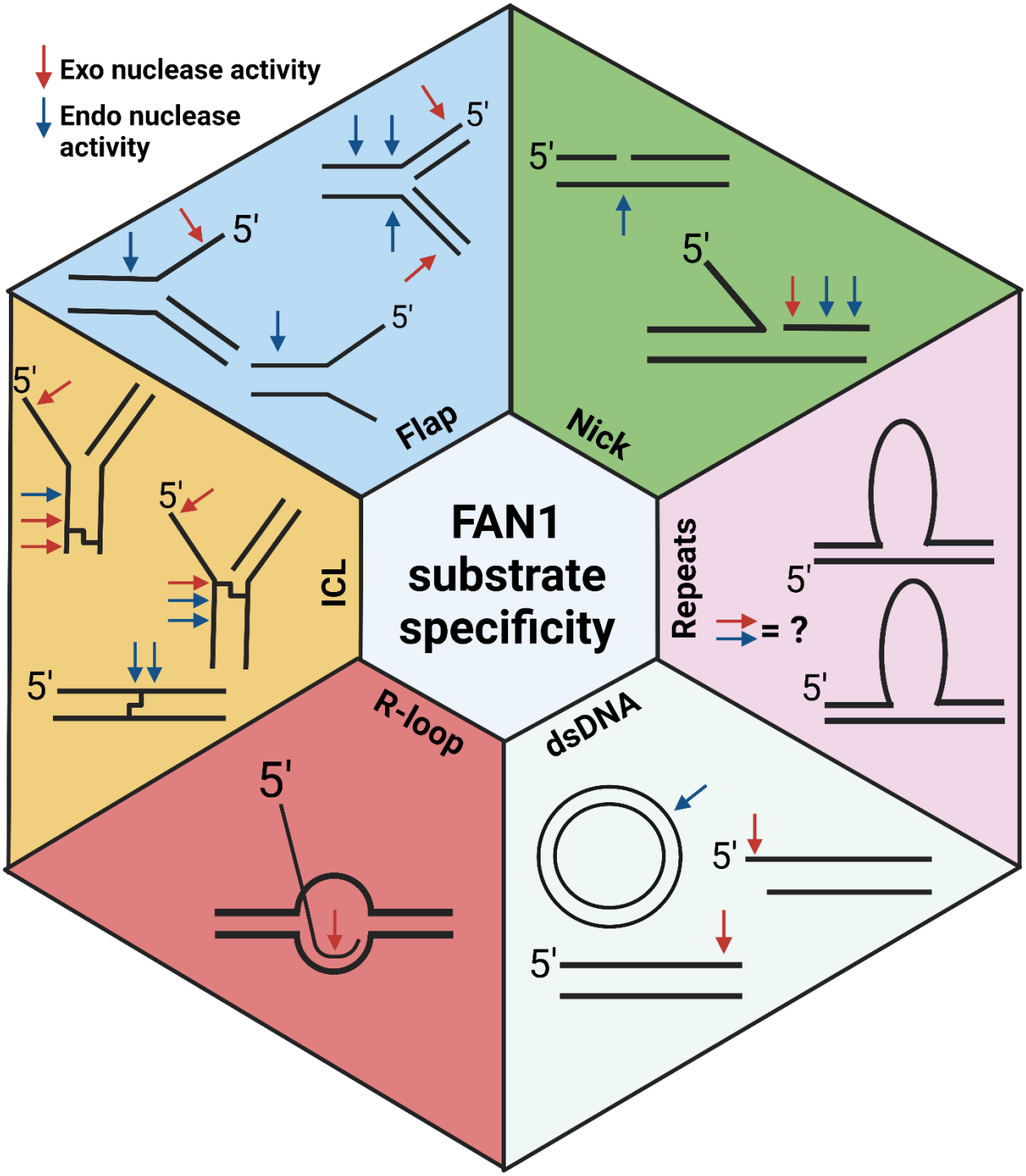


Figure 1.13 FAN1 substrate specificity

Reported nuclease activity of FAN1 across multiple DNA substrates. This image is modified from the following reference [94].

1.8.4 FAN1 structure

Three FAN1-DNA structural complexes have been reported, two involving human FAN1 and one *Pseudomonas aeruginosa* FAN1 (PaFAN1) [102], [105], [106]. The FAN1 protein (lacking the UBZ domain) is approximately 1017 residues in length (~75 kDa) and consists of 25 α-helices and 10 β-sheets [106] (Figure 1.14A). Furthermore, the FAN1 structure bears a bi-lobed morphology distinguishing the N-, and C-termini. The N-terminus contains 223 residues and consists of an N-terminal helical bundle and an A/B Acinus and Pias (SAP) DNA binding domain whereas the C-terminus consisting of 415 residues houses the tetratricopeptide (TPR) protein-protein interaction domain and a viral replication and repair (VRR) nuclease domain [106] (Figure 1.14A &B). All reported structures show FAN1 to resemble a glove-like structure where the SAP and helical domain make up the thumb and wrist region whilst the VRR nuclease and TPR protein interaction domains contributes to the fingers. Structural studies have also presented a shallow basic groove approximately 30 Å wide and 20 Å descending into the protein between the VRR nuclease and TPR domains, proceeding perpendicularly into the N-terminus through to the SAP domain, [102], [105], [106]. Despite FAN1's proposed involvement in ICL repair, no structures of FAN1 in complex with ICL-containing substrates have been achieved. Therefore, discussions regarding FAN1-DNA contacts on the atomic level will be described using structures containing double flapped substrates and nicked structures with a 5'-flap.

1.8.4.1 N-terminal helical bundle and the SAP domain

Binding of the 5'-duplex in human and bacterial FAN1 structures have been demonstrated for both double flap and nicked substrates. In structures containing double flapped substrates, the helical domain makes multiple hydrogen and electrostatic contacts to the 3'-flap through residues Y374, V577 and R581 with additional contacts formed at the SAP domain via Q492 and K493 to the template strand (Figure 1.14C &D). Furthermore, the phosphodiester of the second nucleotide enters a shallow groove and forms a hydrogen bond to the main chain amide group of Y374 [102]. The bacterial FAN1-DNA complex with a 5'-flap and nicked substrate offers more insight into the contacts made towards the 3'-duplex. It is suggested that the SAP domain of FAN1 stabilises DNA binding via hydrogen bonding of basic residues to the phosphodiester backbone of the template strand. This structure also provides an explanation as to how junction bending occurs, and is facilitated by hydrophobic side chain residues facing towards the paired bases at the 3'-duplex whilst other residues including W184 and Y40 position against the 5'-flap and template phosphodiester at the junction respectively to ensure adequate bending [87]. Collectively, these residues co-ordinate to form a hydrophobic wedge which is consistent with other flap-endonucleases [45].

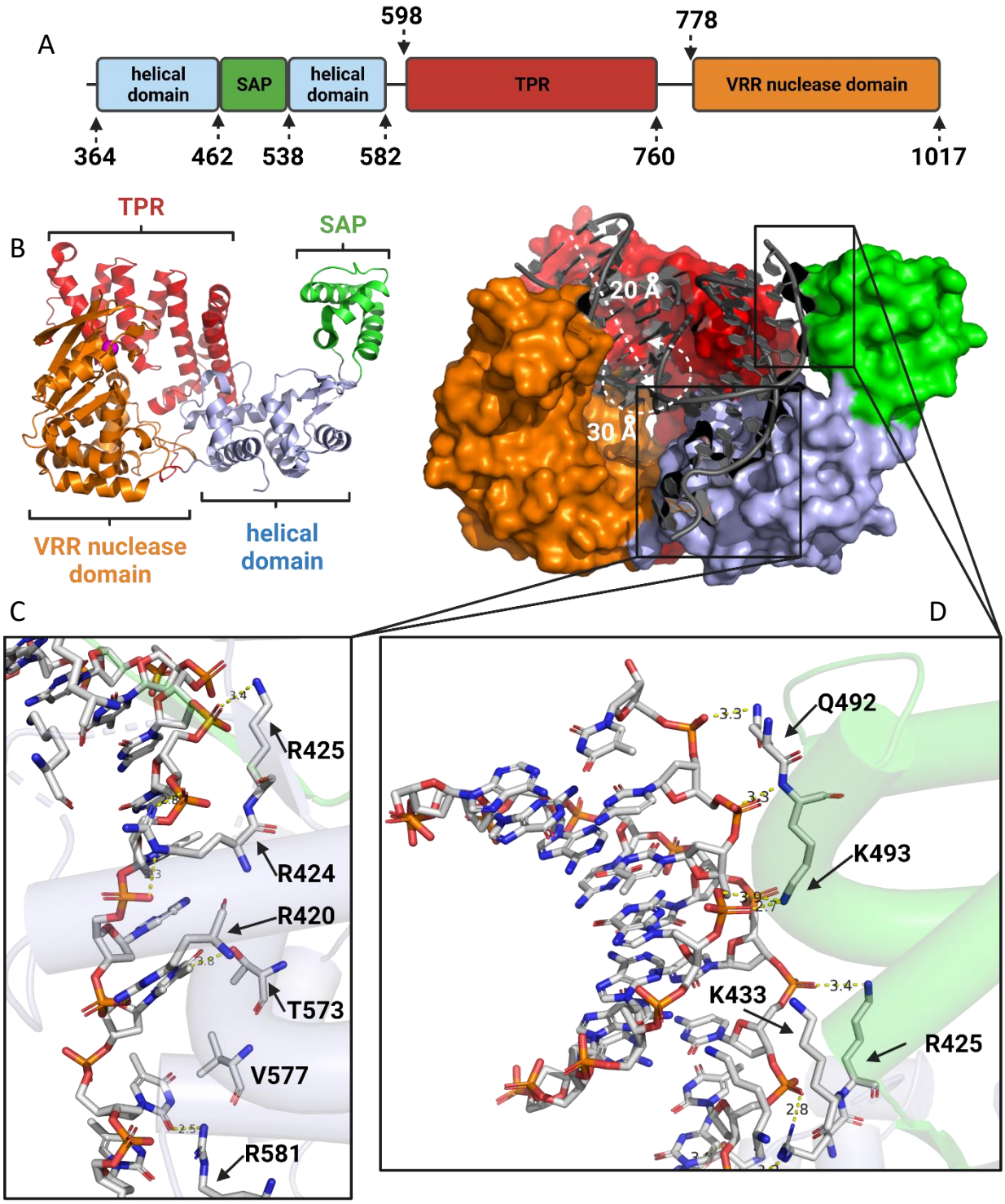


Figure 1.14 N-terminal SAP and helical domain interactions of FAN1

A*. Amino acid and domain organisation of human FAN1.* ***B****. 3-dimensional orientation of FAN1 (PDB: 4RI8) following the same colour coding as A.* ***C****. Magnified residue contacts at the helical domain and the 3'-flap strand; carbons (grey); nitrogen's (blue); oxygen's (red); phosphorous (orange).* ***D****. Magnified residue contacts between the SAP domain and the template strand.*

1.8.4.2 C-terminal basic binding pocket

In human FAN1-DNA structures, a cluster of basic residues form a pocket which was proposed to select for the 5' phosphate of a 1-nucleotide 5'-flap (Figure 1.15A & B). The 5'-flap to extends orthogonally away from the 5'-duplex and enters the basic pocket where hydrogen bonds/electrostatic interactions are formed between the side chain amine groups of R706, R952, H742 and K986 to the three non-bridging oxygens of the 5' terminal phosphate. The following two phosphodiesters in the processed strand display no direct contacts with human FAN1 until the third phosphodiester encounters the Ca^{2+} ion which replaces Mg^{2+} ions to perturb nuclease activity in the active site (Figure 1.15B) [102]. However, the literature lacks functional data as to human FAN1 basic pocket residue mutations to alanine and the subsequent effect on FAN1-mediated catalysis. In contrast, the basic pocket of human FAN1 is not conserved in its bacterial homolog. Instead, a separate basic pocket is located at the site that converges the nuclease, TPR and helical domains [84]. In this structure the 5' strand base displaces from the 5'-duplex and forms a U-shaped conformation with the 5'-flap. A tryptophan residue in the path of the 5'-flap steers the second and third nucleotides of the 5'-flap into the basic pocket where a hole resides, that permits entry of ssDNA. However, the phosphodiester of the first nucleotide negates entry into the hole of the basic pocket and instead protrudes outwards in the opposite direction into the open cavity shared between the N- and C- terminal domains (Figure 1.15C) [106]. Collectively these reports illustrate the substrate flexibility across FAN1 homologs and how the 5'-terminal phosphate is especially important in human FAN1 substrate specificity.

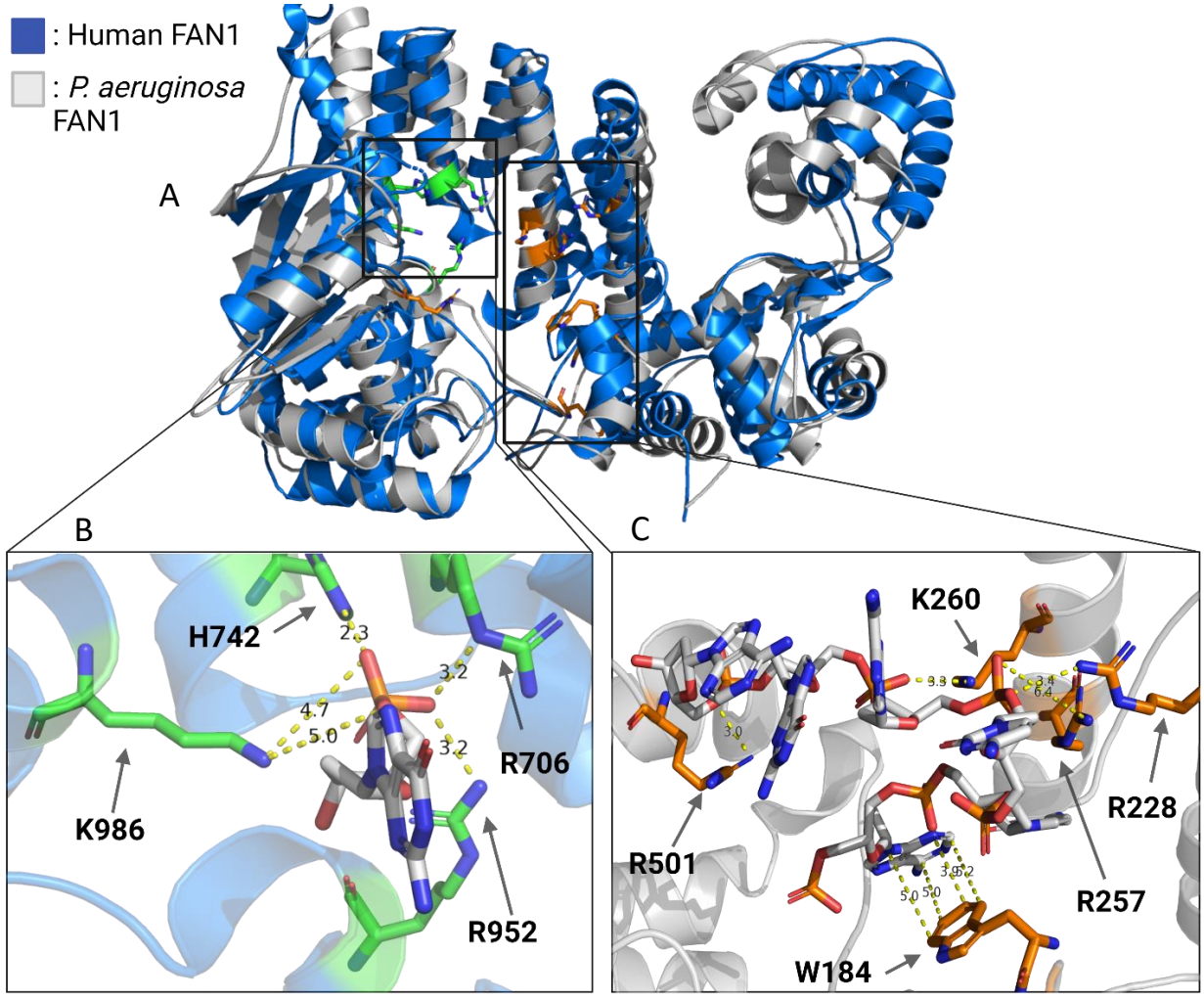


Figure 1.15 5'-flap binding pocket of human and bacterial FAN1

***A**. Superimposed human (blue) and bacterial (grey) FAN1 monomers. **B**. Magnified image of the human 5'-flap basic pocket contacting the 5'-terminal phosphate of a 1-nucleotide 5'-flap (PDB code: 4RI9). Human FAN1 residue colour coding is as follows: Carbons (green); nitrogen's (blue) and DNA is colour coded: carbons (grey); oxygen's (red); nitrogen's (blue); phosphorous (orange). **C**. Magnified image of bacterial FAN1 complexed with a 4-nucleotide 5'-flap (PDB code: 4R89) contacting residues in the lower cavity of the protein. Bacterial FAN1 residue colour coding is as follows: Carbons (orange); nitrogen's (blue) and DNA is colour coded: carbons (grey); oxygen's (red); nitrogen's (blue); phosphorous (orange).*

1.8.4.3 Active site

FAN1 phosphodiesterase activity is facilitated by the active site core within the VRR nuclease domain. This domain contains a PD-(D/E)XK motif and is conserved in multiple endonucleases across bacteria and human model systems [107], [108]. Furthermore, domain overlap of the FAN1 nuclease core has suggested high similarities to the nuclease domain of the Holliday junction resolvase in *Pyrococcus furiosus* [105], [109]. Because of this, FAN1 nuclease activity is undertaken using divalent metal ions such as Mg^{2+} of Mn^{2+} and shows no catalytic activity in the absence of metals or in the presence of Ca^{2+} ions, similar to other endonucleases including members of the FEN superfamily, topoisomerases and RNA processing enzymes [110]. In human FAN1, a single non-catalytic Ca^{2+} ion is co-ordinated by the two carboxylate residues D960 and E975. With relation to the processed strand, the metal ion shows proximity contact to the non-bridging oxygen of the third phosphodiester with additional hydrogen bonding observed between the side chain of G957 and the non-bridging oxygen of the second phosphodiester located on the processed strand in the presence of an 1-nucleotide 5'-flap (Figure 1.16) [102]. Juxtaposed to human FAN1, bacterial FAN1 catalyses its substrate using two divalent Mg^{2+} or Mn^{2+} ions inside the active site core. Metal ion A is coordinated by D507, E522 and V523 and metal ion B is coordinated by E386 and D507 with 3.8 Å between both metal ions. The metal ions contact the non-bridging oxygen of the phosphodiester on the fifth nucleotide on the processed strand at the 5'-duplex (Figure 1.16) [106]. Overall, discrepancies exist in the literature between human and bacterial structures as to whether FAN1-mediated catalysis is supported by 1, or 2 metal ions.

The comparative difference in metal ion positioning in human and PaFAN1 offers possible explanation as to why bacterial FAN1 incises further into the 5'-duplex compared to human FAN1 which performs successive 3-nucleotide incisions from the end of the 5'-flap [84]. Finally, FAN1 exon Sanger sequencing data in KIN patients has also suggested the importance of the metal-coordinating ion D960 (D507 in bacterial FAN1) whereby 8 out of 12 KIN patients possessed missense mutations in the nuclease domain of FAN1, specifically, D960N, Q929P and G937D. Because of these findings, it is claimed that the loss of FAN1 nuclease activity leads to direct cause of multiple kidney diseases [89].

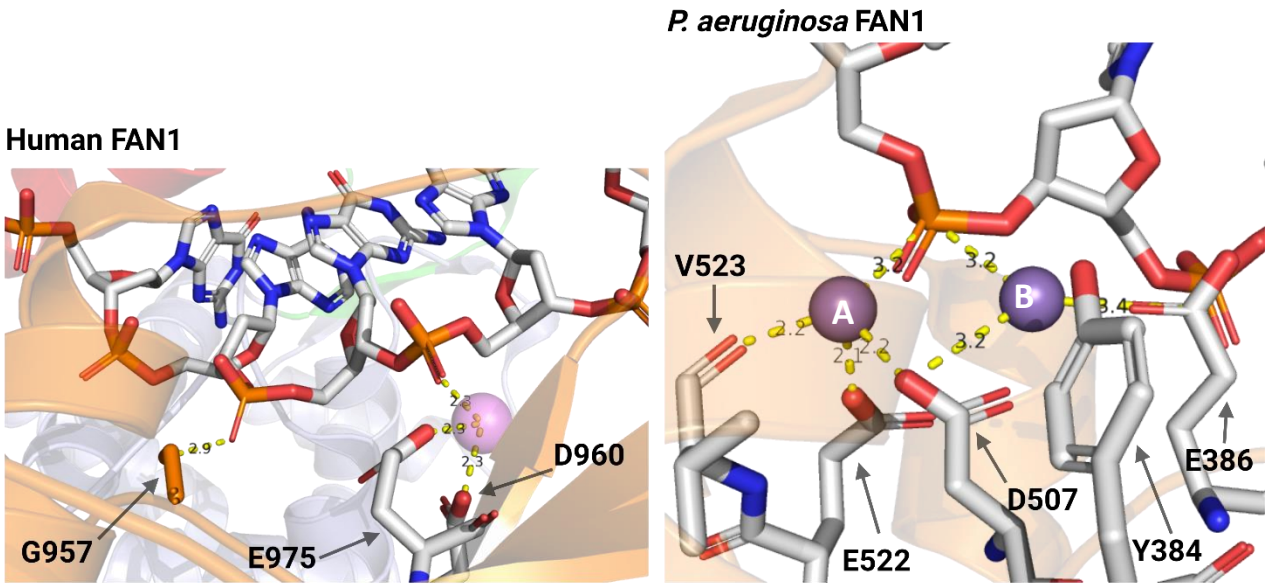


Figure 1.16 FAN1 active site

Active site core residues involved in human and P. aeruginosa FAN1. Carbon atoms are shown in grey; oxygen atoms shown in red; phosphorous atoms are shown in orange.

1.8.5 Proposed models of dimeric FAN1

The interactions between a protein and its substrate are vital for catalysis although protein-protein interactions can also play an important role in facilitating catalysis. Proteins acting alone in solution are considered monomeric, whereas a complex of two identical proteins during the reaction is homodimeric [111]. Crystallographic evidence suggested that human FAN1 catalyses DNA substrates in the monomeric state (section 1.8.4) [102]. In contrast, recent crystallographic and biochemical studies have presented evidence for an asymmetric homodimer of human FAN1 in the presence of DNA containing a 5'-flap, and suggests a model in which FAN1 'scans', 'latches' and 'unwinds' the substrate before catalysis (Figure 1.17A) [105]. The FAN1 homodimer exhibits a head-to-tail arrangement where the primary FAN1 (^{P}FAN1) hydrolyses the 5'-flap and the secondary FAN1 (^{S}FAN1) binds the 3'-duplex of the substrate. Dimerisation is driven by the basic residues K525, R526 and K528 on α9 of the ^{S}FAN1 SAP domain and is further supported through slight conformational adjustments of α7 and α5-β1 to accommodate varying dimeric conformational states of the complex (Figure 1.17A). The phosphodiester backbone of the 5'-duplex forms multiple electrostatic contacts with the ^{P}FAN1 SAP domain (residues K482, N490, Q492, K493, Q494) and helical domain (residues R420, R424, K425, K433) (Figure 1.17B). The importance of these residues to FAN1 activity were demonstrated when they were substituted to glutamic acid, resulting in severely diminished DNA binding,

and endonuclease activity. Additionally, hydrogen bonding is also witnessed between residues Y374 and Y436 to the 5'-duplex nearer the nuclease domain ^{P}FAN1 in which mutations to phenylalanine had no effect on binding, but significantly impaired endonuclease activity. Thus, alluding to their specific role in determining a catalytically competent FAN1-DNA complex prior to catalysis. Finally, hydrogen bonding is also observed on the helical domain of ^{S}FAN1 and the phosphodiester backbone of the 3'-duplex with residue ^{S}K401 and upon latching, and unwinding of the DNA, further contacts are made at this region to the 3'-duplex through ^{S}K409 and ^{S}Q406. All of which again led to substantially decreased endonuclease activity when mutated to glutamic acid (Figure 1.17B) [105].

This structural data led to researchers proposing a DNA interaction model consisting of three key events: substrate 'scanning', substrate 'latching' and substrate 'unwinding'. During substrate scanning, the 5'-duplex binds the ^{P}FAN1 SAP and helical domain residues with no additional contacts made to the 5'-duplex by the helical domain of ^{S}FAN1. The substrate is then latched to the homodimer through minor conformational rearrangements of ^{S}FAN1, affording a 7 Å translocation of the ^{S}helical domain towards the 3'-duplex. Finally, substrate unwinding is undertaken to position the 5'-flap in proximity of the ^{P}FAN1 catalytic metals in the active site. Following these events, it is claimed that FAN1 achieves a catalytically competent state and 5'-flap endonuclease incision is facilitated [84], [105]. Although insightful, the model fails to describe the series of conformational and structural events involved in product release in preparation for the binding a new substrate. FAN1 dimerisation has also been reported in bacterial FAN1 homologs [112]. FAN1 from *Vibrio vulnificus* (VvFAN1) resides in the dimeric state in the presence or absence of substrate although *Pseudomonas aeruginosa* FAN1 (PaFAN1) is monomeric. Both proteins exhibited efficient endonuclease activity against short 5'-flaps and decreased endonuclease activity as the 5'-flap was extended to 15, and 40 nucleotides with PaFAN1 showing the most substantial loss in nuclease activity, implying FAN1 dimerisation increases FAN1's substrate flexibility during DNA repair events. Furthermore, nuclease dimerisation is also observed in other DNA repair nucleases including MUS81-EME1 and MRE11, both of which are involved in resolving Holliday junctions and 3-5' exonuclease activity on damaged DNA respectively [113], [114]. Therefore, it is not out of the realms of possibility that FAN1 follows the same dimeric trend as its associate nucleases during DNA repair events.

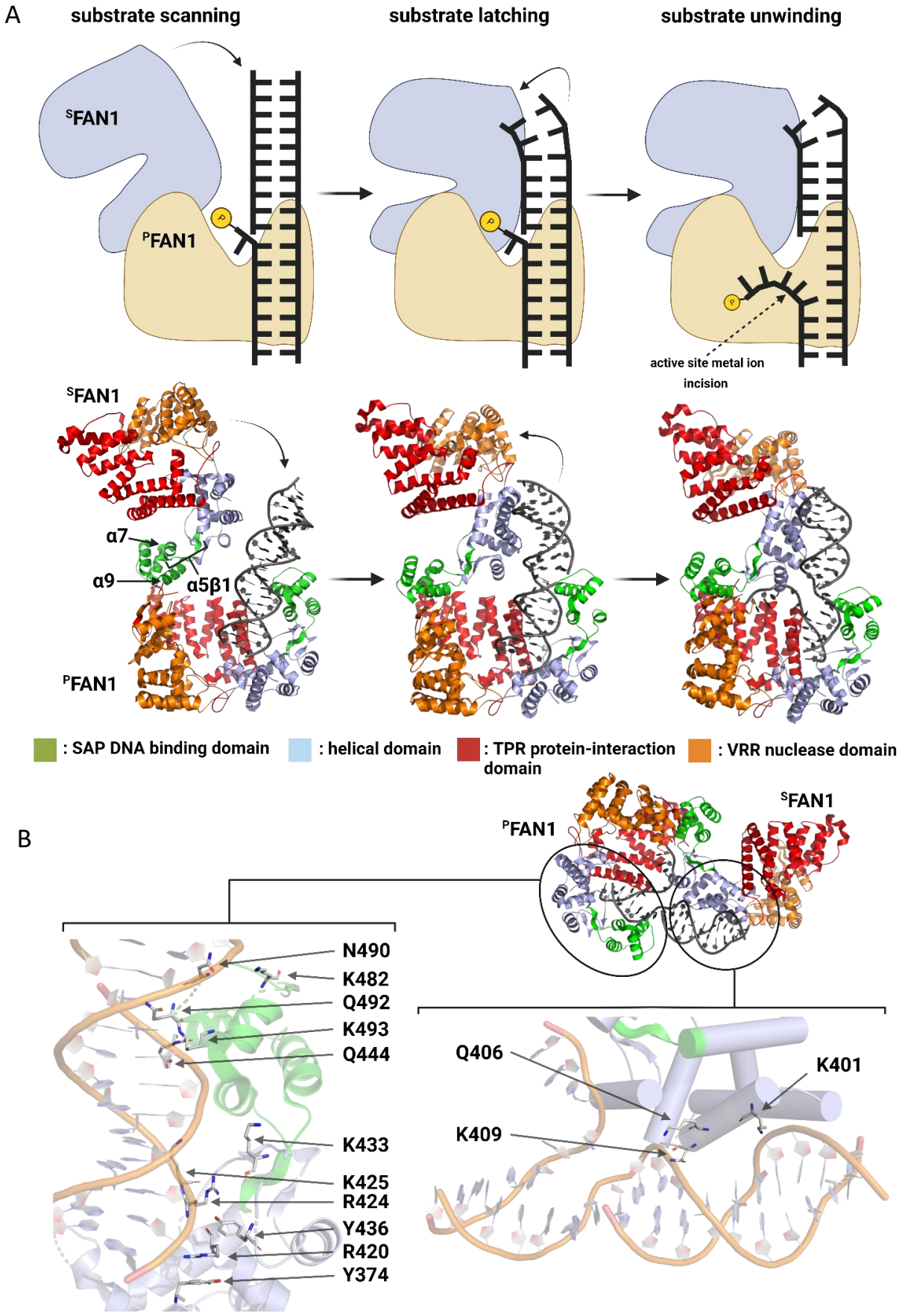

A
substrate scanning
substrate latching
substrate unwinding
SFAN1
PFAN1
active site metal ion
incision
α7
α9
α5β1
: SAP DNA binding domain
: helical domain
: TPR protein-interaction domain
: VRR nuclease domain
B
N490
K482
Q492
K493
Q444
K433
K425
R424
Y436
R420
Y374
Q406
K401
K409

Figure 1.17 Structure and proposed model of dimeric FAN1 function

***A**. Schematic illustration of the proposed FAN1 dimer mechanism. Dimeric human FAN1 traverses the substrate (substrate scanning) making contacts between the ^{P}FAN1 SAP/helical domains (PDB code: 4REC). After scanning ^{S}FAN1 conformationally changes and latches to the phosphodiester backbone of the 3'-duplex (substrate latching) (PDB code: 4REB). The ^{P}FAN1 monomer then unwinds the 5'-flap from the 5'-duplex in preparation for catalysis (substrate unwinding) (PDB code: 4REA). **B**. Magnified image of ^{P}FAN1 SAP/ helical domain residue contacts to the 5'-duplex phosphodiester backbone and ^{S}FAN1 helical domain residue contacts to the 3'-duplex phosphodiester backbone. FAN1 domains are colour coded as shown in the figure.*

1.8.6 Comparison of FEN1 and FAN1 properties

Until now, FEN1 and FAN1 have been discussed separately. Although similarities and differences arise when the two enzymes are compared. Despite the structural difference between the two enzymes, they both employ similar methods in substrate binding. In the presence of branched structures, both enzymes incorporate distinct DNA binding domains to contact the 5' and 3'- duplexes [20], [102]. Moreover, both enzymes profoundly kink the DNA at the ss/dsDNA junction through hydrophobic helical wedges and by doing so, the 5'-flap strand is isolated and subsequently positioned towards the active site. Overall, hydrophobic helical wedges and duplex bending is conserved amongst several 5' and 3' nucleases, indicating its importance in facilitating recognition and catalysis [20], [113], [115]. A major difference between the two enzymes is the degree of domain ordering and conformational changes required to achieve catalytic competency. Reports surrounding monomeric FAN1 postulate a preformed DNA binding sites and active sites. However no detailed biophysical evidence is currently available to support this claim. It is also postulated that dimeric FAN1 also exhibits an ordered state although conformational changes are required to perform the 7 Å translocation towards the 3'-duplex before DNA unwinding [84], [102].

On the other hand, FEN1 is largely disordered in the absence of substrate, 3'-flap binding drives conformational ordering of the helical arch onto the pre-threaded 5'-flap ssDNA and manoeuvres the scissile phosphate during active site transfer [48], [69]. This results in FEN1's ability to achieve extremely high substrate specificity and accurate +1-1 scissile phosphate hydrolysis required to resolve Okazaki fragments during DNA replication [20], [46], [116]. Therefore, the lack of conformational changes in monomeric FAN1 possibly explains its flexible substrate specificity compared to FEN1 [88]. It should be noted that minor conformational changes are observed in the dimeric FAN1-DNA complex however this is only to make contacts with the 3'-duplex, most likely offering stability throughout DNA unwinding [105]. In summary, the

importance of FEN1 and FAN1 is clearly presented through their direct links to disease. Up-regulation of FEN1 results increased susceptibility to a varying cancers whilst the abolishment of FAN1 nuclease activity confers hypersensitivity to ICL damage and KIN disease [96], [117], [118]. In that respect, it proves desirable to uncover the mechanism of FAN1-mediated catalysis as a potential therapy target.

1.9 Aims of the project

In this thesis, attentions will first be drawn to FEN1, a well-researched enzyme with a wealth of literature surrounding its function. It is widely accepted that protein and DNA conformational change are critical in efficient FEN1-mediated catalysis addressed in numerous structural, kinetic, and biophysical studies. A recent report by *Song et al* claimed that DNA conformational change can be observed and is rate-limiting in the mutant D34A. This study used tandem 2-AP substitutions in the +1-1 base positions of the processed strand and observed a burst in fluorescence believed to occur prior to reaction. Therefore, the first aim of this thesis is to replicate the findings of *Song et al* with our less severe active site mutant D34N and to potentially apply this DNA conformational change assay to other proteins.

The majority of *in vitro* FEN1 research focuses on the 3'-flap binding site, helical arch/gateway and the active site core residues as the majority of DNA contacts and reaction driving conformational changes occur in this region. In contrast, less attention has been afforded to the role of the FEN1 H2TH:K^+ domain. The basic environment of this domain (residues R239, K244, R245) promotes hydrogen bonding and electrostatic contacts to the template strand and thus, questions are raised to whether removal of this basic environment affects the rates of the FEN1 reaction. Additionally, it remains enigmatic whether basic residues in the H2TH:K^+ domain affect conformational changes at the +1-1/-1-2 base positions required for active site transfer and ultimately FEN1 catalytic competency. Consequently, the second aim is to perform biochemical kinetic studies on FEN1 H2TH mutants R239A, K244A and R245A to measure changes in multiple and single turnover rates compared to wild-type (WT). ECCD assays will measure changes in DNA conformation when the substrate is complexed with the H2TH mutants using adenine mimicking fluorophores at the +1-1/-1-2 base positions.

When compared to FEN1, FAN1 is significantly understudied. Structural studies have suggested conflicting models of FAN1-mediated catalysis regarding it's monomeric, or dimeric state. It is possible that human FAN1 can function under both states dependent on substrate structure, however definitive evidence confirming this has not yet been achieved. Nuclease and binding assays have proven useful in understanding human FAN1's ability to process a myriad of DNA secondary structures with preference to double-flapped substrates bearing a 1-nucleotide 5'-flap and 8-nucleotide 3'-flap although currently, questions remain unanswered. Firstly, the literature has proposed a plethora of expression and purification strategies of FAN1 however evidence regarding purity and yield has not been disclosed. Furthermore, single, or multiple turnover kinetic parameters of FAN1 substrates are not currently available and thus, limiting our knowledge of the FAN1 reaction mechanism. Therein, FAN1 characterisation will begin by establishing a minimal-step, cost-effective expression and purification strategy in *E.coli*. Following this, buffer and additive optimisation experiments will increase the stability of homogenous FAN1 in solution, resulting in limited sample degradation and proteolysis. The second aim is to characterise FAN1 through nuclease assays against a variety of single and double flap substrates to confirm the number of products formed as described in the literature. Finally, following HPLC gradient testing of FAN1 reaction products, single and multiple turnover kinetic assays will measure the change in kinetic parameters of varying FAN1 substrates.

Chapter 2 Materials and Methods

2.1 WT & mutant FEN1 construct cloning and protein expression

Buffers and reagents for the WT/mutant FEN1 expression can be found in Table 1.

Table 1: Buffers, medias, and gels.

2XYT media (1 L)	16 g tryptone, 10 g yeast extract, 5 g NaCl, pH 7.5 with 10 M NaOH
SOC media (1 L)	20 g Tryptone, 5 g yeast extract, 0.6 g NaCl, 0.19 g KCl, 1.20 g $MgSO_4$, 3.60 g glucose.
LB Agar media (1 L)	10 g Tryptone, 5 g yeast extract, 10 g NaCl, 15 g agar, 50 µg/ml kanamycin, 34 µg/mL chloramphenicol
Lysis buffer (0.5 L)	500 mL of IMAC A1 buffer (20 mM Tris-Base, 40 mM Imidazole, 1 M NaCl, 5 mM , β-mercaptoethanol (2-BME), adjusted to pH 7.0 with 10 M NaOH), 0.1 % lysozyme, protease inhibitor tablet
1 ×PBS (1 L)	8 g NaCl, 0.2 g KCl, 1.44 g $NaHPO_4$, 0.24 g KH_2PO_4, pH 7.4 with 10 M KOH
50 × TAE (1 L)	242 g Tris-base, 57.1 mL glacial acetic acid, 50 mM EDTA, pH 8.0 with 10 M acetic acid and 10 M KOH
Agarose gel (50 mL)	0.4 g agarose, 50 mL 1 ×TAE, 5 µL 10,000 × SYBER safe gel stain
Acrylamide gel	Resolving (12 %): 3.4 mL H_2O, 4 ml Acrylamide/bis-Acrylamide (30 % solution), 2.5 ml resolving gel buffer, 100 µL APS (10 %), 10 µL TEMED, 50 µl TCE. Stacking (4 %): 3.4 ml H2O, 830 µL Acrylamide/bis-Acrylamide (30% solution), 680 µL stacking gel buffer, 50 µL APS (10 %), 10 µL, 5 µL TEMED.
Running buffer (1 ×)	25 mM Tris, 192 mM glycine, 0.1% SDS Dilute 100 ml 10x stock with 900 ml MilliQ water
Sample Buffer (2 ×)	62.5 mM Tris-HCl, pH 6.8, 2% SDS, 25% (v/v) glycerol, 0.01% bromophenol blue, 5% 2-βME or 100 mM DTT

2.1.1 Site directed mutagenesis & PCR amplification

The pET-28b WT FEN1 plasmid (Refer to appendices Figure 8.1) was previously produced by Grasby lab group members. The plasmid was sequenced using the GeneWiz service. The pET28-WT-FEN1 plasmid was used as a template strand in later polymerase chain reaction (PCR) reactions for site directed mutagenesis to produce R239A, K244A & R245A. To perform site directed mutagenesis, individual primers overlapping the mutated region to an alanine codon was produced using the QuickChange primer design feature by Agilent (https://www.agilent.com/store/primerDesignProgram.jsp). Primers designed for each FEN1 mutant are shown in Table 2.

Table 2: R239A, K244A and R245A primers

Primer name	Sequence
R239A-F	5' - CTT GGG CCC AAT ACC CGC GAT ACT CTC ACA GTA GTC- 3'
R239A-R	5' - GAC TAC TGT GAG AGT ATC GCG GGT ATT GGG CCC AAG - 3'
K244A-F	5' – GTC CAC AGC CCG CGC GGG CCC AAT ACC C - 3'
K244A-R	5' – GGG TAT TGG GCC CGC GCG GGC TGT GGA C - 3'
R245A-F	5' – TGA GGT CCA CAG CCG CCT TGG GCC CAA TAC - 3'
R245A-R	5' – GTA TTG GGC CCA AGG CGG CTG TGG ACC TCA - 3'

PCR amplification was performed by first producing 3 µM primer pair solutions diluted in MilliQ water. PCR mutagenesis reactions were performed using a bench-top PCR machine (BIO-RAD) with a total of 25 uL for each reaction (Table 3).

Table 3: PCR mutagenesis reaction for R239A, K244A and R245A

Reagent/Material	Volume (µL)
10 × PFU-ultra hotstart reaction buffer	2.5
PFU-ultra hotstart DNA polymerase (2.5 U/ µL)	0.5
pET28b WT-FEN1 template plasmid (145 ng/ µL)	1.0
Primer pair (3 µM)	2.5
dNTP	2.0
MilliQ water	16.5
Total	25

A program was selected to allow sufficient denaturing, annealing and elongation of the DNA sequences using the pFU ultra-hotstart polymerase reaction system with the temperatures of each stage shown in Figure 2.1. PCR reactions were then digested with 1 µL of DPN1 for 1 hour at 37 °C. Reactions were then analysed by agarose gel electrophoresis (Gels and buffer recipes shown in Table 1). Gels were imaged using SYBER safe stain and a BIO-RAD imaging camera (Image Lab).

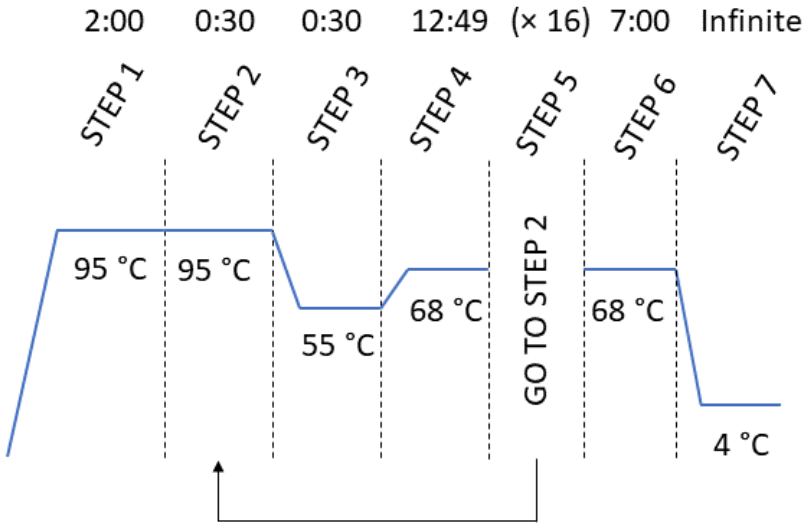


Figure 2.1: PCR mutagenesis temperature and stages

Schematic of the change in temperature for the PCR reaction of R239A, K244A and R245A. Above each stage is the time taken minutes for the stage to complete. Step 1 and 2 disrupted the hydrogen bonds between the plasmid template strands. Steps 5 and 6, annealing temperature of the primers to the template. Step 4. The preferred extension temperature for PFU Ultra hotstart polymerase. Step 4. Holding temperature following completion of the PCR reaction cycles.

2.1.2 Bacterial cell transformation

Bacterial cell transformation was performed to introduce PCR amplified DNA and other plasmid constructs into competent cell lines. 1.0-2.0 µL of DNA was pipetted into 20 µL of DH5α cells and left on ice for 45 minutes. The cells were heat shocked at 42°C for 30 seconds before being placed back on ice for 5 minutes. 1 mL of SOC media was used to rescue the cells. The cells were then incubated at 37 °C for 2 hours at 200 rpm. Cells were then centrifuged at 13,400 rpm for 5 minutes before the supernatant was removed from the mixture. The cells were resuspended in 100-200 µL of fresh SOC media before being plated and spread onto Lb agar plates containing 50 µg/ mL kanamycin for FEN1-containing constructs. FAN1-conatining constructs were plated using 100 µg/mL ampicillin and 34 µg/mL chloramphenicol. Plates were incubated overnight (~16 hours) at 37°C. Successful transformation was achieved if individual colonies were observed.

2.1.3 Plasmid preparation for sequencing

DH5α bacterial cells were used for plasmid preparation whereas BL21 bacterial cell lines were used for protein expression. DH5α cells possesses a RecA recombinase mutation, resulting in higher plasmid yield whereas BL21 bacterial cells are deficient in proteases, resulting in reduced target protein degradation. Following bacterial cell transformation, a single colony was placed into a 5 mL 2XYT starter culture containing 50 µg/mL kanamycin (FEN1 constructs) or 100 µg/mL ampicillin and 34 µg/mL chloramphenicol (FAN1 constructs) and grown overnight at 37 °C at 200 rpm. The starter culture was centrifuged at 13,400 rpm for 30 seconds and the supernatant was removed. The cell pellet was resuspended with 250 µL of ZymoPURE P1 buffer. 250 µL of ZymoPURE P2 buffer was added to the cells and gently mixed by inversion. 250 µL of ZymoPURE P3 buffer was then added to the mixture and mixed thoroughly by inversion. The neutralised lysate was centrifuged for 5 minutes at 16,000 ×g before 600 µL of supernatant was transferred to a fresh microcentrifuge tube. 260 µL of ZymoPURE binding buffer was added to the supernatant and mixed thoroughly for 15 seconds. The mixture was then placed into a Zymo-Spin II-PX column and was centrifuged at 10,000 ×g for 1 minute. 800 µL of ZymoPURE wash 1 buffer was added to the column and centrifuged again at 10,000 ×g for 1 minute. 800 µL of ZymoPURE wash 2 buffer was added to the column and centrifuged again at 10,000 ×g for 1 minute. The column was collected into a fresh collection tube and 25 µL of ZymoPURE

elution buffer was added before centrifugation at 10,000 ×g for 1 minute. 260 nm absorbance was used to identify the concentration of the purified plasmid DNA before checked for sequencing by GeneWiz.

2.1.4 Protein expression and purification of WT-FEN1, R239A, K244A & R245A

E.coli BL21 (DE3) Rosetta 2 were transformed as per the protocol described in section 2.1.1. A 50 mL 2XYT starter culture containing 25 µg/mL kanamycin and 34 µg/mL chloramphenicol was inoculated with a single colony containing the required plasmid. The culture was incubated overnight at 37 °C at 200 rpm (16-18 hours). The starter culture was seeded into 500 mL 2XYT cultures containing 25 µg/mL kanamycin and 34 µg/mL chloramphenicol with a starting concentration of 0.05 OD_{600}. The cultures were kept under 37 °C at 200 rpm until the OD_{600} reading reached 0.6-0.8. The cell cultures were then inoculated with 0.4 mM IPTG and grown for 16-18 hours at 17 °C. Cultures were centrifuged at 5000 ×g for 45 minutes at 4 °C and the pellets were resuspended in 25 mL of 1 ×PBS before being centrifuged again at 10,000 ×g for 10 minutes at 4 °C. Pellets were weighed and then resuspended in ~40 mL of lysis buffer before being placed into -20 °C storage until ready to be purified. Prior to purification of WT-FEN1, R239A, K244A & R245A, the buffers in

Table 4 were produced.

Table 4: Buffers for WT-FEN1, R239A, K244A and R245A purification.

Buffer name	Composition
IMAC A1	20 mM Tris-Base, 5 mM imidazole, 1 M NaCl, 5 mM 2-BME, adjusted to pH 7.0 with 10 M NaOH
IMAC A2	20 mM Tris-Base, 40 mM imidazole, 1 M NaCl, 5 mM 2-BME, adjusted to pH 7.0 with 10 M NaOH
IMAC B1	20 mM Tris-Base, 250 mM imidazole, 1 M NaCl, 5 mM 2-BME, adjusted to pH .0 with 10 M NaOH
DEAE A1	20 mM Tris-Base, 1 mM EDTA, 20 mM DTT, adjusted to pH 8.0 with 10 M NaOH
DEAE B1	20 mM Tris-Base, 1 mM EDTA, 1 M NaCl, 20 mM DTT, adjusted to pH 8.0 with 10 M NaOH
Heparin A1	25 mM Tris-Base, 1 mM EDTA, 20 mM DTT, adjusted to pH 7.5 with 10 M NaOH
Heparin B1	25 mM Tris-Base, 1 mM EDTA, 1 M NaCl, 20 mM DTT, adjusted to pH 7.5 with 10 M NaOH
2 × Storage buffer	100 mM HEPES-KOH, 200 mM KCl, 1 mM EDTA, 10 mM DTT, 0.04 % NaN_3, 20 % glycerol, pH adjusted to pH 7.5
Stripping buffer	20 mM $NaSO_4$, 50 mM EDTA, 0.5 M NaCl, 0.02 % NaN_3, pH adjusted to 7.4 with NaOH

Frozen cell pellets were thawed on ice and sonicated using a SONICS VibraCell probe sonicator with 50 % amplitude, 3 seconds on and 27 seconds off for a total run time of 6 minutes. If lysates were high in viscosity, DNase 1 was added to the lysate. The sonicated lysate was then centrifuged at 20,000 ×g for 1 hour at 4 °C. Protein samples were purified at 4 °C using AKTA FPLC systems (Cytiva) and analysed by 4-20 % SDS-PAGE after each purification step. The lysate was loaded onto a Ni^{2+} charged IMAC FF column equilibrated with 5 CV of IMAC A1 buffer. The column was then washed with 5 CV of IMAC A1 buffer and then IMAC A2 buffer. Isocratic elution was used to elute protein off the column with IMAC B1 buffer. Isolated fractions were diluted 2-fold with DEAE A1 buffer and loaded onto the DEAE HiTrapQ column which was previously equilibrated with 5 CV of DEAE A1 buffer. A gradient elution of 0-1 M NaCl was used to elute any bound protein from the column. FEN1 proteins eluted directly into the flow-through and diluted 1:1 with Heparin A1 buffer. The sample was then loaded onto the heparin column previously equilibrated with 5 CV of heparin A1 buffer. Gradient elution was used (0-1 M) NaCl to elute FEN1 from the column. Isolated fractions were then concentrated to approximately 10 mL using Vivaspin-20 columns (10,000 MWCO) before being loaded onto

the HiPrep 2610 desalting column previously equilibrated with 2 × storage buffer. The sample was concentrated again using Vivaspin-20 columns (10,000 MWCO) until the concentration of the sample reached 200 µM using the extinction coefficient (22920 M^{-1} cm^{-1}). Purified samples were stored in 50 % glycerol and snap-frozen in liquid nitrogen before being stored at – 80 °C.

2.2 Kinetic assays of WT-FEN1, R239A, K244A and R245A

Steady-state and transient kinetic assays were performed using the buffers shown in Table 5.

Table 5: Buffers list for kinetic FEN1 kinetic assays

Buffer name	Composition
10 ×RB (Reaction buffer)	1M KCl, 0.5 M HEPES, 80 mM $MgCl_2$, 1 mg/mL BSA pH 7.5 with 5 M KOH
5 ×RRB (Reduced reaction buffer)	5000 µL 10 ×RB, 50 µL 1 M DTT, 4950 µL MilliQ water
1 ×RRB	2000 uL 5 ×RRB, 8000 µL MilliQ water
10 ×FB (Folding buffer)	1 M KCl, 0.5 M HEPES pH 7.5 with 5 M KOH
1 ×FB	0.1 M KCl, 50 mM HEPES pH 7.5
Quench solution	7 M UREA, 80 mM EDTA
HPLC buffer A	2.5 mM tetrabutylammonium bromide (TBAB), 1 mM EDTA, 0.1 % MeCN, diluted with MilliQ water
HPLC buffer B	2.5 mM TBAB, 1 mM EDTA, 70 % MeCN, diluted with MilliQ water
HPLC buffer C	8 % MeCN, diluted with MilliQ water
HPLC buffer D	80 % MeCN, diluted with MilliQ water

2.2.1 Kinetic assay substrates

Nuclease activity kinetic assays were performed discontinuously using a double-flap substrate DF(p5,1) 5'FAM (Figure 2.2). The substrate was formed by combining substrates in stoichiometric ratios of the strand ssA, and ssB. To monitor nuclease activity ssA was labelled with a 5'-fluorescein label attached to the 5'-phosphate. Strands ssA, ssB and the DF(p5,1)5'FAM substrate were previously ordered and combined by Grasby group members and stored at -20 °C. Oligos were designed using the Oligo analyser tool provided by

IDT-DNA Technologies (https://eu.idtdna.com/pages/tools/oligoanalyzer?returnurl=%2Fcalc%2Fanalyzer). The concentration of the single-stranded, and combined oligos were measured using absorbance at 260 nm with VersaWave 2 spectrophotometer.

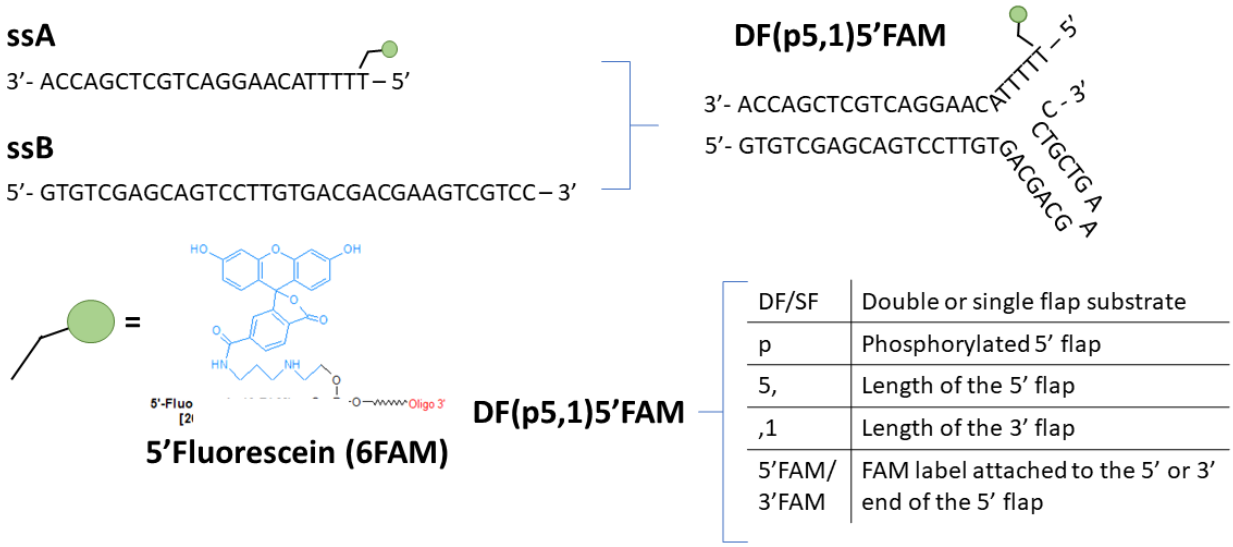


Figure 2.2 Schematic illustration of DF(p5,1)5'FAM for multiple, and single turnover experiments
Schematic of the DF(p5,1)5'FAM substrate used for nuclease activity assays with FEN1. Individual substrates ssA and ssB were combined at 50 µM and 55 µM respectively. The 5'-phosphate was labelled with a 6FAM fluorescent label Image taken from the GeneLink website.

2.2.2 Multiple turnover nuclease activity kinetic assay of FEN1

Preliminary Kinetic experiments of WT FEN1, R239A, K244A and R245A were performed with 100 nM DF(p5,1)5'FAM with final concentrations of 60 pM, 6 pM and 0.6 pM enzyme. Michaelis-Menten experiments were performed with final concentrations of substrate ranging from 0.5-1000 nM and final enzyme concentrations ranging from 1-300 pM. ssA and ssB oligos were mixed and diluted together in stoichiometric ratios to a working concentration of 50 µM and 55 µM using 1 × FB. Combined oligos were heated to 95 °C for 3 minutes and annealed at room temperature before being placed on ice. Proteins stocks were stored at 100 µM and subsequently diluted to 10 × final enzyme concentration using 1× RRB. To the reaction tube, 1600 µL of 1.125 × RRB was added accompanied 120 µL of 10 × final substrate (Table 6). 108 µL was then removed from the reaction mixture which acted as a negative control for RP-dHPLC. 108 µL of 10 × E was added to the reaction mixture and timepoints were taken at 2, 4, 6, 8, 10, 12 and 20 minutes.

Table 6 DNA substrate dilutions

Final [S] (nM)	Working stock of S (nM)	10 × [S]	Vol of S for 120 uL of 10 × S	Vol of 1 ×FB for 120 uL of 10 × S
1000	50,000	10,000	24	96
750	50,000	7500	18	102
500	50,000	5000	12	108
250	5,000	2500	60	60
100	5,000	1000	24	96
75	5,000	750	18	102
50	500	500	120	0
25	500	250	60	60
20	500	200	48	72
10	500	100	24	96
5	500	50	12	108

The assay was run at 37 °C. Control reactions were quenched in 20 µL of 80 mM EDTA/7 M UREA and reaction samples were quenched in 100 µL of 80 mM EDTA/ 8M UREA. Individual timepoints were purified using reverse-phase ion pair chromatography cartridge (OLIGOSepTM) (Hitachi Transgenomic). Chromatograms were measured using fluorescence via 5'FAM (Excitation: 494 nm, Emission: 525 nm). The column was equilibrated using 95 % HPLC buffer A and 5 % HPLC buffer B. Peak integration was performed using the HSM software. The % product was calculated using equation 1 and the concentration of product (nM) was calculated using equation 2. Initial rates were calculated using equation 3 with the linear regression model. Previously recorded values of WT FEN1 k_{cat} values of 150 min^{-1} and 20 nM for the K_M. Initial rates were normalised for enzyme concentration were used to plot a Michaelis Menten curve with non-linear regression. Data sets were weighted using $1/Y^2$ to accommodate for standard error.

$$\%\boldsymbol{P} = \frac{\boldsymbol{Area_{product}}}{\boldsymbol{Area_{product} + Area_{Substrate}}} \times \mathbf{100} \quad \textbf{Equation 1}$$

$$\boldsymbol{nM\ Product} = \frac{\%\ \boldsymbol{Product}}{\mathbf{100}} \times [\boldsymbol{S}] \quad \textbf{Equation 2}$$

$$\boldsymbol{V_0} = \frac{\boldsymbol{kcat}\,[\boldsymbol{E}][\boldsymbol{S}]}{\boldsymbol{K_m} + [\boldsymbol{S}]} \quad \textbf{Equation 3}$$

2.2.3 Single turnover kinetic experiment of WT/mutant FEN1 and DF(p5,1)5'FAM

Discontinuous single turnover assays were performed at 37 °C using the RQF-63 quench flow device (Hi-Tech Sci Ltd). 5 nM DF(p5,1)5'FAM substrates (final concentration) were heated and annealed as per described in section 2.2.2. Reactions were performed under single turnover conditions where [E] >>K_M >>[S]. Enzyme concentrations were 8 × the K_M value presented via steady state multiple turnover assays. Enzyme and substrate concentrations were initially prepared at 2 × final concentration. Substrates were diluted to 2 × using 1 × FB and then 1× RRB following heating and annealing. 2 × enzyme stocks were diluted using 1 × RRB. All enzyme and substrate concentrations were stored on ice until ready to be used for quench flow experiments. The RQF-63 quench flow device was filled with 1 × RRB containing 15 % glycerol in which the glycerol acts to inhibit reactant diffusion. Time points at 9.1, 12.1, 19.4, 30.6, 57.5, 124.2, 240.8, 840.8, 3240.8, 12840.8, 25640.8 and 51240.8 milliseconds were quenched in 7 M UREA and 80 mM EDTA. 100 µL of enzyme and substrate were injected followed by 80 µL of quenching solution. Quenched samples of approximately 160 µL were analysed by RP-dHPLC as previously described in section 2.2.2. The maximal single turnover rate k_{STmax} was determined by plotting the percentage of product formation ($\%\boldsymbol{P} = \frac{\boldsymbol{Area_{product}}}{\boldsymbol{Area_{product} + Area_{Substrate}}} \times \mathbf{100}$ Equation 1) versus time in minutes. Quench flow single turnover data sets were plotted and compared against one phase association (Equation 4) or two-phase association (Equation 5). Due to time-constraints, one replicate value was performed for each mutant. Quench flow data was processed using GraphPad prism.

$$\boldsymbol{P_t = P_\infty(1 - e^{-k_{STmax}{}^t})} \qquad \textbf{Equation 4}$$

$$\boldsymbol{P_t = P_{max1}\left(1 - e^{-k_{ST1}{}^t}\right) + P_{max2}\left(1 - e^{-k_{ST2}{}^t}\right)} \quad \textbf{Equation 5}$$

2.3 Measuring conformational change in FEN1 by 2-aminopurine substitution at the +1-1 position of the hydrolysed duplex.

Buffers and reagents can be found in Table 7.

Table 7 Buffer composition for ECCD experiments

Buffer name	Composition
10 × Ca^{2+}RB	100 mM $CaCl_2$, 1 M KCl, 0.5 M HEPES, pH adjusted to 7.5 with 5 M KOH
10 × FB	0.5 M HEPES, 1 M KCl, pH adjusted to 7.5 with 5 M KOH
1 M DTT	1 M, diluted with MilliQ water
1 × Ca^{2+} RRB	10 mM $CaCl_2$, 100 mM KCl, 50 mM HEPES, 1 mM DTT, MilliQ water

2.3.1 ECCD assay substrates

Changes in DNA conformation at the +1-1 and -1-2 positions of the 5'-flap-containing strand were performed by replacing bases at the +1-1/-1-2 position with tandem 2-AP bases that mimic the structure of adenine. The static structure of the double flap substrate DF(p5,1)$_{+1-1}$PP and DF(5,1)$_{-1-2}$PP aided the removal of flap movement within solution and allowed for isolated measurement of DNA conformational change by the increase, or decrease in 2-AP fluorescence intensity under non-catalytic conditions. Single strands 2AP-ssA and TEMP1 were combined in stoichiometric ratios to produce the DF(p5,1)$_{+1-1}$PP substrate whereas single strands 2AP-ssB and TEMP2 were combine in stoichiometric ratios for the substrate DF(p5,1)$_{-1-2}$PP. The substrates used for ECCD experiments were designed using the Oligo analyzer too by IDT Technologies (https://eu.idtdna.com/pages/tools/oligoanalyzer?returnurl=%2Fcalc%2Fanalyzer) (Figure 2.3).

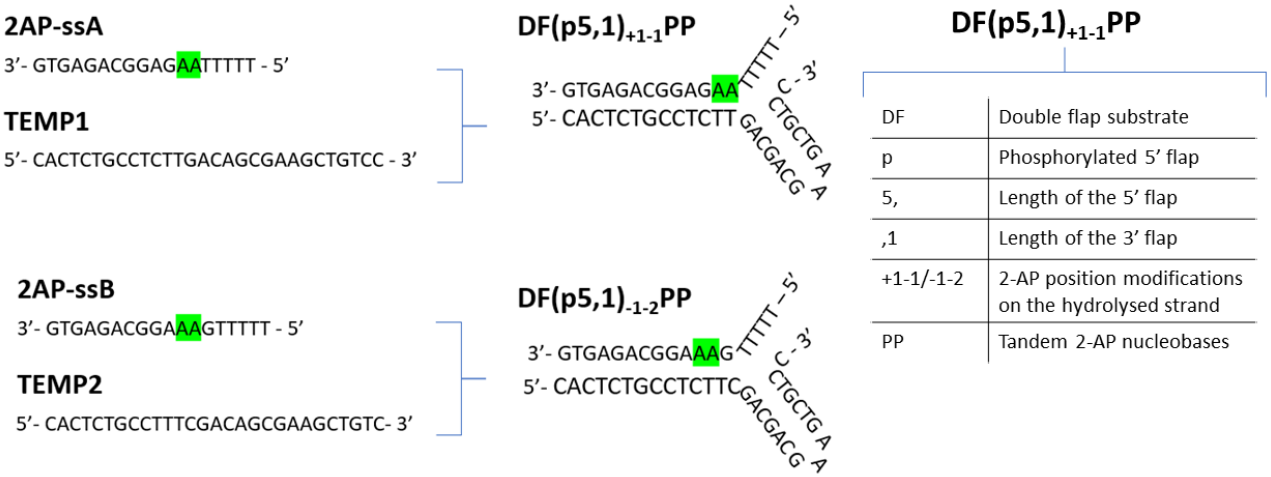


Figure 2.3 ECCD substrates DF(5,1)$_{+1-1}$PP and DF(5,1)$_{-1-2}$PP
Schematic of the DF(p5,1)$_{+1-1}$PP and DF(p5,1)$_{+1-1}$PP substrates used for ECCD assays of FEN1. Individual substrates 2AP-ssA/TEMP1 and 2AP-ssB/TEMP2 were combined at 100 µM and 110 µM respectively. Oligonucleotide concentrations were measured using absorbance at 260 nm by previous Grasby group members.

2.3.2 ECCD assay of WT and mutant FEN1

ECCD measurements were performed using the Chirascan CD spectrophotometer. The spectrophotometer was flushed with nitrogen and a purge check was performed. ECCD experiments were performed at 20 °C at a wavelength of 480 nm and bandwidth of 2 nm. 12 µM protein in 1 × Ca^{2+} RRB was mixed with 10 µM DF(p5,1)$_{+1-1}$PP or DF(p5,1)$_{-1-2}$PP (diluted with 1× FB) and spectra were obtained between 300-480 nm. Following the initial measurements, spectra were also taken upon the addition of EDTA to a final concentration of 25 mM. Baseline subtraction was performed on all samples before being converted to molar ellipticity. The traces were then smoothed. The traces were divided by 6600 due to 2 2-AP molecules present within the substrate before being normalised and analysed on GraphPad Prism 9.0.

2.4 2-AP fluorescence assay of the FEN1 mutant D34N and DF(14,1)$_{+1\text{-}1}$PP

See the Table 8 for the buffers used in the D34N fluorescence assay.

Table 8 Buffers and reagents used the D34N fluorescence assay

1 × RRB	*50 mM HEPES pH 7.5, 100 mM KCl, 0.1 mg/mL BSA and 2.5 mM DTT*
1 × FB	50 mM HEPES pH 7.5, 100 mM KCl

2.4.1 2-AP fluorescence assay

Fluorescence was used to investigate the conformational change and nuclease activity of D34N with the substrate DF(14,1)$_{+1\text{-}1}$PP by placing adjacent 2-AP-probed nucleotides at the +1 and -1 positions of the scissile phosphate on the processed strand. Oligonucleotide strands were designed using the Oligo analzyer tool from the IDT technologies website. The 2-AP labeled DNA was formed by a 3', 5' and template strands at 55, 50 and 55 µM respectively in stoichiometric ratios, resulting in a final 2-AP DNA concentration of 50 µM in 1× FB (DF(14,1)$_{+1\text{-}1}$PP). The substrate DF(14,1)$_{+1\text{-}1}$PP was heated to 95 °C for 5 minutes and allowed to cool until the substrate reached room temperature. D34N or WT FEN1 was mixed with 2-AP DNA with a final concentration of 2.5 µM and 0.5 µM respectively using RRB (Table 8) containing either excess $MgCl_2$ or $CaCl_2$ to a final sample volume of 500 µL [51]. Divalent metal ions were added last to the quartz cuvettes to trigger the reaction. Fluorescence measurements were monitored using the Fluoromax-3® fluorometer using the FlourEssence™ software at 37 °C with a slit length of 5 nm. Signal values were divided by reference values during the experimental run times and samples were quenched with 50 mM EDTA before storing at 4 °C. The excitation and emission wavelengths for the 2-AP containing substrate DF(14,1)$_{+1\text{-}1}$PP were 315 and 375 nm respectively. D34N and WT fluorimetry experiments also employed the method of adding one reagent to the quartz cuvette then taking a measurement. This was also performed to analyse whether fluorescence emission was altered upon the addition of each component of the reaction. A non-linear regression model with one phase association was used to fit the curve data in Graphpad Prism 9.0.

2.5 FAN1-NusA construct cloning and protein expression

Buffer and reagent conditions for the expression and analysis of the FAN-NusA construct can be found in Table 1.

2.5.1 PCR amplification

Molecular cloning procedures for the FAN1-NusA construct were kindly performed by previous Grasby group members. The pET43a FAN1-NusA construct was kindly provided by Professor John Rouse and his research group at the University of Dundee.

2.5.2 Cell transformation and plasmid preparation for sequencing

Transformation, plasmid preparation and sequencing of the FAN1-NusA construct was kindly performed by previous Grasby group members and are described in sections 2.1.2 and 2.1.3.

2.6 Purification of FAN1-NusA

See Table 9 for buffers used to purify the FAN-NusA construct.

Table 9 Buffers used in sections 2.6.1 and 2.6.2 to purify FAN-NusA

Buffer name	Composition
IMAC A1	25 mM Tris, 1 M NaCl, 20 mM imidazole, 10 % glycerol, 0.02 % NaN_3, 5 mM 2-BME, pH adjusted to 7.0 with 5 M NaOH
IMAC A2	25 mM Tris, 500 mM NaCl, 20 mM imidazole, 10 % glycerol, 0.02 % NaN_3, 5 mM 2-BME, pH adjusted to 7.0 with 5 M NaOH
IMAC B1	25 mM Tris, 500 mM NaCl, 300 mM imidazole, 10 % glycerol, 0.02 % NaN_3, 5 mM 2-BME, pH adjusted to 7.0 with 5 M NaOH
Heparin A1	25 mM Tris, 100 mM NaCl, 1 mM $CaCl_2$, 0.02 % NaN_3, 20 mM 2-BME, pH adjusted to 7.5 with 5 M NaOH
Heparin B1	25 mM Tris, 1 mM $CaCl_2$, 0.02 % NaN_3, 1M NaCl, 20 mM 2-BME, pH adjusted to 7.5 with 5 M NaOH

2.6.1 Small-scale purification of FAN1-NusA

For proof of concept that the FAN1-NusA construct could be successfully purified by the poly histidine tag, His spin trap columns containing Ni^{2+} Sepharose beads were employed to purify FAN1-NusA. Lysed and sonicated supernatant was applied to the column and centrifuged at 13,400 rpm for 5 minutes. This was repeated for wash and elution stages before being analysed by SDS-PAGE.

2.6.2 Large-scale expression and purification of FAN1-NusA

A 10 mL starter culture containing 2XYT supplemented with 100 µg/mL Ampicillin and 34 µg/mL chloramphenicol was inoculated on an LB agar plate containing the same antibiotics. The culture was placed under 200 rpm shaking incubation at 30 °C, for 16 hours. The culture was then seeded into 500 mL of 2XYT media containing 100 µg/mL Ampicillin and 34 µg/mL chloramphenicol to a concentration of 0.05 OD_{600} and placed back under 200 rpm shaking incubation at 30 °C until the OD_{600} reached 0.6-0.8. The cultures were then inoculated with 0.4 mM IPTG and grown for 16 hours with 200 rpm shaking incubation at 18 °C. The culture was centrifuged at 5000 ×g for 1 hour at 4 °C and the 2XYT media was decanted. The pellet was resuspended in lysis buffer before being centrifuged again at 10,000 ×g for 20 minutes. The supernatant was then loaded onto a 5 mL cobalt TALON IMAC column previously equilibrated with IMAC A1 buffer. The column was then washed with IMAC A2 buffer. The bound protein was eluted from the column by isocratic elution with IMAC B1 buffer. Membrane dialysis with a 10,000 MWCO was used to dialyse the fractions into Heparin A1 buffer accompanied by the addition of (2 units/mL) Thrombin protease. The sample was dialysed for 16 hours prior to loading onto the Heparin column. The column was equilibrated with Heparin A1 buffer and bound proteins were eluted from the column using gradient elution of 0.1-1 M NaCl. Between every step, 4-20 % SDS-PAGE was performed to identify FAN1-NusA at ~130 kDa.

2.6 FAN1-SUMO/ FAN1-MBP construct cloning and FAN1-MBP protein expression

2.7.1 PCR amplification

PCR amplification was performed by designing specialised primers that bind the template DNA and enable the synthesis of the FAN1 gene by Q5 and PFU hotstart polymerase reaction techniques before ligation into pESUMOpro, and pET-His6-MBP-TEV vector. Due to the length of the SUMO and MBP FAN1 constructs,

internal primers were also produced to achieve optimal sequencing quality. Primers were designed using the QuickChange primer design service provided by Agilent (Table 10).

Table 10 Primers for FAN1-SUMO and FAN1-MBP construct production and amplification

Primer name	Sequence
SUMO_hFAN1_F	5' - GCGGTCTCGAGGTCCTTACTACCTTCGGAGTTTCC - 3'
SUMO_hFAN1_IF	5' – CTGGACAGTCAGTACGAATCT G - 3'
SUMO_hFAN1_IR	5' – AAGCTGCTGGAAGAGGTGCTTGAA C - 3'
SUMO_hFAN1_R	5' - GCGTCTCTAGTTAGCTAAGGCTTTGCCTCTTAG - 3'
MBP_hFAN1_F	5' - TACTTCCAATCCAATGCAAATGGTCCTGGTCAAACAACC- 3'
MBP_hFAN1_IF	5' – AACCATCCAAACCGTTCGTTG – 3'
MBP_hFAN1_IR	5' – GATGCACTTGATAGTCGGTTCAG – 3'
MBP_hFAN1_R	5' – TTATCCACTTCCAATGTTATTAGCTAAGGCTTTGGCTCTTAGC – 3'

The FAN1-SUMO and FAN1-MBP constructs were produced using the PFU hotstart polymerase, and Q5 polymerase reaction protocols using a BIORAD PCR machine with both reactions containing a total volume of 25 (Table 11).

Table 11 PCR reactions for FAN1 SUMO and FAN1-MBP constructs

Reagent	Volume (µL)	Reagent	Volume (µL)
10 × pFU hotstart buffer	2.5	**5 × Q5 Polymerase buffer**	5.0
10 mM dNTP's	2.0	**10 mM dNTP's**	0.5
3 µM Primer pair	3.0	**3 µM Primer pair**	2.5
1 ng/ µL Linearised pET43a FAN1 template plasmid	2.0	**1 ng/µL pcDNAhFAN1 template**	0.5
DMSO	2.5	**Q5 polymerase**	0.25
MilliQ water	12.5	**MilliQ water**	11.25
Total	25	**Total**	25

Varying PCR temperatures were used for DNA denaturation, polymerase attachment and subsequent DNA elongation from the annealed primer. PCR Temperatures were increased or decreased depending on the melting temperature of the annealed primer and whether polymerases required higher temperatures to facilitate strand elongation (Figure 2.4).

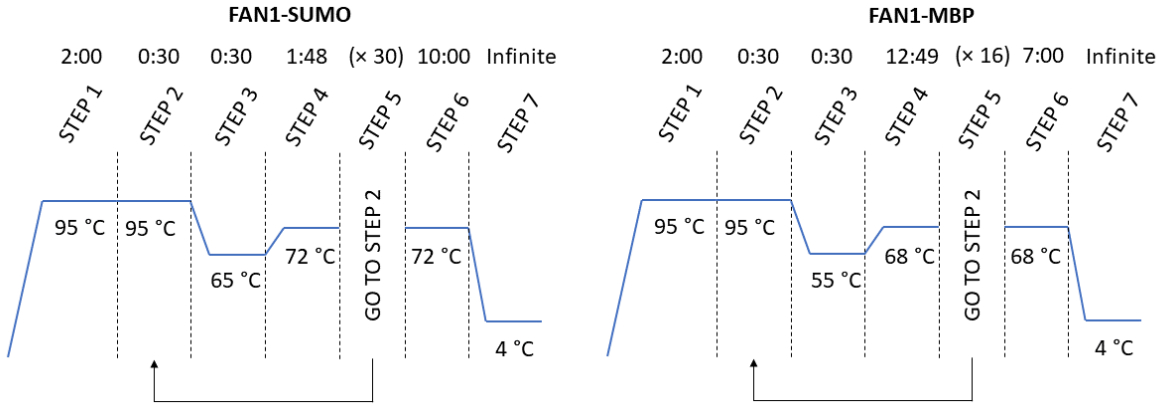


Figure 2.4 PCR temperature reaction stages for the FAN1-SUMO and FAN1-MBP constructs
Schematic of the PCR reaction cycles and relevant temperatures. FAN1-SUMO was performed using the Q5 polymerase reaction protocol (Left) and the FAN1-MBP construct was produced using the PFU hotstart polymerase protocol (Right).

To produce the FAN-SUMO construct, the pESUMO-pro empty vector was digested with Bsa1 restriction endonuclease at 37 °C for 1 hour. The linearised pESUMO vector was then analysed by agarose gel electrophoresis. The PCR amplified FAN1 gene insert was also analysed by agarose gel electrophoresis before gel extraction was performed (Monarch). The PCR product was then cleaved with Bsa1 as previously described. PCR clean up (Monarch) of both the FAN1 gene insert and pESUMO-pro empty vector was performed before both components being ligated together with T4 DNA ligase. The ligated FAN1-SUMO construct was then transformed into DH5α cell lines before plasmid amplification and sequencing.

The FAN1-MBP construct was produced using the ligation independent cloning strategy. The pET-HMT empty vector was linearised by Ssp1 digestion at 37 °C for 1 hour and analysed by agarose gel electrophoresis. The FAN1 gene insert produced using Q5 polymerase PCR and was also analysed by agarose gel electrophoresis and then gel-extracted. Both the pET-HMT empty vector and FAN1 gene insert were treated with T4 DNA polymerase to produce 3'-overhangs before prior to PCR clean-up of both samples. The pET-HMT empty vector and FAN1 gene insert were combined and transformed into DH5α cell lines and grown on LB agar

plates with 100 ug/mL ampicillin. Colony PCR was performed on selected colonies before analysis by agarose gel electrophoresis, plasmid amplification and sequencing.

2.7.2 Cell transformation and plasmid preparation for sequencing

Refer to sections 2.1.2 and 2.1.3 for bacterial transformation and plasmid preparation of ligated plasmids prior to sequencing by GeneWiz.

2.8 Protein expression and Purification of FAN1-MBP

2.8.1 Small-scale purification of FAN1-SUMO and FAN1-MBP

See section 2.5.1 for proof-of-concept experiments for purifying FAN1-MBP by Ni^{2+} IMAC. Also refer to Table 9 for buffer components.

2.8.2 Large-scale Purification of FAN-MBP

Table 12 Buffers for the purification of FAN1-MBP in section 2.8.2

Buffer name	Composition
Lysis Buffer	*25 mM Tris pH 7.5, 1.0 M NaCl, 5 mM imidazole, 5 % ethylene glycol, 0.02 % NaN_3, 0.1 % lysozyme, 1 protease inhibitor tablet, 5 mM 2-BME*
IMAC A1	*25 mM Tris pH 7.5, 1.0 M NaCl, 5 mM imidazole, 5 % ethylene glycol, 0.02 % NaN_3, 5 mM 2-BME*
IMAC A2	*25 mM Tris pH 7.5, 1.0 M NaCl, 40 mM imidazole, 5 % ethylene glycol, 0.02 % NaN_3, 5 mM 2-BME*
IMAC B1	*25 mM Tris pH 7.5, 1.0 M NaCl, 300 mM imidazole, 5 % ethylene glycol, 0.02 % NaN_3, 5 mM 2-BME*
Heparin A1	*25 mM MES pH 6.0, 5 % ethylene glycol, 0.1 mM EDTA, 20 mM 2-BME*
Heparin B1	*25 mM MES pH 6.0, 1.2 M NaCl, 5 % ethylene glycol, 0.1 mM EDTA, 20 mM 2-BME*
FAN1 TEV buffer	*25 mM Tris pH 7.5, 200 mM NaCl, 5 % ethylene glycol, 0.1 mM EDTA, 5 mM 2-BME*
2 × FAN1 Storage buffer	100 mM HEPES-KOH, 200 mM KCl, 1 mM EDTA, 10 mM DTT, 0.04 % NaN_3, 20 % glycerol, pH adjusted to pH 7.5

A 50 mL culture of 2XYT media containing BL21 DE3 PlysS cells were grown for 16 hours at 37 °C. The cultures were seeded into 5 litres of 2XYT media and incubated under 37 °C shaking incubation until an OD_{600} of 0.6-0.8 was reached, the cultures were inoculated with 0.4 mM IPTG and grown for 16 hours at 17 °C. The culture was centrifuged at 5000 ×g for 1 hour at 4 °C to separate the cell mass from the 2XYT media. The pellet was resuspended in 1 ×PBS and centrifuged again at 10,000 ×g for 20 minutes to remove any further debris or dead cell matter. The washed pellet was resuspended in lysis Buffer before sonication at 50 % amplitude, 10 seconds on, 20 seconds off, and a total time of 6-8 minutes. The culture was centrifuged at 10,000 ×g for 30 minutes before being loaded on a 20 mL IMAC FF column following previously described IMAC procedures with buffers from Table 12. Fractions containing FAN1-MBP were diluted 4-fold using Heparin A1 buffer. The sample was then loaded onto a 20 mL Heparin column and purified as described in previous procedures. Fractions containing FAN1-MBP were desalted into FAN1 TEV buffer using the 2610 desalting column. 1 mL of 40 µM TEV protease was added to the desalted fractions and allowed to cleave overnight at 4 °C. IMAC FF purification was repeated to remove the MBP fusion protein containing the poly-histidine tag. Flow-through fractions identified to be FAN1 through SDS-PAGE were desalted into FAN1 storage buffer and then concentrated to 4 µM (5 mL) before being diluted 2-fold with 50 % glycerol. The concentration of the solution prior to diluting with 50 % glycerol was measured by the absorbance at 280 nm. Purified FAN1 stocks were snap frozen in liquid nitrogen and stored at -20 °C.

2.8.3 Revised method for the purification of FAN-MBP after differential scanning fluorimetry

Table 13 Buffers used for the revised purification of FAN1-MBP

Buffer name	Composition
Lysis Buffer	500 mL of IMAC A1, 0.1 % Lysozyme, 8 µg/mL DNase1, pH adjusted to 7.0 with 5 M HCl
Amylose B1	1000 mL of IMAC1, 10 mM maltose, pH adjusted to 7.0 with 5 M HCl
IMAC A1	50 mM Sodium phosphate, 5 mM Imidazole, 0.02 % NaN_3, 5 mM BME, pH adjusted to 7.0 with 5 M HCl
IMAC B1	50 mM Sodium phosphate, 250 mM imidazole, 0.02 % NaN_3, 5 mM BME, pH adjusted to 7.0 with 5 M HCl
2 × FAN1 Storage buffer	100 mM PIPES, 200 mM NaCl, 2 mM $CaCl_2$, 0.02 %, 10 mM DTT, pH adjusted to 7.0 with 5 M HCl

A 50 mL starter culture of 2XYT containing 100 µg/mL carbenicillin and 34 µg/mL chloramphenicol was inoculated with a single colony grown from an LB agar plate containing the same antibiotics. The culture was grown for 16 hours at 37 °C, at 200 rpm. The culture was then seeded into 2 L of 2XYT media containing 100 µg/mL carbenicillin and 34 µg/mL chloramphenicol to a final concentration of 0.05 OD_{600} using a bench-top spectrophotometer. When the culture reached 0.6-0.8 OD_{600}, the culture was inoculated with 0.4 mM IPTG and grown for 16 hours with 200 rpm shaking incubation at 17 °C. The culture was centrifuged at 5000 ×g for 1 hour at 4 °C to remove the 2XYT media and the pellet was resuspended in 1 ×PBS. 1 ×PBS was removed by centrifugation (10,000 ×g, 20 minutes) and the pellet was resuspended in lysis buffer. Prior to sonication (5 seconds on, 20 seconds off, 70 % amplitude) and centrifuged at 10,000 ×g for 20 minutes.

The supernatant was loaded onto the NEB amylose resin equilibrated with IMAC A1 buffer (1CV = 15 mL). Due to the size of the gravity flow column, repeated additions of lysate were added to the column and the flow-through fractions were collected for SDS-PAGE. The amylose resin was washed three times with IMAC A1 buffer to remove any protein contaminants in between the amylose resin beads. On column cleavage was performed by adding 500 µL of TEV protease and allowed to cleave the serine protease site for 16 hours. The

next morning another 1000 µL of 40 µM TEV protease was added to the amylose resin and allowed to cleave for one hour before the column being eluted with 3 CV of amylose B1 buffer. Fractions containing predominantly cleaved FAN1 were pooled together and loaded onto the IMAC FF column previously equilibrated with IMAC A1 buffer. The column was then washed with a combination of 92 % IMAC A1 and 8 % IMAC B1 (25 mM total imidazole concentration) before isocratic elution with IMAC B1 buffer. Fractions were analysed by SDS-PAGE and selected fractions were pooled for desalting into 2 ×FAN1 storage buffer using the 2610 desalting column. Pooled fractions were concentrated to approximately 18 µM after 2-fold dilution with 100 % glycerol. Samples were snap-frozen with liquid nitrogen and stored at -80 °C.

2.9 Differential scanning fluorimetry screening

DSF screening was performed to identify optimum buffer conditions for FAN1. DSF screening experiments were performed using a 96-well RT-PCR machine (Agilent). SYPRO-Orange fluorescent dye was used to monitor the change in fluorescence following change in the folded state of FAN1-MBP. Melt curves were recorded using the MxPro software with excitation and emission parameters set to 492, and 610 nm respectively. Sample plates were produced and loaded into the RT-PCR machine with a starting temperature of 25 °C. The temperature was incrementally increased by 1 °C every minute until the temperature reached 95 °C. The raw data was smoothed and normalised before processing melt curves using the simpleDSFviewer software [119]. The half maximal temperature (T_M) was obtained through simpleDSFviewer. The normalised, smoothed curves and T_M values were plotted using GraphPad prism 9.0.

2.9.1 Screening for optimum enzyme and SYPRO-Orange dye concentration

Preliminary experiments for DSF are required to identify the optimal concentration of enzyme and SYPRO-Orange dye concentration to prevent fluorescent signal saturation during DSF experiments. The volume of each sample well was 25 µL. The final concentration of SYPRO-Orange dye in the well was 1 ×, 2 ×, 4 ×, and 6 ×. For each dye concentration, a 25 × working stock was produced for each (E.g., 1 x final dye concentration: 25 × working dye stock). The final enzyme concentrations ranged from 0-1 µM diluted in 1 × FAN1 storage buffer without glycerol (section 2.8.2). The SYPRO-Orange dye was added to the sample wells last before

centrifugation at 500 ×g for 1 minute at 4 °C before loading onto the RT-PCR machine and performing the DSF experiment.

2.9.2 Commercially bought screens and in-house produced screens

From section 2.8.1, 0.1-0.2 µM FAN1 and 1 × SYPRO-Orange dye was identified as the optimal final concentrations for future DSF screening experiments. For screening buffer components using DSF, a total reaction volume of 25 µL was prepared. Commercially bought screening plates including the RUBIC Additive screen (master stock: 1.56 ×), Durham pH buffer screen (master stock: 2 ×), and Durham osmolyte screen (master stock: 2 ×) were diluted to 1 × upon the addition of enzyme to a final concentration of 0.2 µM and SYPRO-Orange dye to a final concentration of 1 ×. Screens controlled for ionic strength; broad range pH buffer screens were all produced in-house to a master plate concentration of 2 ×.

2.10 FAN1 oligo production and kinetic assays

2.10.1 DMT on production of FAN1 oligos

Table 14 Buffers used for phosphoramidite synthesis of FAN1 oligonucleotide substrates.

Buffer	Composition
Buffer A	100 mM triethylammonium acetate (TEAA) pH 7.0, 0.1 % acetonitrile
Buffer B	100 mM TEAA pH 7.0, 25 % acetonitrile (MeCN)
Buffer C	8 % acetonitrile (MeCN)
Buffer D	80 % (MeCN)
BTP Buffer A	345 mM bis-tris propane (BTP), pH 11.5
BTP Buffer B	10 mM BTP pH 9.0, 1 mM EDTA

Oligonucleotide synthesis was performed using an ABI 394 DNA synthesiser by Dr David Finger. To process the single-stranded products, 5 mL of 10 % diethanolamine (DEA) was pushed through the columns slowly over the course of 5 minutes to remove the cyanoethyl protecting groups from the phosphates. The columns were then slowly washed with 5 mL of MeCN. The column was then flushed with air for 2-5 minutes before the column being flushed slowly with argon for 5 minutes. Samples containing a 5’ dimethoxytrityl (DMT) protecting and oligonucleotides containing 5’FAM labels were removed from the column pushing through 4 mL of concentrated ammonium hydroxide, which removed the base protecting groups and cleaved the 3’-hydroxyl ester from the column beads. Oligonucleotide removal from the column was performed very slowly over 60 minutes and the flow-through was collected into Eppendorf tubes. Oligonucleotide products were left at room temperature for 16 hours. 100 µL of BTP buffer A ensure the DMT protecting group was not prematurely removed from the oligonucleotides. Samples were then vacuum centrifuged at 22 °C for 16 hours to remove the concentrated ammonium hydroxide from the sample. The dried solid was then resuspended in 500 µL 10 mM BTP buffer B. The samples were then purified using RP-dHPLC using with an MeCN gradient of 5 – 55 % buffer B (with 100 mM TBAB) between 1-9 minutes. Samples containing a 5’-DMT

protecting groups were freeze dried before being resuspended in 500 µL of sodium acetate buffer pH 5.2 and incubating at 65 C to removal the DMT protecting group from the 5'-end. The samples were then diluted 1:3 with ethanol to precipitate the DNA out of solution. The precipitates were washed with 70% Ethanol to remove any excess salts and centrifuged at 10,000 ×g for 15 minutes to recover the samples. Oligonucleotides that had the 5'-DMT removed by acid were treated with ammonium hydroxide again to remove the residual 5'-protecting group from the solid chemical phosphorylation reagent II. All oligonucleotide products were then desalted using NAP-25 size exclusion columns by gravity flow into MilliQ water. Samples were finally characterised by Mass spectrometry; this was performed by Professor Mark Dickman.

2.10.2 FAN1 nuclease assays

Refer to Table 15 for buffers and gels used for kinetic assays and RP-HPLC method development.

Table 15 Buffer table for FAN1 nuclease assays

Buffer name	Composition
10 × RB	500 mM PIPES-KOH, 1M KCl, 80 mM $MgCl_2$, 1.0 mg/mL BSA, pH adjusted to 7.0 with 5 M KOH
5 × RRB	250 mM PIPES-KOH pH 7.0, 500 mM KCl, 40 mM $MgCl_2$, 0.5 mg/mL BSA, 25 mM DTT
1 × RRB	50 mM PIPES-KOH pH 7.0, 100 mM KCl, 40 mM $MgCl_2$, 0.1 mg/mL BSA, 5 mM DTT
10 × FB	500 mM PIPES-KOH, 1M KCl, pH adjusted to 7.0 with 5 M KOH
1 × FB	50 mM PIPES-KOH pH 7.0, 100 mM KCl
2 × Loading buffer	95 % v/v formamide, 0.02 % bromophenol Blue, 0.01 % xylene cyanol, 0.02 % SDS, 0.1 mM EDTA
5'FAM substrate gel recipe	20 % acrylamide (19:1), 1 ×TBE, 7 M UREA
3'FAM substrate gel recipe	15 % acrylamide (19:1), 1 ×TBE, 7 M UREA
TBAB buffer A	2.5 mM TBAB, 0.1 % MeCN, 1 mM EDTA
TBAB buffer B	2.5 mM TBAB, 70 % MeCN, 1 mM EDTA
TEAA Buffer A	100 mM TEAA, 0.1 % MeCN, 1 mM EDTA, pH adjusted to 7.0 with concentrated TEA
TEAA Buffer B	100 mM TEAA, 25 % MeCN, 1 mM EDTA, pH adjusted to 7.0 with concentrated TEA
TEAA buffer B V2	100 mM TEAA, 50 % MeCN, 1 mM EDTA, pH adjusted to 7.0 with concentrated TEA
PAGE running buffer	1 ×TBE/ 7 M UREA

FAN1 substrates were produced as previously described before being combined in stoichiometric ratios to produce double or single flap structures. Revert to section 2.2.1 to see how single stranded oligonucleotides were designed and analysed. Single-stranded oligonucleotide sequences and substrates can be in Figure 2.5. SB5,1,HT2 was produced by previous Grasby group members.

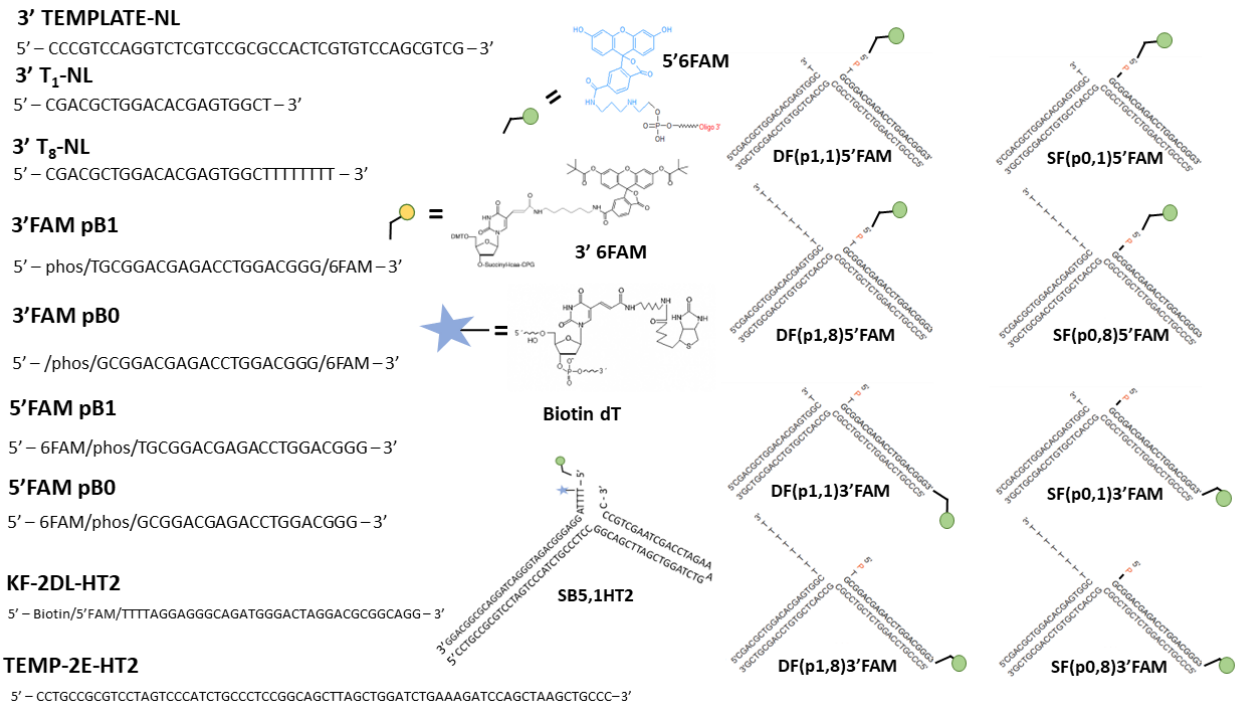


Figure 2.5 FAN1 substrates used in nuclease activity assays.

Schematic of substrates and their individual strands used for FAN1 nuclease assays. Substrates labelled with 'NL' indicate no attached 6FAM label on the strand. Lower case 'p' indicates the strand contains a 5'-phosphate group. Strands containing a capital 'T' followed by a subscript number indicates the number of thymine bases on the 3'-flap strand. Strands containing 'B1' or 'B0' correspond to a 1-nucleotide 5'-flap is present or no 5'-flap is present respectively. Single strands KF-2DL-HT2 and TEMP-2E-HT2 were combined to produce the SB5,1,HT2 substrate; DF(p1,1)5'FAM consists of 3'T_1-NL, 3'FAM pB1 and TEMPLATE-NL; SF(p0,1)5'FAM consists of 3'T_1-NL, 5'FAM pB0 and TEMPLATE-NL; DF(p1,8) 5'FAM consists of 3'T_8-NL, 5'FAM pB1 and TEMPLATE-NL; SF(p0,8)5'FAM consists of 3'T_8-NL, 5'FAM pB0 and TEMPLATE-NL; DF(p1,1)3'FAM consists of 3'T_8-NL, 3'FAM pB1 and TEMPLATE-NL; SF(p0,1)3'FAM consists of 3'T_1-NL, 3'FAM pB0 and TEMPLATE-NL; DF(p1,8)3'FAM consists of 3'T_8-NL, 3'FAM pB1 and TEMPLATE-NL; SF(p0,8)3'FAM consists of 3'T_8-NL and 3'FAM pB0.

FAN1 nuclease assays using the SB5,1,HT2 substrate were performed at 37 °C with timepoints taken from 0.5-256 minutes in a total reaction volume of 3000 µL. To a reaction tube was added 600 µL of 1.25 ×RRB, 6 µL of 5000 nM SB5,1HT2 made up with MilliQ water. 600 µL of the mixture was removed as the negative control sample and 600 µL of 2000 nM FAN1 was added to the mixture to initiate the reaction. Reactions were quenched by removing 100 µL of reaction and quenching in 150 µL of 500 mM EDTA. Reaction samples were analysed by RP-dHPLC as previously described in section 2.2.2. The final concentration of FAN1 was 400 nM and the final concentration of substrate was 10 nM. Reaction buffers used for the FAN1 nuclease assay with SB5,1,HT2 can be found in Table 5. Quenched reactions were analysed by RP-dHPLC

FAN1 nuclease assays of DF(p1,8)3'FAM and DF(p1,8)5'FAM were performed at 37 °C with final enzyme concentrations of FAN1 between 0.4-400 nM. Enzyme stocks were diluted using 1 × RRB to a working concentration of 10 × final enzyme concentration. FAN1 substrates were prepared in stoichiometric ratios to a final concentration of 50,000 nM using 1 × FB followed by heating at 95 °C before allowing to anneal at room temperature for 20 minutes. Substrates were diluted to 5000 nM with 1 × FB (5 × the final substrate concentration). Into each reaction was pipetted 240 µL of 5 × RRB, 2.4 µL of 500 nM DF(p1,8)3'FAM or DF(p1,8)5'FAM and 837.6 µL of MilliQ water to a total reaction volume of 1200 µL. 120 µL of 10 × FAN1 was added to initiate the reaction and timepoints were taken between 0.5-64 minutes. A separate reaction tube replacing 120 µL FAN1 for MilliQ acted as the negative control. Upon mixing, the final concentration of substrate was diluted to 10 nM. Reactions were performed in duplicate. 100 µL volumes of reaction were quenched in 3 µL of 500 mM EDTA.

Two individual reactions were produced for each of the oligos: DF (p1,1)5'FAM, DF(p1,1)3'FAM, SF(p0,1)5'FAM, SF(p0,1)3'FAM, SF(p0,8)5'FAM and SF(p0,8)3'FAM reactions where one reaction was quenched at 16 minutes and the second reaction was quenched at 32 minutes with 3 µL of 500 mM EDTA. Substrates were again produced to 50,000 nM using 1 × FB. Into each reaction tube was pipetted 160 µL of 1.125 × RRB. 20 µL of 100 nM substrate was mixed into the reaction tube. 18 µL of the mixture was removed from the reaction tube and used as a negative control. 18 µL of 4.0 nM FAN1 was then pipetted into the reaction tubes to initiate the reaction resulting in 0.4 nM final enzyme concentration and 10 nM final substrate concentration.

2.10.2.1 Denaturing UREA PAGE analysis of quenched FAN1 reactions.

Quenched reactions were analysed by denaturing UREA PAGE. 15 and 20 % acrylamide (19:1) gels with 1 ×TBE and 7 M UREA were diluted using PAGE running buffer. The mixture was then degassed for 10 minutes. 280 µL of 10 % APS and 14 µL of TEMED were added to the mixture before decanted into the cassette and allowed to set for 1 hour with the well comb placed into the gel mixture. Once set, the comb was removed, and each well was syringe washed with PAGE running buffer. The gel was run at 35 Watts until the heat plate reached a temperature of 55 °C. Quenched reactions were diluted 1:1 in 2 × loading buffer. Samples were then heated at 95 °C for 5 minutes and ~100 ng of sample (10 µL) was loaded into each well. The gel was run

until the bromophenol blue line reached approximately 60 % of the way down the gel. The gel was stained with SYBERGOLD stain and imaged using the BIO-RAD imaging camera.

2.10.2.2 RP-dHPLC analysis of FAN1-SB5,1HT2 reactions

FAN1 nuclease activity assays with the SB5,1,HT2 substrate were performed using identical RP-dHPLC conditions as section 2.2.2. Gradient experiments were performed using RP-HPLC analysis with 300-400 fmols of sample being injected onto the column. Varying acetonitrile gradients were tested for TBAB and TEAA-containing buffers before the column was washed with 100 % TBAB, or TEAA elution buffer.

Chapter 3 2-aminopurine fluorimetry assay measuring changes in D34N conformational changes and reaction rate

3.1 Introduction

Several research groups have performed kinetic, biophysical, and structural studies of FEN1 to unearth its mechanistic properties regarding substrate processing and wider role amongst DNA maintenance and repair. Multiple models have been proposed as to the series of events that unfold during 5'-flap processing. However, it is universally agreed that FEN1 uses divalent metal-ions to perform nucleophilic attack of water at the +1-1 scissile phosphate with a varying length 5'-flap. Furthermore, FEN1 nuclease activity is significantly increased upon the addition of a 1-nucleotide 3'-flap whereas reaction is not observed or extremely slow in the presence of single-stranded DNA or Holiday junctions compared to other FEN nucleases [45], [48], [120]. Precise, and reproducible incision of the +1-1 scissile phosphate (Figure 3.1A) and active site transfer by FEN1 is a critical event during DNA replication and repair. Non-specific hydrolysis of the 5'-flap strand will no longer produce ligatable products and so threaten the efficiency and fidelity of replication and repair (Figure 3.1C) [121].

To achieve a catalytically competent FEN1-substrate complex, substantial DNA and FEN1 conformational changes are required. Unpublished NMR studies have suggested that FEN1 substrates adopt multiple conformational states including co-axial stacking and bent structures (Figure 3.1A) until initial binding to FEN1 enforces a 100° bend at the 5' and 3'-duplexes and thus, permits the 5'-flap to be threaded through the disordered helical gateway. In addition to this, smFRET data has demonstrated gapped DNA substrates actively transisiton between coaxially stacked, and bent conformations prior to complex formation with pol-β [122]. However, with regards to FEN1, the 5'-flap is still not ready to be hydrolysed (Figure 3.1B) [40], [123]. Active site transfer of the scissile phosphate is reliant on 3'-flap binding to the 3'-flap loop wedge motif almost 20 Å away from the active site. Following binding of the 3'-flap, L53-moderated conformational ordering of the helical gateway occurs (Figure 3.1B) [48]. Moreover, DNA distortion of the +1-1 positions occurs in a 'rolling' or 'twisting' motion to correctly position the scissile phosphate near the Mg^{2+} ions. This was shown through ECCD experiments using 2-AP labelled double flaps at the +1-1/-1-2 positions (Figure 3.1D). After

entering the catalytically competent state following FEN1 ordering, the 5'-flap is then exposed to the active site residues, all of which are highly conserved within the FEN superfamily. Active site residues D181 and D233 co-ordinate metal ion A whereas D34 and D86 co-ordinate metal ion B towards the scissile phosphate. The metal ions are accompanied by a water molecule utilised in for nucleophilic attack. K93 and R100 form ion pairs with the scissile phosphate to facilitate hydrolysis and stabilise the 5'-flap. Lastly, Y40 stacks against the +1 nucleobase [68]. It is reported that mutations severely impede FEN1 function [67]. However, FRET data has revealed the rate of DNA bending is not affected by active site residue mutations to alanine in the presence, or absence of Ca^{2+} ions. Moreover, DNA bending at the junction is not affected by flap length and DNA conformational change is reduced when mismatched bases are incorporated into the 5'-duplex towards the junction [67] (Figure 3.3A).

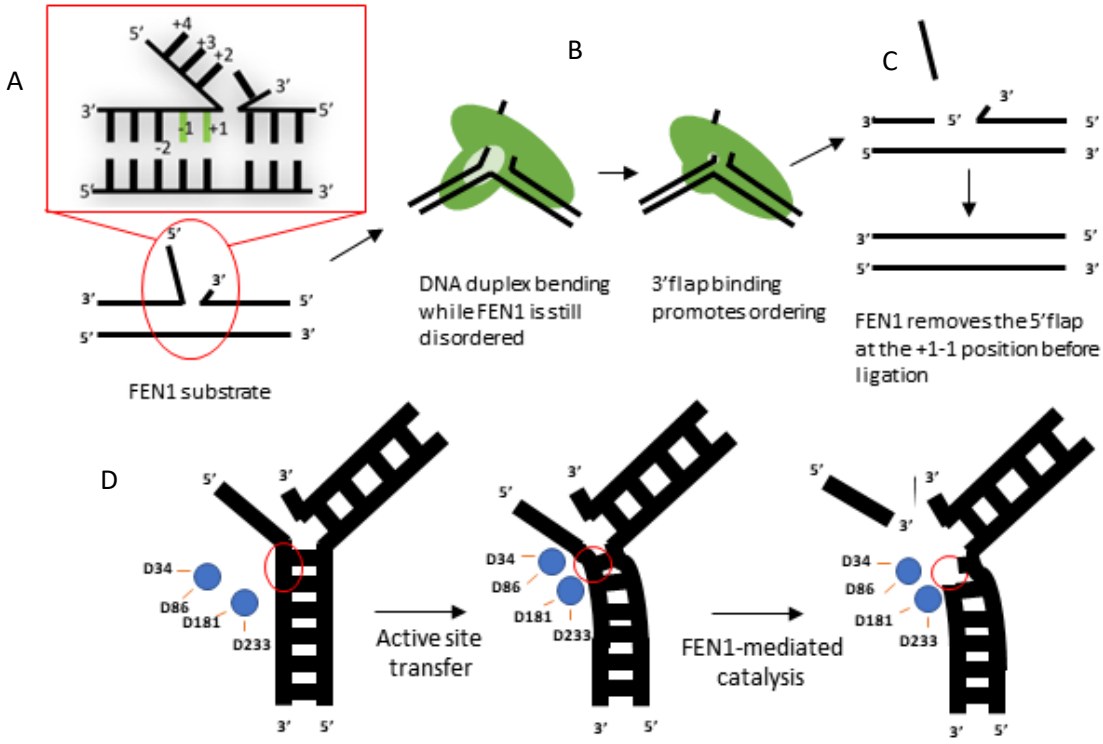


Figure 3.1 FEN1-mediated catalysis and active site transfer

***A**. FEN1 double flap substrate with a magnified view of the +1-1 bases located on the processed strand are highlighted in green. **B**. Substrate binding to disordered FEN1 results in duplex bending. 3'-flap binding to the loop-wedge motif leads to conformational changes and ordering of the helical gateway/cap region following threading of the 5'-flap into the active site. Active site transfer through DNA junction distortion produces a catalytically competent complex. **C**. FEN1 hydrolysis of the 5'-flap at the scissile phosphate before release of the 5'-flap and ligatable product. **D**. 5'-flap positioning in the active site prior to active site transfer. DNA conformational change occurs through distortion at the +1-1 position to orientate towards the active site metal ions. Following this, FEN1 catalysis of the 5'-flap once catalytic competency has been achieved. The +1-1 scissile phosphate is highlighted with a red circle.*

Enzyme kinetics has been pivotal in gaining greater insight into FEN1-mediated catalysis of flap-containing DNA structures. Kinetics measures the rate-limiting step of a reaction from conformational change to enzyme product release. When the reaction is saturated with substrate, enzyme product release is determined to be the multiple turnover rate (k_{cat}) in the FEN1 reaction of a double flap substrate. If the reaction is saturated with enzyme, then the single turnover rate can then be calculated (k_{STmax}) of the FEN1 reaction. This determines the rate-limiting step of the decay of the enzyme-substrate complex in the absence of substrate turnover. It has been suggested that one of these rate-limiting steps under single turnover conditions is conformational change of both the enzyme and substrate during FEN1-DNA reactions [124], [125].

Multiple turnover kinetics can be performed using continuous, and discontinuous methods. A FRET-based assay is an example of a continuous assay that involves labelling the substrate with fluorescent dyes on different areas of the substrate which results in a FRET signal. Following mixing of the reactants the substrate is processed leading to the abolishment of the FRET signal [67]. Multiple turnover assays are also performed discontinuously by manual mixing and quenching of the reaction mixture enabling analysis of sample composition including substrate and product concentrations. On the other hand, single turnover conditions can result in fast reactions that must be monitored using robotic equipment including stopped-flow and quench-flow. Stopped flow is a continuous assay where spectroscopic measurements of the reaction mixture are performed however quench-flow is performed discontinuously by rapid mixing and quenching of the reaction mixture before analysis of sample composition. From this, kinetic studies of WT-FEN1 showed that double flap substrates with a variable length 5'-flap and 1-nucleotide 3'-flap presented the highest k_{cat} value of 150-170 min^{-1} and a k_{STmax} value of 860-1400 min^{-1} indicating that enzyme-product release is rate-limiting in WT-FEN1 [126], [127]. Enzyme product release was also rate-limiting with single flap substrates and substrates containing three double-stranded regions although with reduced multiple, and single turnover rates.

Song et al have used 2-AP as a fluorescent label to monitor the kinetics of FEN1 reactions. 2-AP, an analogue of adenine, was used by the Song group because it can be readily introduced into DNA substrates and is a fluorescent nucleobase (Figure 3.2A &B). *Song et al* used two adjacent 2-APs in their experiments located either side of the scissile phosphate diester bond (Figure 3.1A). Furthermore, 2-AP fluorescence is quenched

by nucleobase stacking such as that found in the stacked duplexes of DNA substrates but is maximal in free solution [128]. The combination of stopped flow, quench flow and 2-AP labelling for assessing DNA conformation was used by the Song group to monitor the progression of FEN1-DNA reactions under excess enzyme concentration and to reach conclusions about rate-limiting conformational changes and product formation.

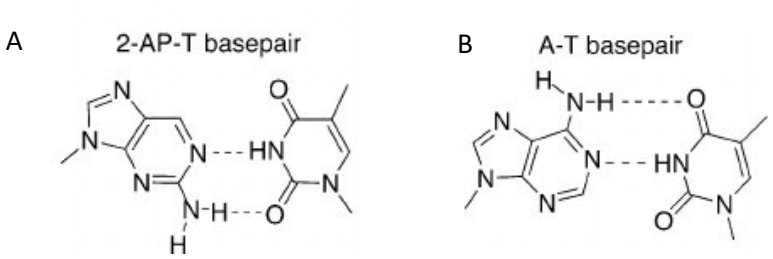


Figure 3.2 2AP structure
A*. 2-AP structure with Watson and Crick base pairing to thymine.* ***B****. Watson and Crick base pairing between adenine and thymine. Figure 6.2 is an altered image from the following reference* [129]*.*

The Song group reported new biophysical and kinetic evidence suggesting that conformational change is rate-limiting for the FEN1 mutant D34A. The residue D34 is an active site residue involved in divalent metal ion coordination. Reactions were monitored using 2-AP labelling at the +1-1 position (Figure 3.1A) around the scissile phosphodiester of a double flap substrate (Figure 3.3B). In the presence of Mg^{2+} ions, it is claimed that the rate of DNA bending was reported as $2.6x10^{-9}$ M^{-1} s^{-1} $k_{on\text{-}bend}$ and $2.2x10^{-9}$ M^{-1} s^{-1} for D34A. Samples in the presence of Ca^{2+} ions displayed no significant change in $k_{on\text{-}bend}$ compared to Mg^{2+} conditions. Stopped flow data in excess enzyme displayed a drastically slower DNA distortion rate ($k_{distort}$) for WT compared to D34A (WT $k_{distort}$: 25 s^{-1}, D34A $k_{distort}$: 0.1 s^{-1}) in the presence of Mg^{2+} ions resulting in a ~2-fold increase in fluorescence after 30 seconds (half-life 7 seconds). Contrastingly, quench flow data showed no product formation by D34A after 100 seconds using a 5'-fluorescein label whereas product formation was observed in other active site mutants Y40A, K93A, R100A and D181A (Figure 3.3A). Therefore it is suggested that the residue D34 may play a role in facilitating conformational change prior to hydrolysis [68].

Until the Song group publication in 2018, enzyme-product release was believed to be the rate-limiting step of FEN1-mediated hydrolysis of branched DNA substrates under multiple turnover conditions although the rate-limiting step under single turnover conditions remains relatively enigmatic. Therefore, the primary aim

of this chapter was to replicate the Song group experiments and investigate whether the burst phase was observed in the FEN1 active site mutant D34N. As produced by *Song et al*, we performed a kinetic assay under excess enzyme concentration conditions with 2.5 µM [E] and 0.5 µM [S]. However, in contrast to the Song publication, instead of replacing D34 with a hydrophobic alanine, we used a mutant containing the polar, uncharged amino acid Asparagine (D34N). The substrate used in our kinetic assay was a double flap structure with a 1-nucleotide 3'-flap and 14- nucleotide 5'-flap labelled with 2-AP at the +1 and -1 position (DF(14,1)$_{+1-1}$PP) (Figure 3.3B).

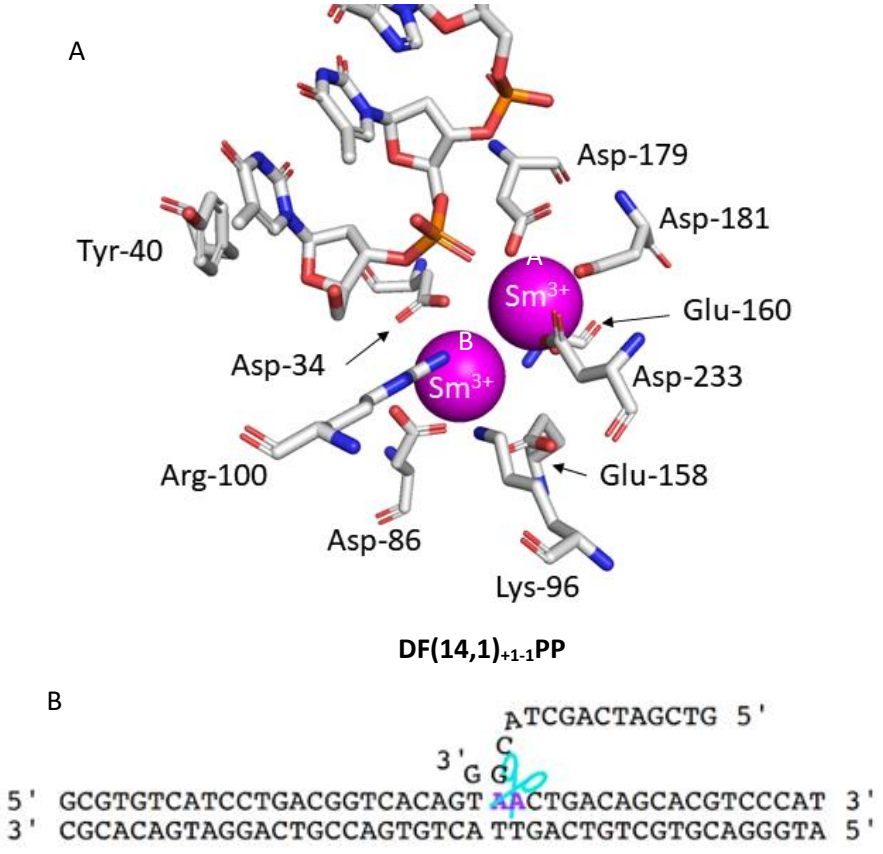


Figure 3.3 FEN1 active site residues and DF(14,1)$_{+1-1}$PP

***A**. Active site residues during FEN1 hydrolysis of the 5'-flap. D34 and D86 co-ordinate metal-ion B whilst D181 and D233 co-ordinating metal-ion A. R100 and K93 form contacts at the scissile phosphate during catalysis. Y40 stacks against the +1 base (PDB code: 3Q8K). Double bonded oxygens shown by two red lines; amino groups shown in blue; phosphate diesters shown in orange. **B**. Double flap substrate used by the song group with embedded tandem 2-AP bases. The same substrate was used to perform FEN1 2-AP assays within this research chapter. 'DF' represents a double flap; the left number in the bracket represents the length of the 5'-flap and the left number is the length of the 3'-flap; $_{+1-1}$ corresponds to the position of the 2-AP on the 5'-flap strand and 'PP' corresponds to two 2-AP bases within the substrate. The DF(14,1)$_{+1-1}$PP image in Figure 6.3 was taken from the following reference* [68].

3.2 Measuring the rate of 2-AP fluorescence change in D34N with DF(14,1)$_{+1-1}$PP

3.2.1 D34N and WT 2-AP assay

As previously described by *Song et al*, a 2-fold increase in fluorescence was observed between 0-30 seconds using the 2-AP probed substrate in Figure 3.3B upon mixing of DF(14,1)$_{+1-1}$PP, mutant D34A and Mg^{2+} ions. To replicate this, an assay was designed to monitor the 2-AP fluorescence intensity at 375 nm between 5-100 seconds using the substrate DF(14,1)$_{+1-1}$PP. The fluorescence intensity of the substrate in the absence of protein was approximately 35,000-40,000 AU independent of catalytic, or non-catalytic metal ions. Negligible fluorescence was recorded of solutions containing protein in the absence of substrate. When D34N was added to the substrate in either Ca^{2+}, or Mg^{2+} ion conditions, a small (<2-fold) enhancement in fluorescence intensity was recorded compared to the free substrate. This is likely to be the result of interactions of the substrate with the enzyme, where bending may occur at the DNA junction for the protein to accommodate the 5'-flap. However, the fluorescence intensity remains constant over 100 seconds implying minimal reaction has occurred in either metal condition on this timescale (Figure 3.4A). The same is also true when Ca^{2+} ions are added to the reaction containing WT protein. However, when Mg^{2+} ions is added to the WT reaction, an increase in fluorescence intensity was recorded with 450,000 AU after 5 seconds although this is expected as the substrate is then being hydrolysed resulting in unquenching of the +1-1 2-AP bases in the substrate. From this, there is no evidence of a 2-fold burst in 2-AP fluorescence intensity upon formation of the D34N-Mg^{2+} complex as reported by *Song et al* [68] (Figure 3.4B).

Although the data shown in Figure 3.4A has already shown a lack of increase in 2-AP fluorescence over 100 seconds with D34N compared to the stopped flow data with D34A from the Song group, we sought further confirmation of our findings by monitoring the fluorescence intensity after the sequential addition of each reaction component. Following the addition of Mg^{2+}, the reaction was monitored over 240 minutes to witness the reaction reach completion. As previously shown in Figure 3.4A, a 2-fold increase in fluorescence intensity was recorded following the addition of D34N. Other research groups have also reported the small 2-fold increase in 2-AP fluorescence intensity after the addition of protein although it was suggested that the increase is due to minor, non-specific distortions in the 5'-duplex of the substrate, resulting in partial unquenching of the 2-AP bases [130]. The fluorescence intensity then decreases slightly upon the addition of

Ca^{2+} ions to the reaction. Following the addition of Mg^{2+} ions, the fluorescence intensity decreased after two minutes. The straight-line in Figure 3.4C between the sequential addition of D34, Ca^{2+} and Mg^{2+} ions was speculated to be dilution of the reaction, however after normalising the fluorescence intensity for volume, the same decrease in fluorescence intensity was observed. Monitoring the reaction over 240 minutes began with a lag phase until the fluorescence intensity increased to ~320,000 AU after 240 minutes. This was denoted as 5'-flap hydrolysis and subsequent unquenching of the tandem 2-AP bases. Non-linear regression with one-phase association was poorly fitted to the data set with a rate of 0.002 min^{-1} (Figure 3.4D). This was expected as the D34N mutant severely impedes the reaction rate of 5'-flap hydrolysis.

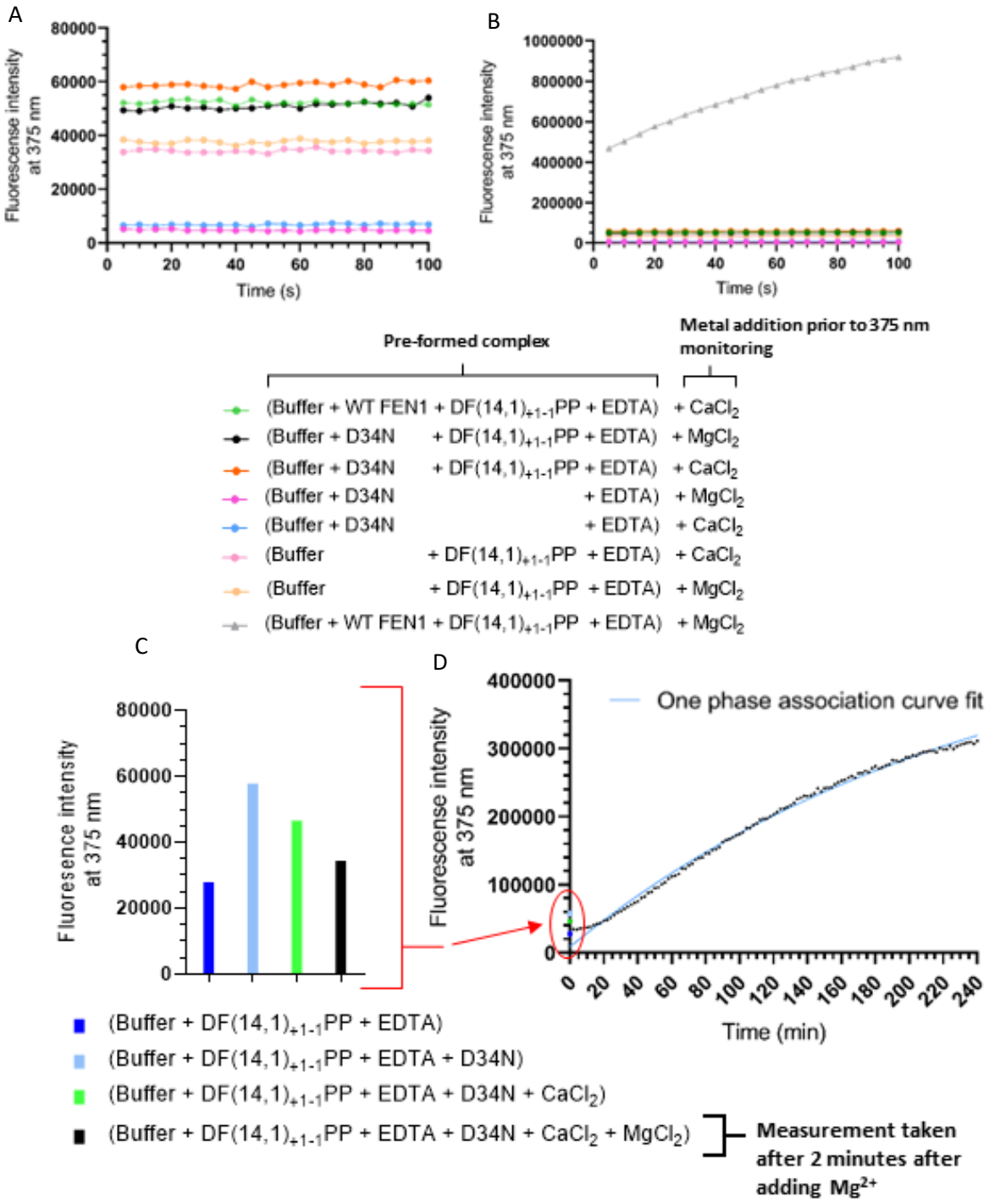


Figure 3.4 D34N assay over 100 seconds and 240 minutes

***A**. 5-100 second assay of 2.5 µM D34N and 0.5 µM DF(14,1)$_{+1-1}$PP. 50 mM HEPES pH 7.5, 100 mM KCl, 0.1 mg/mL BSA and 2.5 mM DTT ('buffer'), DF(14,1)$_{+1-1}$PP and 1 mM EDTA were mixed to form a pre-formed complex before the sequential addition of either 8 mM $MgCl_2$ or $CaCl_2$ prior to measurements being taken every 5 seconds at 375 nm. **B**. The 5-100 second assay compared to 2.5 µM WT and 0.5 µM DF(14,1)$_{+1-1}$PP upon the addition of 8 mM magnesium chloride. **C**. Individual measurements during the sequential addition of 2.5 µM D34N, 2 mM $CaCl_2$ and 16 mM $MgCl_2$. **D**. 240-minute assay of 2.5 µM D34N and 0.5 µM DF(14,1)$_{+1-1}$PP. The reaction was fitted to a non-linear regression model with one-phase association in GraphPad prism 9.0.*

An interesting finding was the lag-phase between 2-20 minutes for the D34N-DF(14,1)$_{+1\text{-}1}$PP reaction (Figure 3.4D). As previously mentioned, for FEN1 to be considered in a catalytically active state, conformational changes of both enzyme and substrate must occur. It is speculated that D34N leads to conformational states that are not catalytically viable for substrate processing. This results in FEN1 having to unfold and reposition the substrate into the correct orientation. This may explain why the fluorescence intensity does not rise immediately. Another explanation is non-specific hydrolysis of the 5'-flap by D34N. It is shown throughout the chapter that hydrolysis of the +1-1 position (Figure 3.1A) is performed by D34N as shown by unquenching of the 2-AP bases after 20 minutes although our assay does not directly describe where earlier hydrolysis events of the 5'-flap phosphodiester backbone occur prior to 2-AP hydrolysis. D34N may adopt a catalytically competent state but in fact incise at different phosphodiester positions of the 5'-flap prior to hydrolysis of the phosphodiester bond at the +1-1 position (Figure 3.1A). If non-specific hydrolysis occurred, our assay would not be able to monitor the non-specific hydrolysis activity due to the +1-1 2-AP bases being quenched. This is concordant with quench flow experiments where 5 nM DF(p5,1)5'FAM (Figure 3.5C) was mixed with 400 nM and 1 µM D34N, both conditions produces k_{ST} rates of 0.013 and 0.014 min^{-1} (Figure 3.5A). RP-dHPLC analysis of the products between 10-360 minutes show multiple products formed between 2–5-minute retention times (Figure 3.5B). Previous literature has suggested that 6-nucleotide products have retention time of 5.3-5.5 minutes using RP-dHPLC. This implies the remaining products after 5, 4 and 3-nucleotides in length with single nucleotide resolution columns [48]. Contrastingly, steady-state experiments of WT with 5 nM DF(p5,1)5'FAM (quenched at 10 minutes) show the 6-nucleotide product eluting at ~4.5 minutes (Figure 3.5D). Thus, allocations of exact product lengths cannot be confirmed for the D34N-SB5,1 quench flow data without denaturing PAGE analysis with molecular weight standards. Product drift and peak widening can also occur upon the overuse of RP-dHPLC columns.

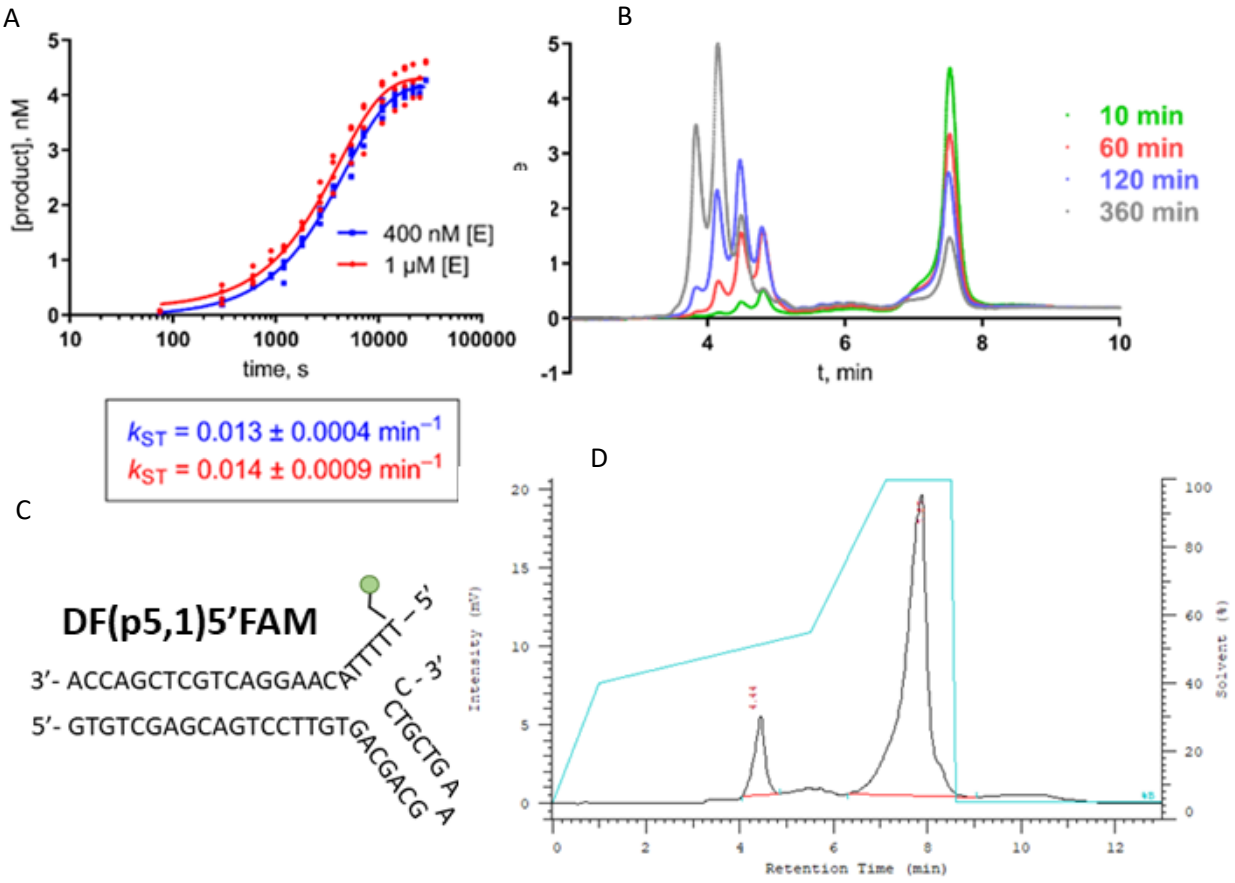


Figure 3.5 D34N single turnover experiments compared to WT

***A**. Quench flow experiments of 400 nM & 1 µM D34N with 5 nM DF(5,1)5'FAM. Reactions were performed in 50 mM HEPES pH 7.5, 100 mM KCl, 0.1 mg/mL BSA and 2.5 mM DTT. Quench flow experiments were kindly performed Dr Mark Thompson. **B**. RP-dHPLC analysis of product formation from 400 nM D34N and 5 nM DF(5,1)5'FAM after 10, 60, 120, and 360 minutes. Products were eluted using 2.5 mM TBAB and eluted with acetonitrile. RP-dHPLC analysis was kindly performed by Dr Mark Thompson. **C**. Static double flap structure of DF(p5,1)5'FAM. **D**. RP-HPLC analysis of 1 pM WT with 5 nM DF(p5,1)5'FAM after being quenched by 7M UREA, 80 mM EDTA. The substrate peak with a retention time of ~7.5 minutes and the 6-nucleotide product at ~4.5 minutes. RP-dHPLC analysis was performed using 2.5 mM TBAB and eluted with acetonitrile.*

3.3 Summary

To summarise, the 2018 publication by *Song et al* provoked interest due to DNA conformational change being presented as the rate-limiting step in human FEN1 catalysis. Because of this surprising finding, the primary focus of this chapter was to replicate the findings of the Song group using fluorescence-based assays under reagent conditions identical to the publication in 2018. Furthermore, the Song group presented a 2-fold enhancement in 2-AP fluorescence intensity should be observed with mixing D34A and substrate in the presence of Mg^{2+} ions (half-life: 7 seconds). This was attributed to a magnesium dependent substrate conformational change prior to reaction. However, through our assays, we were unable to replicate the results of the Song paper with our mutant D34N measuring time-points between 5-100 seconds and did not observe a progressively increase in fluorescence but instead observed steady but enhanced fluorescence which may be attributed to the 2-AP substrate binding by D34N. It is important to note that a different D34 mutant was used in this chapter in the form of D34N, rather than D34A. However, D34N could be argued as a less structurally deleterious mutant compared to D34A due to retaining polarity and being of similar size with respect to the carboxylic acid. The data provided in Figure 3.4D and Figure 3.5B would suggest that the D34N mutant may produce multiple products due to incorrect positioning of the 5'-strand in the active site. This follows a similar trend to the L53A mutant on double flap, and pseudo-Y substrates where 4-mer, 5-mer and 6-mer products are formed [48].

Chapter 4 Kinetic and biophysical characterisation of WT FEN1 and H2TH mutants

4.1 Introduction

FEN1 and its superfamily members share structural characteristics that allow them to function as efficient DNA processing enzymes, despite their inherent difference in substrate specificity. Firstly, All FENs induce a sharp bend in the substrate through the hydrophobic wedge (α2-α3) to enforce separation between the 5' and 3'-duplexes at the junction [115]. Moreover, the active site is highly conserved amongst the superfamily with carboxylate residues allowing divalent metal ion binding to propagate phosphodiester bond hydrolysis. The helical gateway at the base of α4 is also widely conserved amongst the family members. However, proposes a restrictive issue for bulky DNA bubbles and holiday junction substrates which cannot be threaded [115]. It is suggested that bubbles and holiday junctions' loop around the helical gateway, therefore allowing the substrates to be processed by XPG and GEN1 respectively. Finally, the family members also exhibit a downstream duplex H2TH:K^+ binding site with a high concentration of basic residues which is suggested to function as the primary DNA-protein binding site [115].

The overlap in FEN superfamily structural properties have allowed greater insight into the complex series of events regarding FEN1 substrate recognition, substrate orientation and hydrolysis of the 5'-flap. However, less information is available regarding the primary interaction of FEN1 and its substrate before 5'-flap threading, conformational ordering, and AST. As previously mentioned, the primary contact point between a FEN enzyme and its respective substrate is facilitated by the H2TH domain. The H2TH domain of FEN1 consists of 2 α-helices (α10-α11) separated by 2-turns in the human FEN1 structure, followed by a mono-valent metal binding site (K^+ ion) [131] (Figure 4.1A). Interestingly, archaeal FEN1 possess a, 3-turn motif (H3TH) although both human and archaeal FEN1 show high specificity towards K^+ ions due to K^+ affording intimate DNA contact [131]–[133] (Figure 4.1B).

Basic residues in the H2TH domain of human FENs and somewhat in Archaeal FENs have been suggested to play a role in positioning the 5'-flap towards the active site. Sequence alignments of human FENS (Figure

4.1C) demonstrated conserved arginine residues in FEN1 and EXO1 at position 239 on α10. Moreover, arginine and lysine residues are also present at position 245 for EXO1 and FEN1 respectively although no basic residues are observed at position 244 for all the family members except K244 in FEN1 on α11. (Figure 4.1C). Sequence alignments of the Archaeal H3TH domain also demonstrate basic residues on α10-α11 between positions 231-243 with additional tandem lysine or lysine-arginine residues along the sequence alignment (Figure 4.1C). The conserved basic residues in the H2TH:K^+ domain in FEN1 and EXO1 accompanied by both their substrates exhibiting similar threading and active sight transfer methods, it may be proposed that the basic residues conserved in these regions play a specific role in processing double flap substrates and nicks [115], [131], [134], [135]. Crystal structures of human FEN1 have also implied that residues R239, K244 and R245 create a track, affording electrostatic interactions between R245 and the minor groove of the template strand at 4.0-6.0 Å (Figure 4.1D). The basic side chains of R239 are 2.5 Å from a non-bridging oxygen of the substrate template strand whereas K244 sits between 4.5-5.0 Å from the phosphodiester backbone and therefore allowing electrostatic interactions to occur.

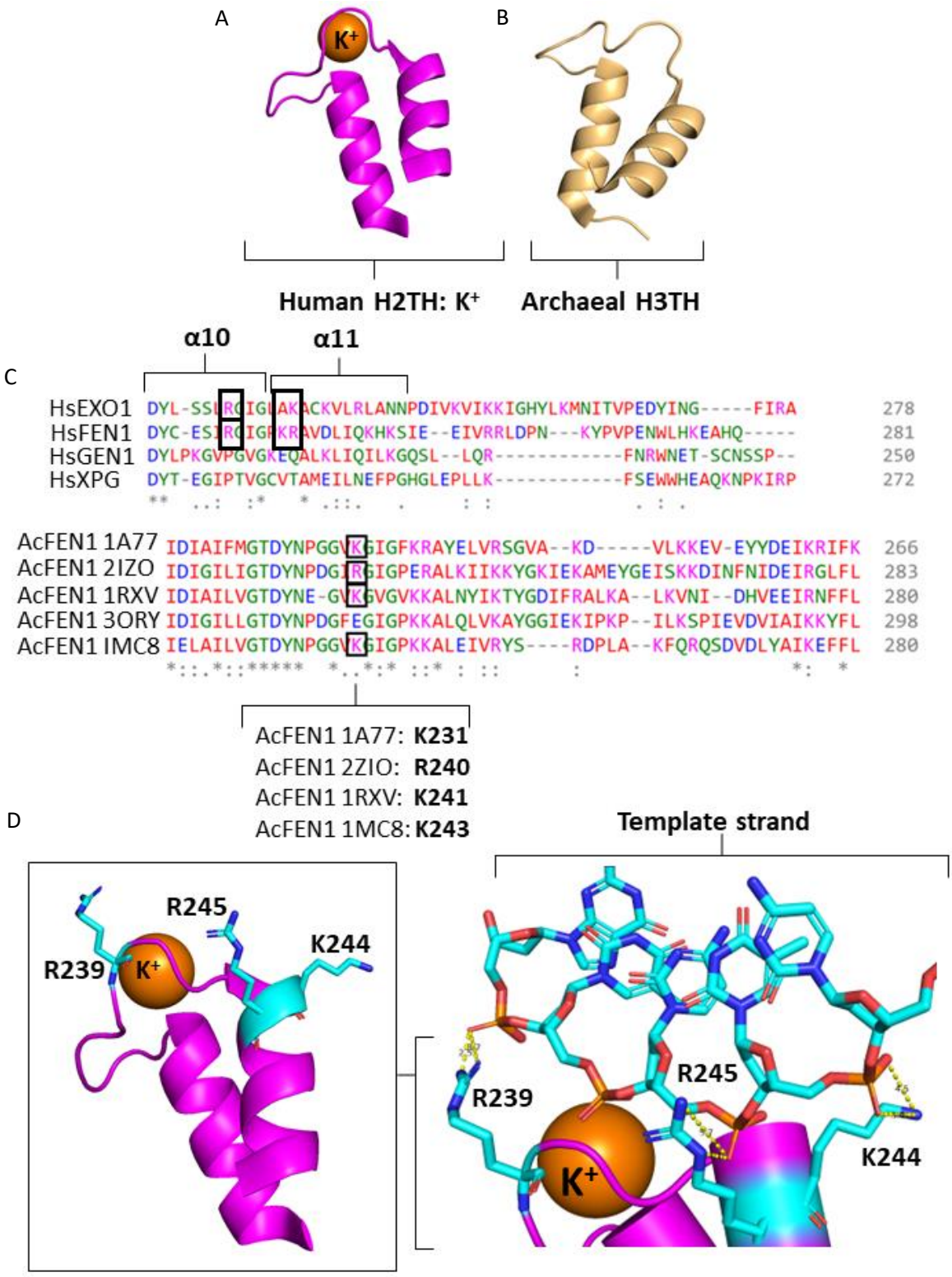


Figure 4.1 The varying H2TH domains of FEN1 across different areas of life

***A**. H2TH: K⁺ domain of human FEN1 (PDB code 3Q8K).* ***B**. Archaeal FEN1 H3TH domain from D. amylolyticus (PDB code: 3ORY)* ***C**. Clustal Omega protein sequence alignment of the human FEN and archaeal FENs (1A77: M. jannaschi; 2IZO: S. solfataricus; 1RXV, A. fugidus; 3ORY: D. amylolyticus; 1MC8: P. horikoshii) at the H2TH/H3TH domain. Basic residues are highlighted by black boxes. Fully conserved residues are shown by an asterisk whereas partially conserved residues are shown by one, or two dots.* ***D**. Structural image of the basic conserved residues R239A, K244A and R245A (PDB code: 3Q8K) and the phosphodiester backbone of the template strand. Phosphorus atoms shown in orange; oxygens shown in red; nitrogen's shown in blue, and carbons shown in turquoise.*

As previously described in section 3.1, enzyme kinetics is a useful tool when characterising an enzyme and how its mutants affect activity. Kinetic characterisation of enzyme-substrate reactions can be performed under multiple-turnover conditions to measure changes in the turnover number k_{cat} of a specific substrate. By performing multiple turnover kinetics under varying substrate concentrations across multiple orders of magnitude, the normalised initial rates (v/E) can be used to calculate the Michaelis-Menten constant K_M. The K_M is described as the substrate concentration at which half maximal velocity is achieved and depicts the stability of all enzyme-substrate/enzyme-product bound species in the reaction. Therefore, the lower the K_M, the higher the affinity an enzyme possesses for a specific substrate. Furthermore, normalisation of k_{cat} by the K_M (k_{cat} /K_M) allows comparison of the catalytic efficiency an enzyme possess for its substrate [136].

Prior work on potential DNA binding residues including lysine and arginine were performed on the human FEN1 at the H2TH domain. The residues R239 and K244/R245 were substituted to alanine before being mixed with varying 5'-single flaps and nicked substrates, which are now known to be sub optimal substrates compared to DF(p5,1) from later reports [137] [46]. It was shown that the nuclease activity decreased with K244A/R245A whereas nuclease activity of R239A remained similar to the WT protein. In addition to nuclease activity, binding affinity studies were also performed, from this it was suggested that R239A displays similar binding to the WT protein and K244A/R245A inherently decreased [137].

Structural studies of the FEN superfamily have provided greater insight into the role of the H2TH domain. Intermediate structures of the human EXO1-DNA reaction led to the proposal of the H2TH domain functioning in a processive, sliding motion during catalysis. Their findings demonstrated hydrogen bonds between the main chain amides of residues 232-237 and the phosphate oxygens of the template strand [50], [115]. moreover, the K^+ binding site was co-ordinated by the carbonyl groups of S222, I233 and the hydroxyl of S229 to the template strand. Reports also suggested that the K^+ binding site was not formed in the absence of substrate. This implies the binding site is repeatedly reformed during sliding the phosphodiester backbone across the H2TH domain throughout the reaction. However, it is unclear whether the binding site can be reconstituted in the absence of potassium or sodium ions. A similar site to the H2TH domain of FEN1 and EXO1 has been identified in DNA polymerases with the 'anchoring' and 'sliding' motion of this DNA binding domain assisting in substrate processing [50], [115], [138].

The literature has proposed that arginine and lysine residues in the H2TH may play role an important role in FEN-mediated substrate processing and catalysis through activity assays and X-ray crystallographic models. Although, these claims remain largely untested. Therefore, in this chapter, discontinuous kinetic experiments will be performed under multiple turnover conditions and analysed by RP-dHPLC. This will provide Michaelis-Menten constants of the WT and H2TH mutant FEN1 proteins with the optimal substrate DF(p5,1)5'FAM. Furthermore, discontinuous single turnover experiments using quench flow will be undertaken to gain the k_{STmax} of the WT and H2TH mutants. The combination of both multiple, and single turnover kinetics rates will elucidate whether enzyme-product release is rate limiting in R239A, K244A and R245A. This chapter will also assess the change in DNA conformation at the scissile phosphate of the 5'-flap for R239A, K244A, R245A and WT. To do this, exciton-coupled circular dichroism (ECCD) was performed under non-catalytic reaction conditions using 2-AP labelled, double flap substrates at the +1-1 (DF(p5,1)$_{+1\text{-}1}$PP) and -1-2 (DF(p5,1)$_{+1\text{-}1}$PP) positions. The use of tandem 2-APs within the substrate produces an exciton pair that will fluoresce at 326 nm exclusively upon change in DNA conformation at the +1-1 or -1-2 positions of the substrate.

4.2 Site directed mutagenesis of and construct production of R239A, K244A and R245A

Site directed mutagenesis with ligation independent cloning was performed to substitute arginine and lysine residues for alanine in the H2TH domain of FEN1. Agarose gel electrophoresis was used to analyse the molecular weights of all three constructs. This showed all three constructs displayed bands at ~6000 bp before (pre) and after (post) digestion with DPN1 which was believed to be R239A, K244A & R245A in the linearised form (Appendices Figure 8.2). The concentration of each construct was determined by the absorbance at 260 nm prior to sequencing (See section 9.1.1 for H2TH mutant construct sequencing).

4.3 Expression and purification of WT, R239A, K244A & R245A

The expression and purification of WT and its associated mutants has been well documented [72], [139]. Constructs were expressed using Rosetta 2 DE3 cell lines. All constructs were grown to a pre-IPTG OD_{600} value of 0.6-0.9. The clarified lysates of WT and mutant FEN1 were purified with Ni^{2+} IMAC chromatography. WT and mutant FEN1 constructs contained an his-tag on the C-terminus and therefore bound to the Ni^{2+} charged column and eluted upon the addition of imidazole. The elution fractions were diluted 2-fold and then purified by DEAE anion exchange chromatography to remove DNA contamination from the sample whereas the FEN1 protein would elute into the flow-through. The sample was diluted 2-fold to reduce the salt concentration of the sample prior to purification with heparin which mimics negatively charged macromolecules like DNA and thus, promotes binding of FEN1. FEN1 was eluted with a salt gradient of 0.1-1.0 M. The heparin elution peak fractions were concentrated and then buffer exchanged using a 2610 desalting column. Peak fractions containing FEN1 were concentrated to a concentration of 100 µM and stored in 50 % glycerol. To test the purity of the sample, 10 µg of sample were analysed by SDS-PAGE and assessed for additional contaminant bands (Figure 4.2) and resulted in similar molecular weight bands proposed to be WT and mutant FEN1.

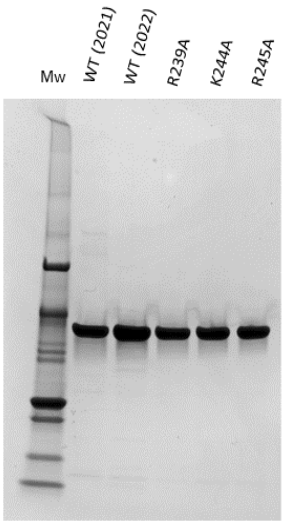


Figure 4.2 SDS-PAGE analysis of previously produced WT FEN1 stocks and H2TH mutants.
4-20 % Tris-Glycine SDS-PAGE analysis of WT FEN1 stocks from 2021 and 2022 accompanied by R239A, K244A and R245A enzyme stocks. Exactly 10 ug was loaded into each well to identify the relative purity of each sample. Molecular markers ranging from 250-10 kDa were loaded onto the left most lanes as a standard to compare sample bands within each enzyme stock.

4.4 Kinetic characterisation of WT, R239A, K244A and R245A

4.4.1 Characterisation of WT and H2TH mutants under steady-state conditions

The rate of reaction catalysed by FEN1 and the H2TH mutants was measured using a double flap substrate DF(p5,1)5'FAM (Figure 4.3A) which contains a 5'-fluorescein label on the 5'-flap. A discontinuous assay was employed to measure the rate of reaction whereby incubated samples were quenched using a solution comprised of 8 M UREA and 80 mM EDTA. Sample analysis was performed using RP-dHPLC equipped with a fluorescence detector. The faster eluting product peak and substrate peaks were quantified before being plotted against time to gain the initial rates of reaction. Throughout these experiments, two batches of WT FEN1 were compared, one batch had been stored at -20 °C for over a year and another batch produced three months prior to the nuclease assays stored at -80 °C. The initial rate of reaction was measured using 100 nM substrate and 6 pM of enzyme at pH 7.5 prior to normalising for enzyme concentration (v/[E]). The normalised initial rates were calculated as 213.3 min^{-1} for the older sample and 292.6 min^{-1} for the newer WT sample (Figure 4.3B) with residual plots showing similar deviation between both samples (Figure 4.3C). Loss in activity was still within the same order of magnitude, although a small decrease in 5'-flap processing was still observed with the old WT FEN1 stock. This was most likely due to enzyme degradation after being stored at -20 °C for over a year whereas the new WT stock was flash frozen in liquid nitrogen before -80 °C storage, accompanied by a reduced storage time. Therefore, to gain Michaelis-Menten parameters, the new WT FEN1 stock will be used.

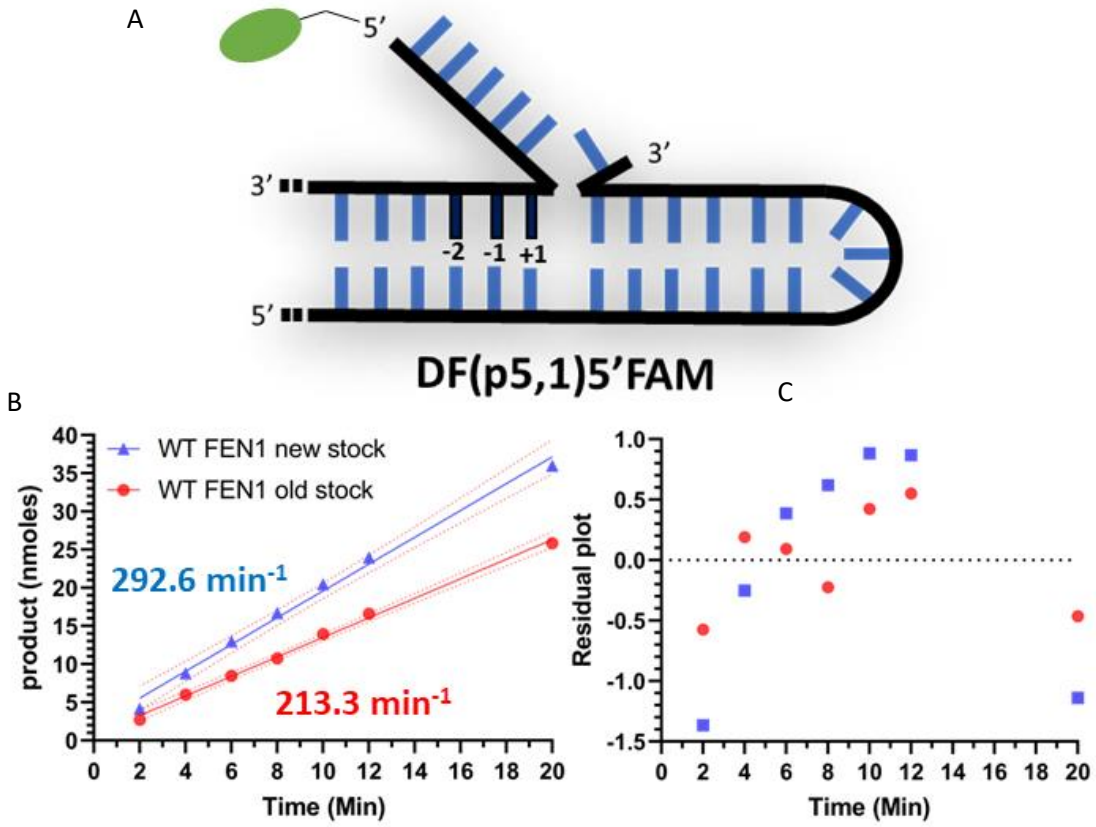


Figure 4.3 Normalised initial rates of reaction for the old and new WT stocks

***A**. DF(p5,1)5'FAM substrate structure with 5'FAM attached to the 5'-flap shown by a green oval. **B**. Initial rate graph of WT FEN1 stock from 2021 and 2022 at 6 pM enzyme mixed with 100 nM DF(p5,1)5'FAM between 2-20 minutes. The initial rates were calculated using simple linear regression affording 1.280 nM/min^{-1} for the old WT and 1.750 nM/min^{-1} for the newer WT stock. To conserve column resolution for future kinetic assay analysis, the experimented was repeated n=1 for each timepoint. **C**. Residual plot against the X axis was also plotted to analyse the spread against the fitted curve by linear regression. Calculated initial rate values as described in the materials and methods chapter and V/E k_{cat} turnover numbers.*

Following completion of the WT nuclease assay, the same preliminary assay was performed for the H2TH mutants R239A, K244A and R245A at 6 pM enzyme and 100 nM DF(p5,1)5'FAM. Normalised initial rates of reaction were 110, 62 and 97 min^{-1} respectively (Figure 4.4A, B & C) with the residual plots all within the same magnitude order of deviation from the curve fit (Figure 4.4D).

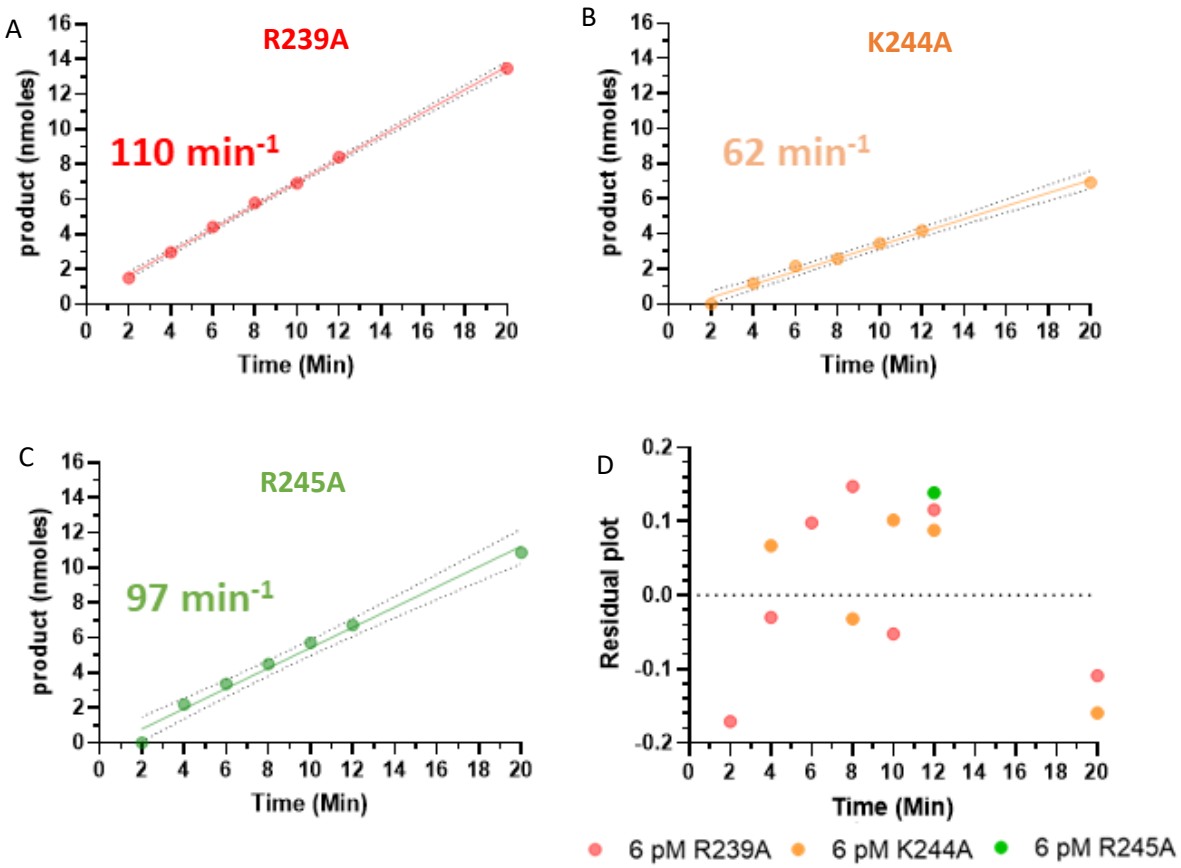


Figure 4.4 Initial rates of reaction for R239A, K244A and R245A

***A**. Preliminary nuclease assay of 6 pM R239A and 100 nM DF(p5,1)5'FAM with an initial rate of 0.662 nM/min^{-1}. **B**. Preliminary nuclease assay of 6 pM K244A and 100 nM DF(p5,1)5'FAM with an initial rate of 0.373 nM/min^{-1}. **C**. Preliminary nuclease assay of 6 pM R245A and 100 nM DF(p5,1)5'FAM with an initial rate of 0.580 nM/min^{-1}. To preserve column resolution for future nuclease assay analysis, zero replicates were undertaken for the preliminary initial rates of the H2TH mutants. **D**. Residual plot of part A, B & C. All datasets were fit using simple linear regression. For these initial rate scouting experiments, data points were repeated n=1 to preserve column resolution for future kinetic assays.*

Once the initial rates were obtained, a full Michaelis-Menten analysis curve of WT, R239A, K244A & R245A FEN1 was undertaken to determine K_M and k_{cat} values for each mutant in comparison to WT. For this experiment, reactions were performed in triplicate (N=3). The Michaelis-Menten parameters determined for the three mutants are shown in

Figure 4.5A, B, C & D. Both R239A and K244A had similar K_M and k_{cat} values compared to WT. Contrastingly, R245A led to a 6-fold increase in K_M and a relatively unaffected k_{cat} Ultimately, what can be extrapolated from the Michaelis-Menten curves of all three mutants is that the catalytic turnover of the double-flap substrate DF(p5,1)5'FAM is not dramatically affected by the H2TH substitutions for alanine. However, it can be suggested that the removal of basic residues from the H2TH domain can result in a lack of electrostatic anchoring of the template strand to FEN1 via non-bridging oxygens of the phosphodiester backbone and therefore result in FEN1's reduced affinity towards DF(p5,1)5'FAM without severely impeding the reaction. It could be proposed that the K^+ binding site acts as the main enzyme-DNA binding region at the H2TH domain and consequently, basic residues on α10-α11 increase the binding efficiency through additional electrostatic interactions between FEN1 and the template strand [140].

WT FEN1 **R239A**

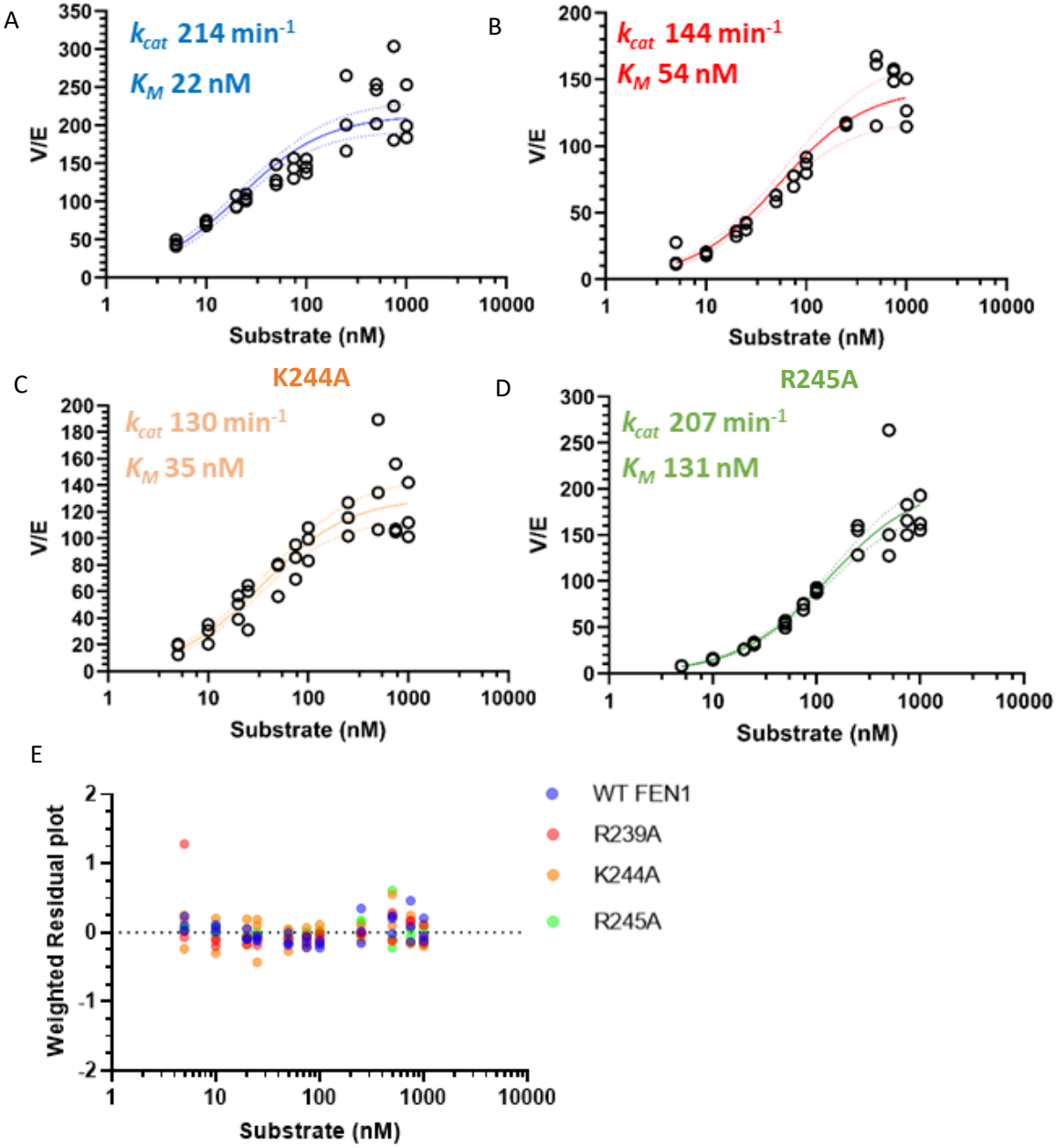

A
k_{cat} 214 min^{-1}
K_M 22 nM
V/E
Substrate (nM)
B
k_{cat} 144 min^{-1}
K_M 54 nM
V/E
Substrate (nM)
C
K244A
k_{cat} 130 min^{-1}
K_M 35 nM
V/E
Substrate (nM)
D
R245A
k_{cat} 207 min^{-1}
K_M 131 nM
V/E
Substrate (nM)
E
Weighted Residual plot
Substrate (nM)
WT FEN1
R239A
K244A
R245A

Figure 4.5 Michaelis-Menten parameters of WT, R239A, K244A & R245A

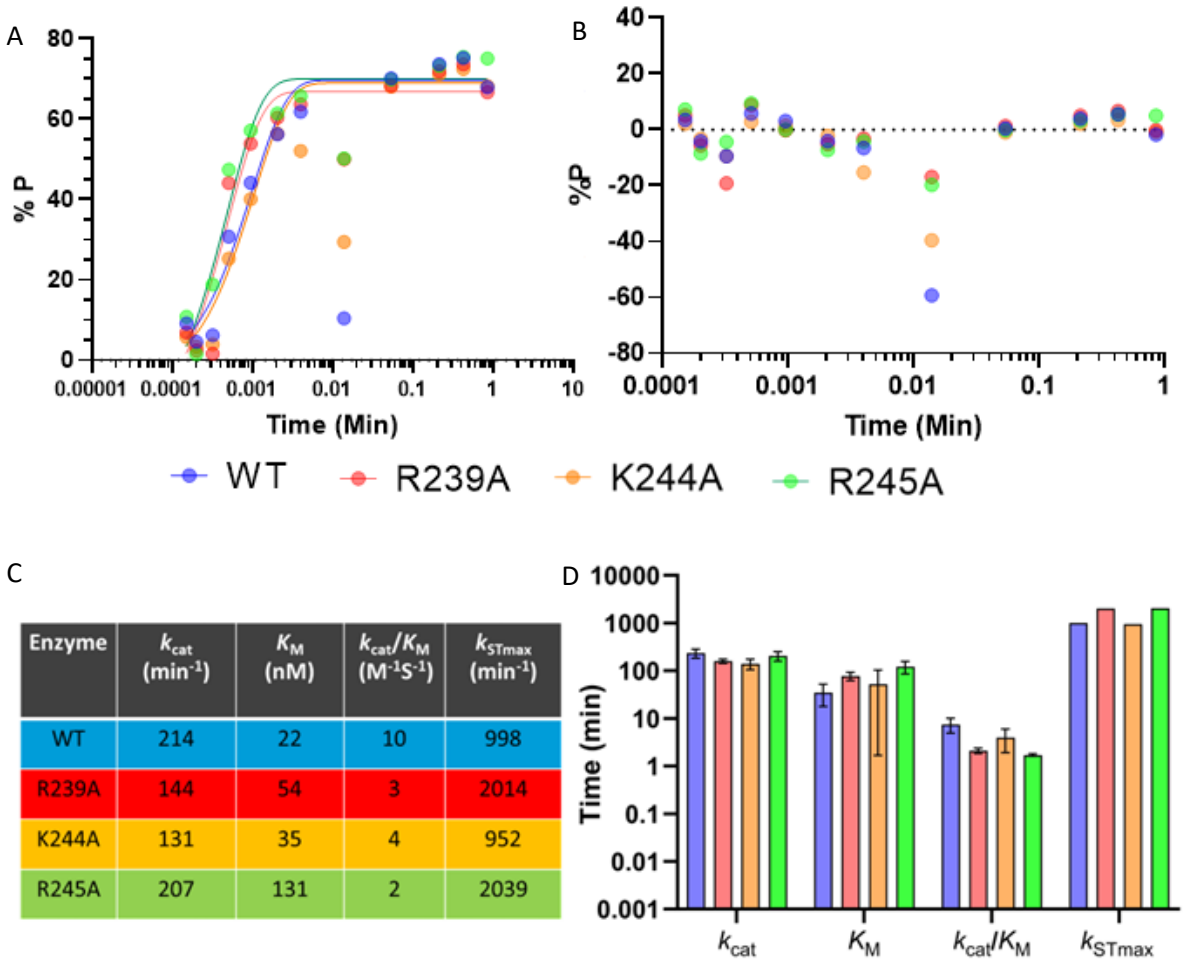
***A**. Michaelis Menten of WT FEN1 (2022 stock) (blue) with a fitted curve and 95 % confidence limit. **B**. Michaelis-Menten of R239A (red) with a fitted curve and 95 % confidence limit. **C**. Michaelis-Menten of K244A (orange) with a fitted curve and 95 % confidence limit. **D**. Michaelis-Menten of R245A (green) with a fitted curve and 95 % confidence limit. **E**. Residual plot against the substrate concentration to analyse the spread against the fitted curve. WT FEN1 (2022 stock) displayed in blue; R239A displayed in red; K244A displayed in orange, and R245A displayed in green. Michaelis-Menten experiments were performed in triplicate and analysed via RP-dHPLC as described in the materials and methods section. Michaelis-Menten graphs were plotted with the k_{cat} (min^{-1}) (also shown as V/E) and log_{10} substrate concentration (DF(p5,1)5'FAM) between 5-1000 nM. The datasets were plotted using non-linear regression and the Michaelis-Menten parameter in GraphPad Prism. All graphs were weighted for $1/Y^2$ to account for error in higher substrate concentration data points. Parts A, B, C and D were repeated n=3 using real repeats to calculate initial rates before calculating Michaelis Menton parameters. Refer to appendices Figure 8.3* for Michaelis Menton curves of individual replicates using a linear x-axis.

4.4.2 Characterisation of WT and H2TH mutants under single turnover conditions

Steady-state kinetics of WT and mutant FEN1 produced marginal changes in the k_{cat}. Although multiple turnover rates of WT and the H2TH mutants produced similar results, it is possible that this masks changes in the rates of other steps in the FEN1 reaction pathway, as product release is the rate-limiting step in the FEN1 reaction. Quench flow experiments under single-turnover conditions allow maximal single turnover rates (k_{STmax}) to be determined. Determination of the maximal single turnover rates was used to investigate whether enzyme-product release remains the rate limiting step during FEN1-mediated catalysis of double flap substrates following alanine mutations in the H2TH domain. Quench flow experiments were performed under single-turnover conditions with 5 nM DF(p5,1)5'FAM and 8 × K_M for the enzyme concentration of each mutant from

Figure 4.5. The reactions were performed using rapid liquid handling apparatus and the reactions were quenched with 8 M UREA/80 mM EDTA to unfold the FEN1 protein and chelate the Mg^{2+} metal ions respectfully to ensure the reaction was stopped. The composition of the sample was then analysed by RP-dHPLC.

The single-turnover reaction curves and rates are shown in Figure 4.6. All four curves displayed anomalous data points at 8.4 $x10^{-2}$ minutes and remain outside the trend of product formation from all four samples (Figure 4.6A). This may have occurred by user error via improper sample injecting and mixing into the quench flow device or partial contamination from previous samples using earlier timepoints shown in the residual plot (Figure 4.6B). It should be noted that old WT stocks were used for the quench flow experiments as the positive control as the newer WT batch was exhausted when conducting ECCD experiments. Nevertheless, WT k_{STmax} rates from Figure 4.6C are near identical to rates reported in the literature [127]. Comparison of all the maximal single turnover rates demonstrated that the H2TH mutants did not dramatically alter maximal activity compared to the WT enzyme. A summary of all the kinetic results (Figure 4.6C & D) demonstrate that k_{cat} and k_{STmax} remained within the same order or magnitude although the catalytic efficiency of the mutants was affected. Notably, for both the WT and mutant proteins k_{STmax} is 5-10-fold greater than k_{cat} indicating that product release is rate limiting in all three H2TH mutants.



C

Enzyme	k_{cat} (min^{-1})	K_M (nM)	k_{cat}/K_M ($M^{-1}S^{-1}$)	k_{STmax} (min^{-1})
WT	214	22	10	998
R239A	144	54	3	2014
K244A	131	35	4	952
R245A	207	131	2	2039

Figure 4.6 Single-turnover experiments of WT and H2TH mutants using quench flow
***A**. Single turnover kinetic reaction of WT and the H2TH mutants and DF(p5,1)5'FAM. Graphs were fit to a single exponential curve. Due to time constraints and descreasing column resolution, single turnover experiments were performed n=1.* ***B**. Residual plot of part A.* ***C**. In figure table showing all rates collected from the steady state and single turnover experiments of WT and the H2TH mutants with DF(p5,1)5'FAM. This chart shows one of n=3 replicates and the n=1 data set replicate for the k_{STmax} rates.* ***D**. Bar chart comparison of kinetic rates measures on a log_{10} scale.*

In summary, from these experiments under multiple and single turnover conditions, substitutions of H2TH positively charged residues to alanine do not impede the FEN1-mediated catalysis of DF(p5,1)5'FAM to the same extent as other mutants within the active site or at the 3'-flap binding site. Despite this, R245A did show a 6-fold increase in K_M and the most proficient decrease in catalytic efficiency implying its importance over K244A and R245A in forming the electrostatic contacts with the template strand.

4.5 Measuring the effects of the H2TH mutants on AST using ECCD and 2AP labelled substrates

4.5.1 ECCD experiments with DF(5,1)$_{+1\text{-}1}$PP and DF(5,1)$_{-1\text{-}2}$PP

Following determination of Michaelis-Menten and single turnover parameters of the H2TH mutants, it was shown that catalytic turnover was minimally affected compared to initial enzyme-substrate binding. Previously, the DNA conformational changes that occur upon untwisting of the DNA to transfer it to the FEN1 active site have been measured using substrates containing tandem 2-APs. To monitor changes in DNA conformation, 2-AP adenine bases were positioned at the +1-1 and -1-2 positions of a double flap substrate containing a 5-nucleotide 5'-flap and a 1-nucleotide 3'-flap Figure 4.7A. Furthermore, ternary complexes of metal ion-enzyme-substrate were formed under non-catalytic conditions in the presence of Ca^{2+} ions to ensure phosphodiester backbone hydrolysis was fully inhibited [51], [67].

Tandem 2-APs in the DNA substrate produce an exciton pair that is sensitive to the orientation of the two nucleobases [67], [141]. The substrates used in these experiments are shown in Figure 4.7A. The 2-AP exciton coupling (EC) has a maxima at approximately 326 nm, a region of the CD spectra that is transparent to protein and other nucleobases. ECCD results showed there was no change in signal in either the DF(p5,1)$_{+1\text{-}1}$PP or DF(p5,1)$_{-1\text{-}2}$PP in the presence of Ca^{2+} ions or EDTA without protein (Figure 4.7B). When WT enzyme was added in the presence of Ca^{2+} ions, the signal at 326 nm was reduced with both DF(p5,1)$_{+1\text{-}1}$PP and DF(p5,1)$_{-1\text{-}2}$PP (Figure 4.7C &D). Thus, in the presence of Ca^{2+} ions, significant DNA distortion is observed in complex with WT FEN1. This is likely due to placement of the scissile phosphate towards the non-catalytic Ca^{2+} ions in the FEN1 active site.

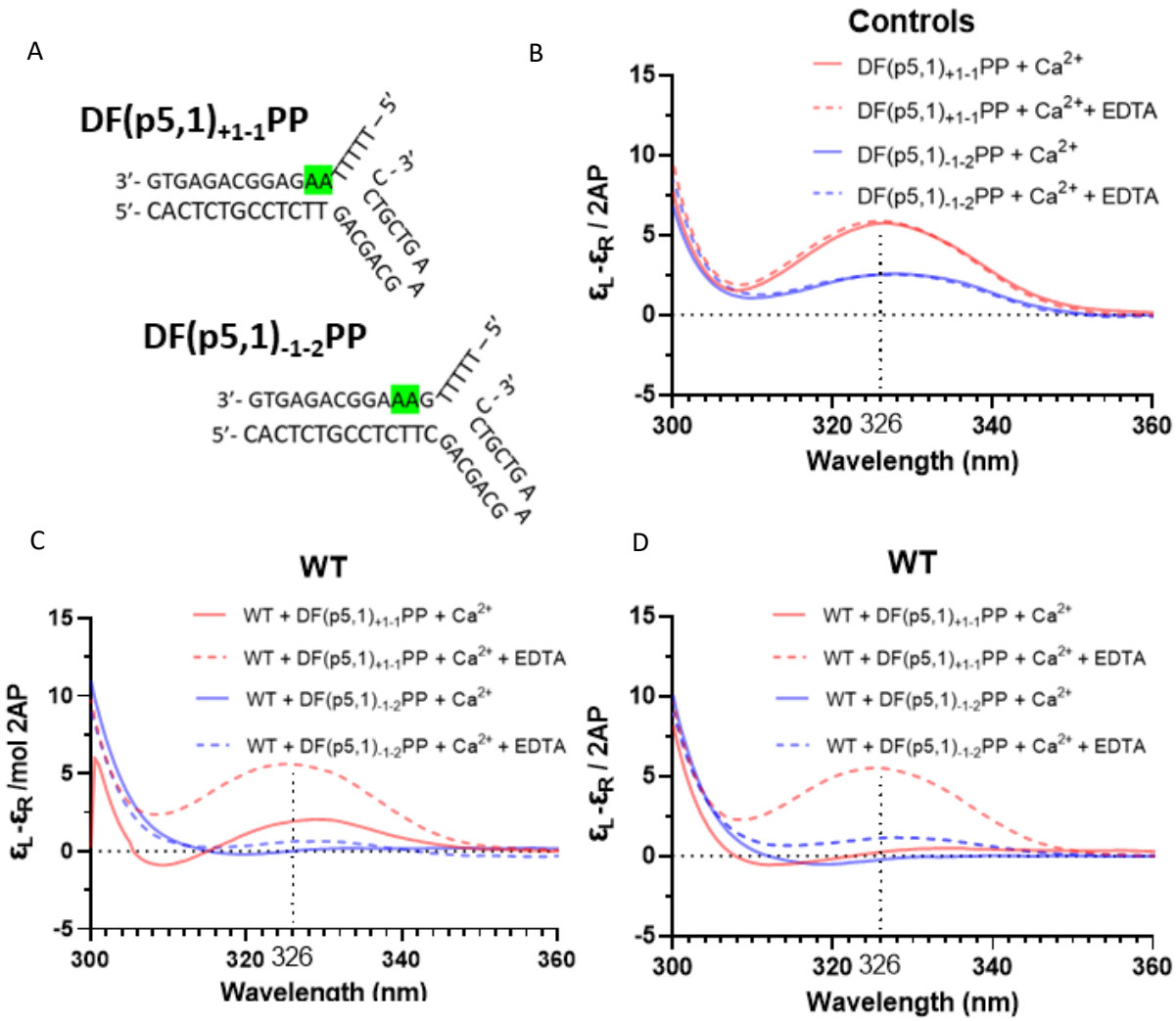


Figure 4.7 ECCD experiments of WT with DF(5,1)$_{+1-1}$PP and DF(5,1)$_{-1-2}$PP

A. *DF(p5,1)$_{+1-1}$PP or DF(p5,1)$_{-1-2}$PP substrates used to assess the changes in DNA conformational change. DF represents the presence of a double flap; numbers in parentheses correspond to the length of the 5'-flap (left) and 3'-flap (right); +1-1 or -1-2 corresponds to the placement of the DNA label on the 5'-strand and PP corresponds to the tandem 2-AP adenine molecules in the substrate.* ***B***. *ECCD Spectra of DF(p5,1)$_{+1-1}$PP or DF(p5,1)$_{-1-2}$PP in the presence of Ca^{2+} and EDTA. C. ECCD spectra of WT with DF(p5,1)$_{+1-1}$PP or DF(p5,1)$_{-1-2}$PP in the presence of Ca^{2+} ions and EDTA.* ***D***. *WT replicate spectra of part C.*

ECCD spectra were recorded for the H2TH mutants with DF(p5,1)$_{+1-1}$PP or DF(p5,1)$_{-1-2}$PP in the presence of Ca^{2+} ions. Following this, EDTA was added to chelate the Ca^{2+} ions and produce a spectra without divalent metal occupancy of the FEN1 active site. In the presence of DF(p5,1)$_{+1-1}$PP and Ca^{2+} ions, the R239A and R245A signal at 326 nm reached approximately 3.0 and 3.5 εL-εR /mol 2-AP respectively (Figure 4.8A & C), a near 50 % reduction in signal compared to WT under the same substrate and metal conditions. ECCD spectra of the H2TH mutants in the presence of DF(p5,1)$_{-1-2}$PP with Ca^{2+} ions produced similar results to the WT protein with only R239A showing higher signal compared to the WT positive control (~0.2 εL-εR /mol 2AP) (Figure 4.8A). K244A exhibited the same change in 2-AP signal within standard error of the WT FEN1 control (Figure 4.8B) and R245A exceeded the WT FEN1 control in levels of signal reduction below baseline (Figure 4.8C). Plotting the magnitude of exciton coupling signal at 326 nm shows reduction in 2AP signal at the +1-1 position of the ss/ds junction with all three H2TH mutants (Figure 4.8D) however increased DNA distortion and bending at the -1-2 position in mutants K244A and R245A (Figure 4.8E). Nevertheless, although the magnitude of changes is altered with respect to WT, the same type of change is observed in each case with, or without divalent metals.

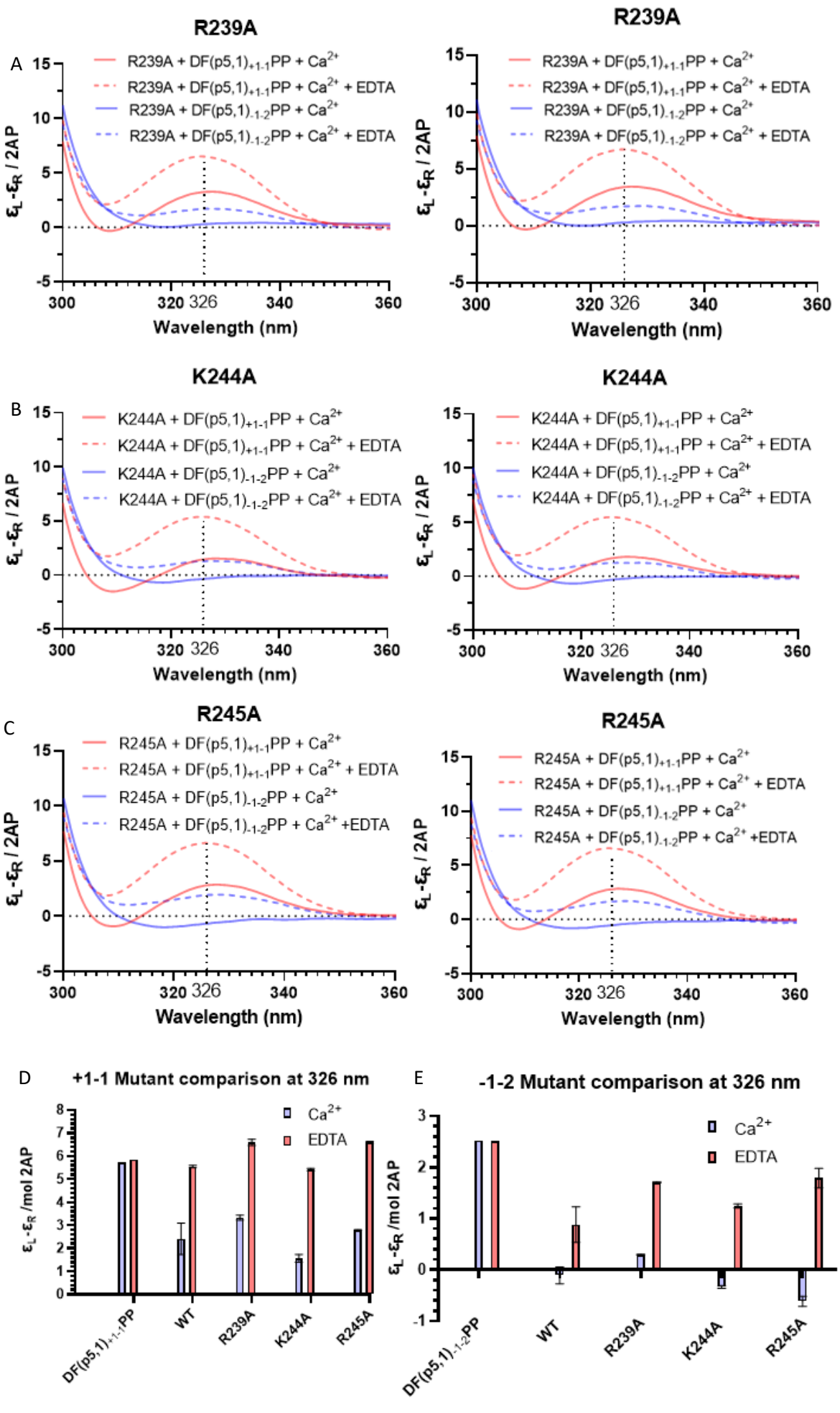

A
R239A
R239A + DF(p5,1)+1-1PP + Ca2+
R239A + DF(p5,1)+1-1PP + Ca2+ + EDTA
R239A + DF(p5,1)-1-2PP + Ca2+
R239A + DF(p5,1)-1-2PP + Ca2+ + EDTA
εL-εR / 2AP
Wavelength (nm)
B
K244A
K244A + DF(p5,1)+1-1PP + Ca2+
K244A + DF(p5,1)+1-1PP + Ca2+ + EDTA
K244A + DF(p5,1)-1-2PP + Ca2+
K244A + DF(p5,1)-1-2PP + Ca2+ + EDTA
C
R245A
R245A + DF(p5,1)+1-1PP + Ca2+
R245A + DF(p5,1)+1-1PP + Ca2+ + EDTA
R245A + DF(p5,1)-1-2PP + Ca2+
R245A + DF(p5,1)-1-2PP + Ca2+ +EDTA
D
+1-1 Mutant comparison at 326 nm
Ca2+
EDTA
εL-εR /mol 2AP
DF(p5,1)+1-1PP
WT
R239A
K244A
R245A
E
-1-2 Mutant comparison at 326 nm
DF(p5,1)-1-2PP

Figure 4.8 ECCD spectra of the H2TH mutants with DF(p5,1)$_{+1-1}$PP and DF(p5,1)$_{-1-2}$PP
***A**. ECCD replicate spectra of R239A at equilibria with DF(p5,1)$_{+1-1}$PP and DF(p5,1)$_{-1-2}$PP in the presence of Ca^{2+} ions and subsequently rescued by EDTA. Spectra were normalised and readings at 326 nm were used to monitor the changes in DNA conformation. **B**. ECCD replicate spectra of K244A at equilibria with DF(p5,1)$_{+1-1}$PP and DF(p5,1)$_{-1-2}$PP in the presence of Ca^{2+} ions and subsequently rescued by EDTA. Spectra were normalised and readings at 326 nm were used to monitor the changes in DNA conformation. **C**. ECCD replicate spectra of R245A at equilibria with DF(p5,1)$_{+1-1}$PP and DF(p5,1)$_{-1-2}$PP in the presence of Ca^{2+} ions and subsequently rescued by EDTA. Spectra were normalised and readings at 326 nm were used to monitor the changes in DNA conformation. **D**. Bar chart comparison of the substrate DF(p5,1)$_{+1-1}$PP against the WT and H2TH mutants. **E**. Bar chart comparison of the substrate DF(p5,1)$_{-1-2}$PP against the WT and H2TH mutants.*

4.6 Summary

In summary, the aim of this chapter was to uncover the kinetic parameters of the H2TH mutants R239A, K244A, and R245A accompanied by their effect on DNA distortion at the ss/ds junction. After performing kinetic experiments under multiple and single turnover conditions. Mutation of basic residues of the H2TH domain produces limited changes in the catalytic turnover of substrate. The largest impact was shown by R245A which increased the K_M compared to the WT protein (20 nM) and suggests that out of the three H2TH mutants, R245A observes the highest reduction in affinity for DF(p5,1). However, the H2TH mutants demonstrated only a marginal decrease in nuclease activity, despite being conserved amongst multiple FEN superfamily members. This was further shown through single-turnover experiments of the H2TH mutants whereby enzyme product release remained rate-limiting throughout the reaction. Despite individual H2TH motif mutations having minimal effect on FEN1 reactions, one possibility is that while individual residues alone are not vital to FEN1 catalysis they act together to achieve optimal substrate processing. From this chapter, the H2TH mutants show a modest decrease in DNA conformational change accompanied by their small alterations in multiple turnover, and single turnover kinetic rate values. However, the scale in which these mutations alter catalytic function is near null compared to other alanine mutations of FEN1. For example, a key residue linked to conformational change is L53 located in the 3'-flap binding region of FEN1 whose mutation to alanine results in a 1000-fold decrease in nuclease activity and DNA conformational change at the ss/dsDNA junction. This compared to the findings of both kinetics and ECCD experiments with the H2TH mutants further show that the H2TH mutants have nowhere near this level of impact on FEN1 catalytic function of DNA conformational change towards the active site.

Chapter 5 Expression and purification of FAN1

5.1 Introduction

The production of target proteins is vital for kinetic and biophysical investigation of their properties. Poor expression of the target protein, in this case FAN1, would necessitate repetitive and time-consuming protein production. In addition to this, impure protein samples may lead to incorrect results of downstream assays. This is a major concern in kinetic studies where protein and substrate concentrations must be accurately measured. Therefore, to allow characterisation of FAN1, optimal expression conditions and purification strategies were identified.

For kinetic and biophysical investigation, proteins must be soluble in solution. To help solubilise insoluble proteins, fusion proteins such as NusA, MBP and SUMO can be attached to the N-terminus of the target protein during recombinant protein production. Both NusA (55 kDa) and MBP (42 kDa) increase target protein solubility in a similar manner, despite their difference in structure and physio-chemical properties. In addition to solubility enhancement, NusA interacts with various DNA maintenance proteins whereas MBP forms complexes with proteins involved in the maltose transport system. When NusA and MBP are utilised in recombinant protein expression, they both perform protein-protein interactions which reduces the level of self-association of the target protein resulting in reduced aggregation. Despite this, it is suggested that NusA and MBP do not play a role in target protein folding [142], [143]. SUMO fusion tags (11 kDa) are also used extensively in recombinant protein production. It has been suggested that the SUMO fusion protein aids in the folding of the attached, target protein [144]. Before kinetic and biophysical assays can be undertaken, the fusion protein must first be removed by protease targeted hydrolysis. These hydrolysis events are dictated by specific amino acid sequences within the target protein. In the presence of NusA, Thrombin protease can be employed to hydrolyse between arginine and glycine residues whereas TEV protease hydrolyses between glutamine and serine residues within an MBP-containing target protein [145]. In the case of SUMO, the protease ULP1 hydrolyses the amide linkage at the diglycine motif [144]. Finally, embedded tags can also be employed to purify the target protein. For example, his tags (consisting of 6-8 tandem histidine residues) allow binding to Co^{2+} or Ni^{2+} charged columns (denoted as IMAC). His-tagged

fusion proteins via the N-terminus can allow separation of the target protein and the fusion protein via IMAC. Purification procedure permits binding of the fusion protein to the column whereas the target protein elutes into the flow-through due to the fusion protein possessing the his-tag following cleavage by proteases.

Multiple expression and purification strategies for the large-scale production of FAN1 have been reported. [102], [105], [146]. FAN1 was successfully expressed using the baculovirus expression system which favours proteins requiring post-translational modifications. The FAN1 gene was cloned into the pFB vector system containing an MBP fusion protein on the N-terminus and was subsequently purified by amylose resins and heparin chromatography [88], [147]. A human FAN1-green fluorescent protein construct (FAN1-GFP) has been produced in human embryonic kidney cells (HEK293) and purified using GFP binding resins. The incorporation of insect and mammalian systems would require expensive, specialist equipment accompanied by extensive training before recombinant protein expression can be undertaken. Therefore, it would be favourable to focus on optimising bacterial systems to express the FAN1 protein. Furthermore, no post-translational modifications are required when experimenting with the truncated FAN1 protein lacking the UBZ domain. [146], [148]

The successful expression and purification of FAN1 has also been reported using bacterial systems (*E.coli*). One study purified FAN1 coupled with an N-terminal NusA fusion protein originating from the pET43a vector. The lysate was purified by DEAE anion exchange chromatography prior to removing NusA with thrombin Finally, the protein was purified using heparin chromatography [88]. Notably, the final volume and concentration of FAN1 was not reported for these expression and purification methods, making it difficult to compare the overall effectiveness of each strategy. Other reports have expressed and purified FAN1 prior to X-ray crystallographic studies using *E.coli*. One report suggests that N-terminal fusion proteins are not required (pET15b vector system), and purified FAN1 by Ni^{2+} charged IMAC columns, followed by on-column cleavage of the N-terminal his-tag by TEV protease before purification with ion exchange and size exclusion chromatography, resulting in a final FAN1 concentration of 20 mg/mL [102]. Whereas another study presents evidence that FAN1 with an N-terminal his-tag and MBP fusion protein could be successfully purified with amylose resin and size exclusion chromatography, bearing a final concentration of 5 mg/mL after removal of the MBP fusion protein by the SARS-CoV protease [105], [149]. A similarity amongst all the FAN1 production

strategies is the use of the TRIS-HCl buffer accompanied by NaCl (100-1000 mM NaCl) and glycerol for cryo-protection before long-term cold storage. After reviewing the literature, multiple strategies have been performed to produce FAN1 although it is unclear which fusion protein is the most effective to increase FAN1 solubility, or whether fusion proteins are required at all.

After successful expression and purification of FAN1, identifying optimum buffer conditions was subsequently explored. However, testing buffer and additive conditions through repetitive expression and purification protocols is time-consuming. Incorporating high-throughput screening strategies to test the change in thermal stability of the protein in its local environment is a more effective method to improve buffer conditions. Differential Scanning Fluorimetry (DSF) is a technique utilising fluorescent dyes that increase in fluorescence when in close contact to hydrophobic regions of macromolecules. DSF enables calculation of the half maximal melting temperature (T_M) of the target protein under varying buffer, pH, salt, and other additive conditions. By periodically increasing the temperature of the solution using a RT-PCR machine, the target protein will unfold (Figure 5.1A). During this process, the dye SYPRO-Orange (Figure 5.1B) can access the hydrophobic regions of the protein and the fluorescence will increase. Therefore, the larger the increase in T_M compared to existing conditions, the higher the thermal stability of the target protein within its environment [150], [151]. DSF and SYPRO-Orange were utilised to measure the change in FAN1 T_M in varying conditions. It was hypothesised that combining conditions that increase the T_M of FAN1 will enable higher concentration of FAN1 following purification due to an increase in stability, and in consequence, solubility.

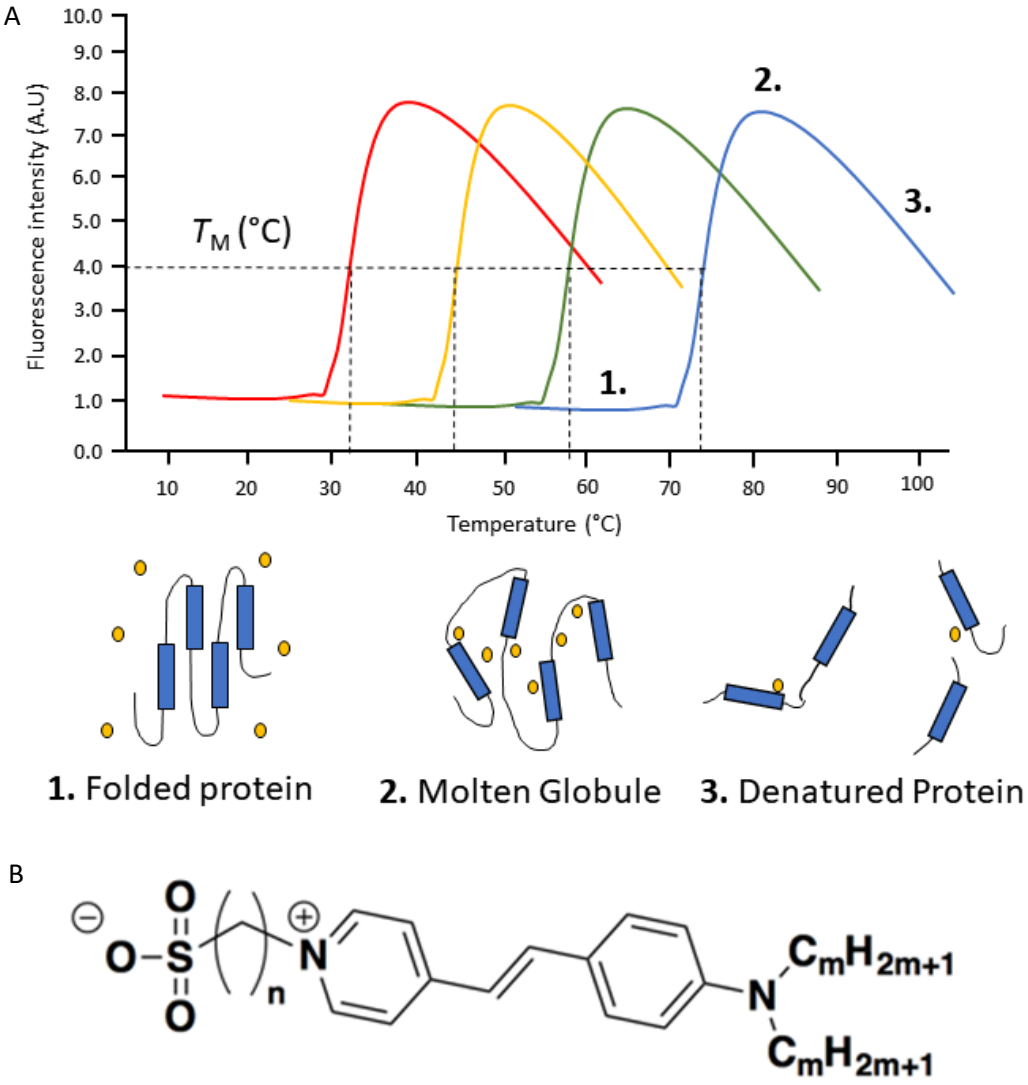


Figure 5.1 Differential scanning fluorimetry and SYPRO-Orange

***A**. Schematic of the DSF experiment where different theoretical buffer conditions are shown by the colours red, yellow, green, and blue. A numbered overlay shows the folded state of the protein as the temperature increases. The target protein is first folded prior to the increase in temperature (1.) and once the temperature increases, the protein unfolds enters a molten globule state whereby the tertiary structure of the protein is lost whilst retaining secondary structure. From this, the hydrophobic regions are exposed, resulting in an increase in fluorescence signal through interactions between the hydrophobic regions and the SYPPRO-Orange dye (2.). Finally, as the protein denatures at high temperatures, the fluorescence decreases again (3.).*
***B**. The structure of SYPRO-Orange dye is a modified image from Volkova et al., 2011* [152].

5.2 PCR amplification and plasmid preparation

5.2.1 Production of the FAN1-NusA construct

Molecular cloning, plasmid preparation and sequencing of the FAN1-NusA construct was performed by Dr Mark Thompson using the pET-His6-TEV-LIC cloning vector. The FAN1-NusA construct consists of a his-tag and a NusA fusion protein for IMAC purification and solubility enhancement respectively. The total linearised length of the pET-His6-NusA empty vector construct was 6486 bp. This, combined with the linearised FAN1 373-1017 gene insert of 1937 bp, results in a final linearised product of 8783 bp. (Figure 5.2A). The sequenced plasmid was bacterially transformed into Dh5α cells and mini-prepped. Agarose gel electrophoresis analysis observed a band at 5000 bp which was thought to be the FAN1-NusA construct in the supercoiled form (Figure 5.2B). The plasmid was stored until ready to be transformed into BL21 competent cells for recombinant protein expression.

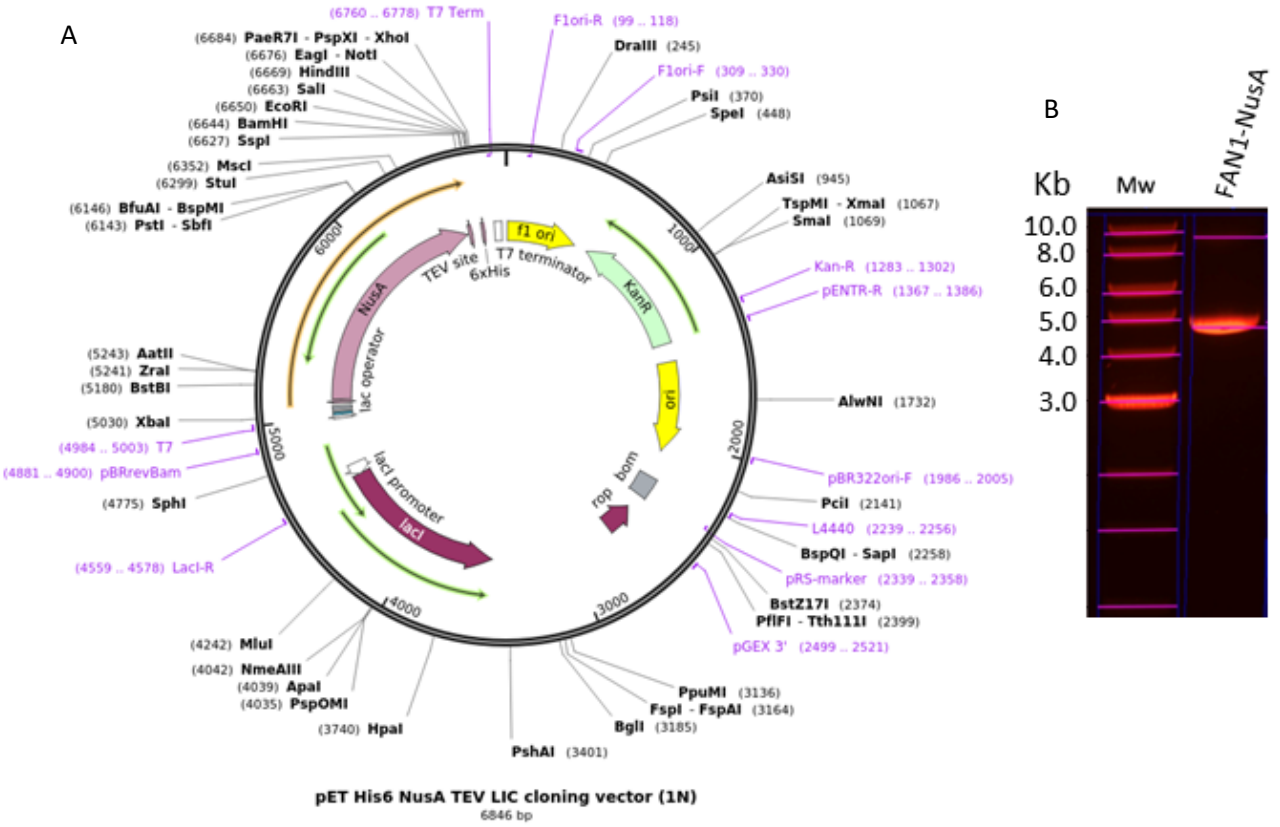


Figure 5.2 FAN1-NusA construct

***A**. Plasmid map of the pET-His6-NusA plasmid construct and available restriction sites. The plasmid map for this construct was taken from the Addgene website. **B**. Agarose gel electrophoresis of the FAN1-NusA construct following plasmid mini-preparation. The construct obtained a concentration of 138 ng/µL. Molecular markers are shown in kilobases.*

5.2.2 Production of the FAN1-SUMO and FAN1-MBP constructs

The FAN1-SUMO construct (FAN1 residues 373-1017) was produced using traditional cloning. The combined pE-SUMO-pro vector containing the his-tag, SUMO fusion tag, ULP1 cleavage site, and the FAN1 gene has a total linearised length 7673 bp (Figure 5.3A). The empty pESUMO-pro vector was first digested with Bsa1 restriction endonuclease. The digested vector was analysed by agarose gel electrophoresis before gel-extraction in preparation for ligation. pFU hotstart polymerase PCR accompanied by GC enhancers led to a FAN1 PCR product of ~2000 bp (Figure 5.3B). The gene insert was analysed by agarose gel-electrophoresis and gel extracted before Bsa1 digestion. PCR clean-up was performed on the digested FAN1 gene and ligated into the digested pESUMO-pro vector with T4 DNA ligase. The ligated FAN1-SUMO construct was transformed into bacterial DH5α cells prior to mini-prep amplification and sequencing (Figure 5.3C) (See sections 9.1.2 for FAN1-SUMO construct sequencing).

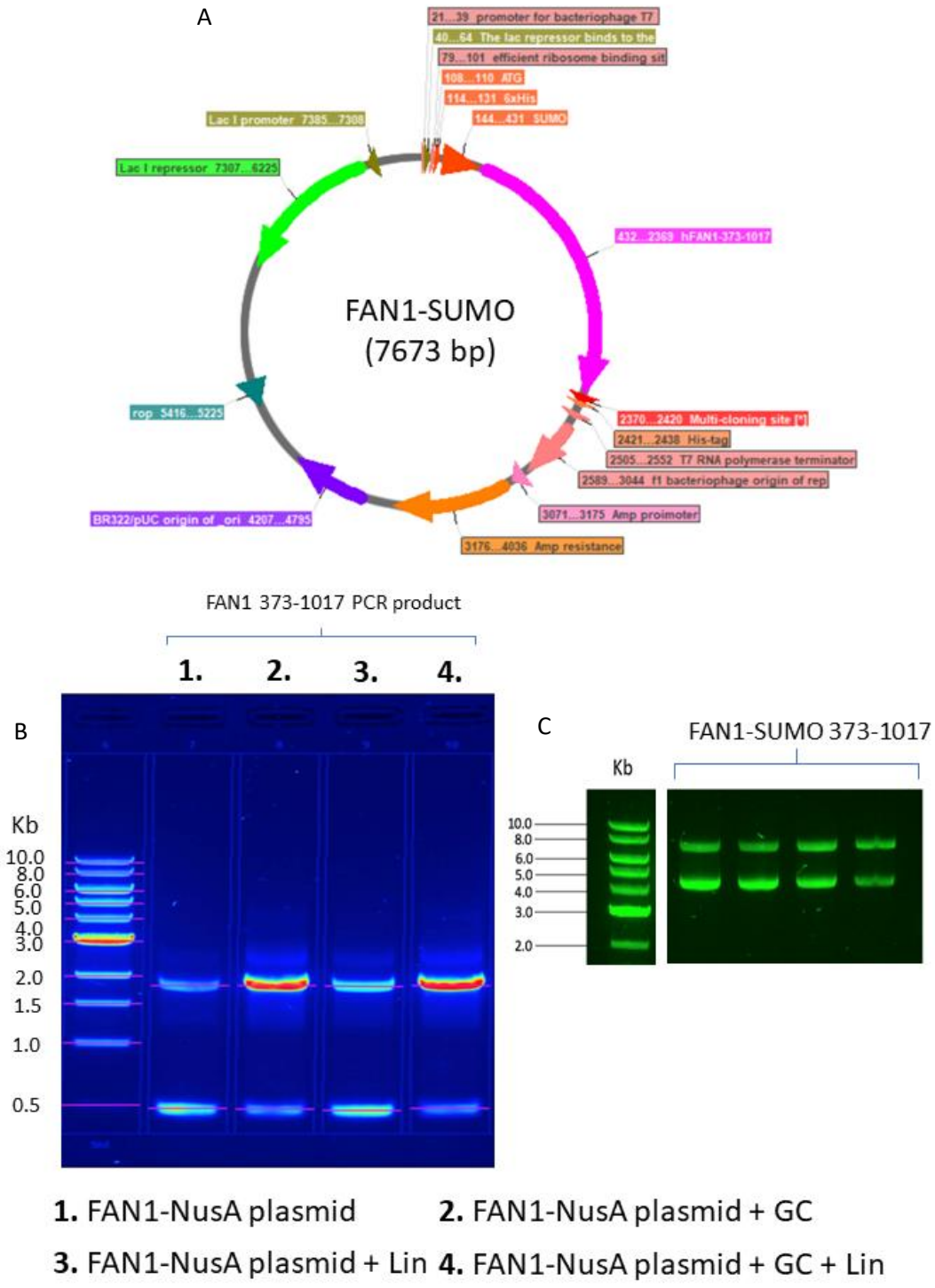


Figure 5.3 Traditional cloning of the FAN1-SUMO construct

***A**. Plasmid map of the FAN1-SUMO construct. **B**. pFU hotstart polymerase reaction of the FAN1 373-1017 gene insert. The incorporation of FAN1-NusA plasmid, linearised FAN1-NusA plasmid and GC enhancers produced more FAN1-373-1017 gene insert following PCR. Composition of the wells are shown below the agarose gel. **C**. Plasmid mini prep of multiple colonies following transformation of FAN1-SUMO into Dh5α cells with concentrations (from left to right) 185 ng/µL, 164 ng/µL, 170 ng/µL and 162 ng/µL. Molecular markers are shown in kilobases*

Ligation-independent cloning was used to produce the FAN1-MBP construct (FAN1 residues 364-1017). The linearised vector containing the his-tag, MBP and TEV cleavage site has a total length of 3138 bp. This accompanied by the linearised FAN1 gene of 1962 bp should equate to an 8427 bp linearised product (Figure 5.4A). The pET-HMT empty vector in the supercoiled morphology showed bands at 3000 bp (Figure 5.4B) and was linearised by Ssp1 digestion (Figure 5.4C). The Q5 polymerase PCR reaction produced the FAN1 gene insert. Agarose gel electrophoresis was performed on the digested pET-HMT vector and FAN1 gene insert prior to gel-extraction with the linearised pET-HMT vector presenting bands at ~8000 bp and the FAN1 gene insert at ~2000 bp (Figure 5.4D). The pET-HMT vector and FAN1 gene insert were treated with T4 DNA polymerase to produce 3'-overhangs. PCR clean up procedures were performed before combining the pET-HMT vector and FAN1 gene insert. The combined sample was transformed into DH5α cells and grown on LB agar plates with ampicillin. Colony PCR of colonies 1 and 2 showed bands which were proposed to be the ligated and supercoiled FAN1-MBP construct (Figure 5.4E). Colony 1 was plasmid mini-prepped and analysed by agarose gel-electrophoresis, showing a band just over 8000 bp and subsequently sent for sequencing (Figure 5.4F) (See sections 9.1.3 for FAN1-MBP construct sequencing).

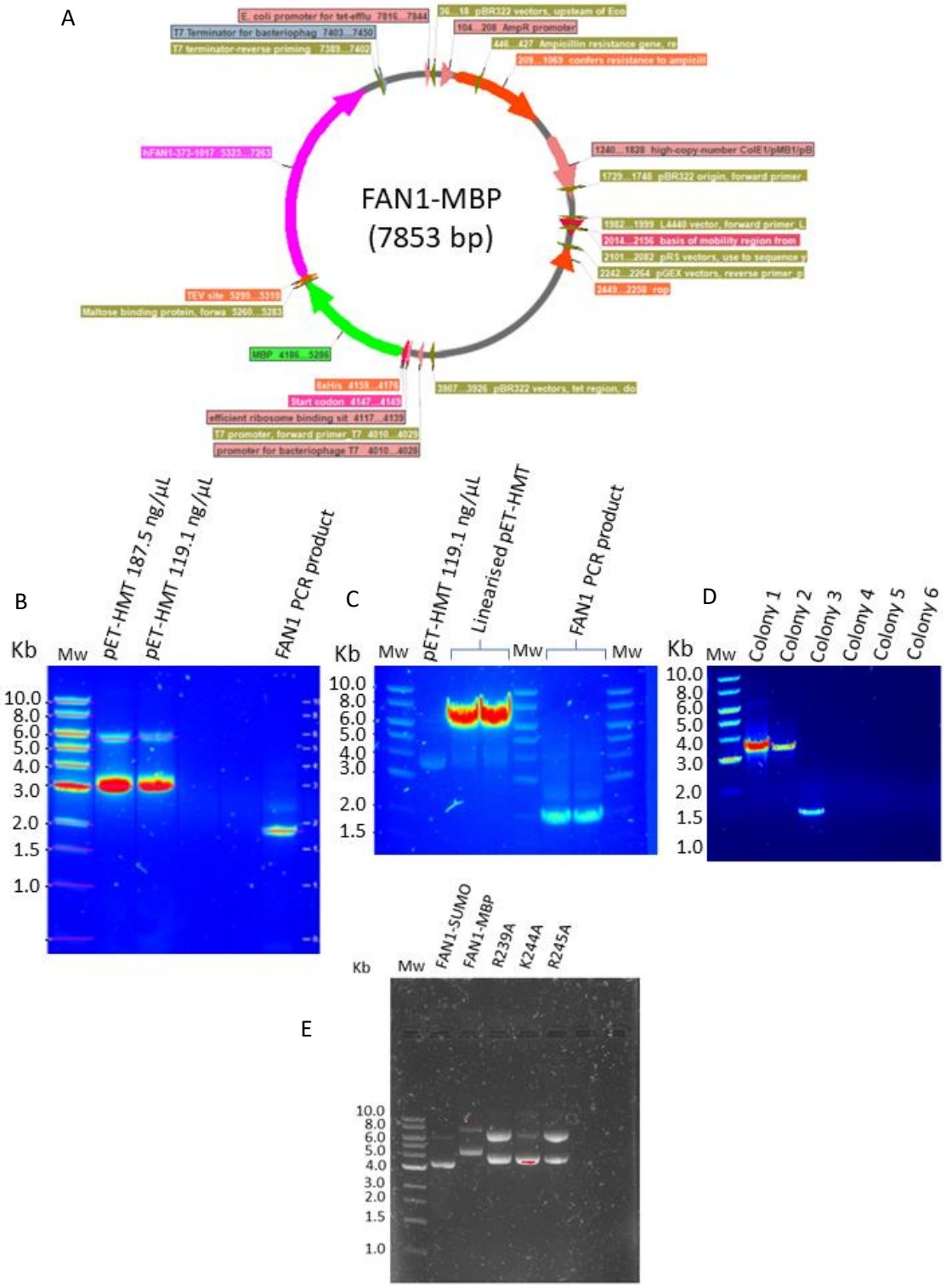


Figure 5.4 Ligation independent cloning of the FAN1-MBP construct

***A**. Plasmid map of the FAN1-NusA construct. **B**. PCR amplification of the linearised FAN1 gene insert at ~2000 bp. The gel also shows the pET-MBP empty vector in the circular (3000 bp) and linearised form (~6000 bp) prior to Ssp1 restriction endonuclease digestion. **C**. Agarose gel electrophoresis showing the pET-MBP empty vector, linearised pET-MBP empty vector and FAN1 PCR gene insert at ~3000, 6000, and ~2000 bp respectively. **D**. Colony PCR showing colonies 1 and 2 with the circularised form of the fully constructed FAN1-MBP construct prior to plasmid mini-prep and sequencing. **E**. Agarose gel electrophoresis of the FAN1-MBP construct at approximately 8000 bp. Molecular markers are shown in kilobases.*

5.3 FAN1 production

5.3.1 Small-scale expression and purification of NusA, SUMO and MBP FAN1 constructs.

After successful cloning and review of the plasmid maps, the theorised molecular weights of FAN1-NusA, FAN1-SUMO and FAN1-MBP were 130, 86 and 117 kDa respectively. Expression tests were performed and FAN1 was identified in the soluble, and insoluble fraction for all constructs (Figure 5.5A & C). Over-expression was observed in TY media with auto-induction for FAN1-NusA, however after lysis by both sonication and French press, FAN1-NusA remained almost entirely in the insoluble pellet fraction. Following expression tests, IPTG induction with probe sonication was chosen as the primary method for cell lysis to increase the yield of soluble protein. As described previously, all three constructs contained his-tags and therefore, small-scale IMAC proof of concept experiments were performed to ensure the expressed proteins bound to a Ni^{2+} IMAC column. Starter cultures (5 mL) were produced, and the soluble fraction was purified using bench-top 1 mL His-Spin trap columns. All three constructs bound to the Ni^{2+} column and eluted upon the addition of buffer containing 300 mM imidazole. FAN1-NusA and FAN1-MBP showed the highest band intensity compared to FAN1-SUMO (Figure 5.5E) and therefore, FAN1-SUMO was removed from the experimental design (Figure 5.5B, D & E).

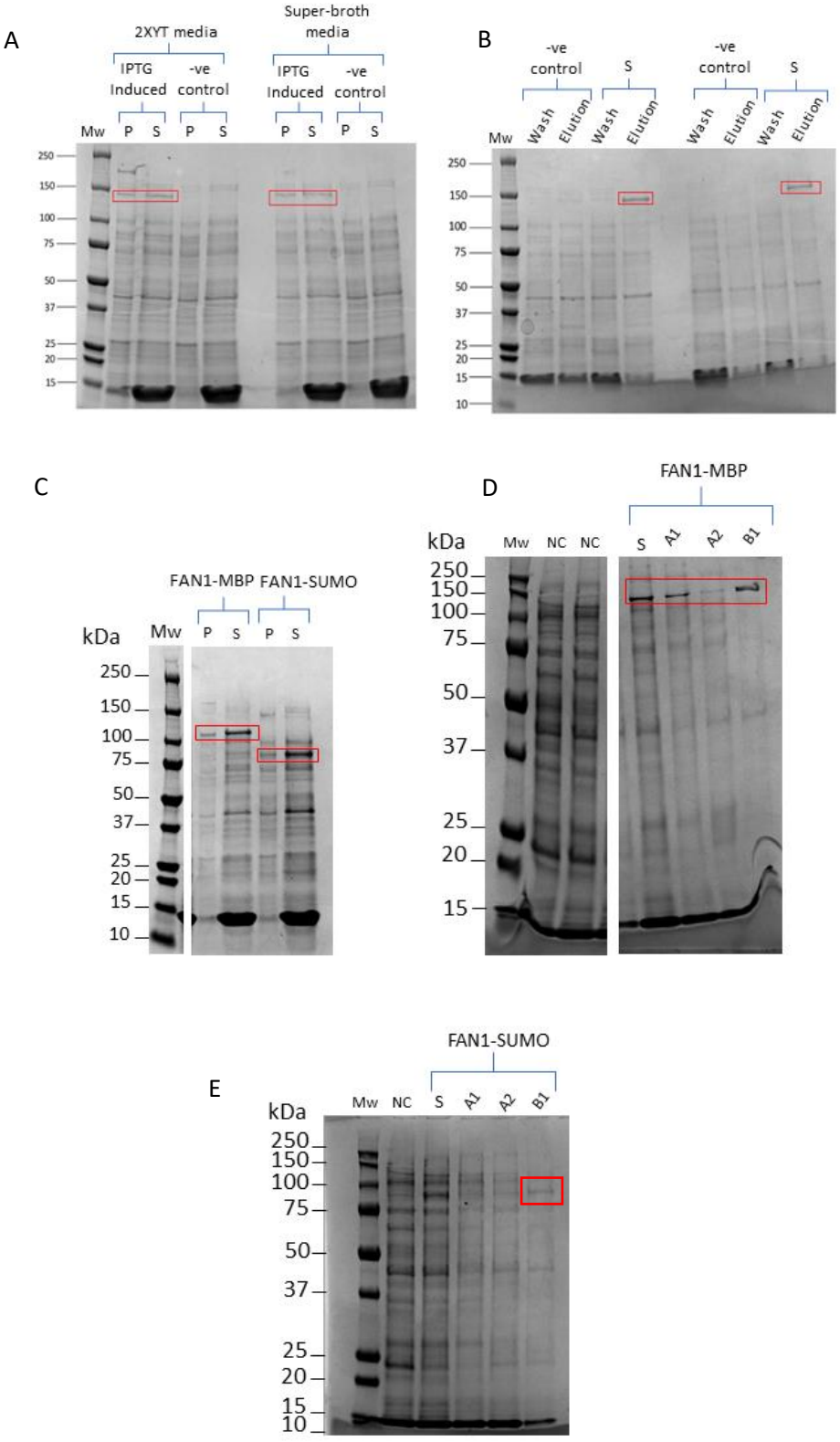

A
2XYT media
Super-broth media
IPTG Induced
-ve control
Mw
P
S
250
150
100
75
50
37
25
20
15
B
-ve control
S
Wash
Elution
10
C
FAN1-MBP
FAN1-SUMO
kDa
D
NC
A1
A2
B1
E

Figure 5.5 Expression test and small-scale purification of FAN1-SUMO and FAN1-MBP

***A**. Expression tests of FAN1-NusA from BL21 DE3 PlysS cells grown in 2XYT and superbroth. The Soluble fraction (S) and pellet (P) were analysed by 4-20 % SDS-PAGE with FAN1-NusA measuring at ~130 kDa and compared to negative controls. **B**. Soluble fractions of FAN1-NusA purified using His-spin trap Ni^{2+}purification of FAN1-NusA. Soluble fractions (S) were applied to the column and washed with buffer containing 40 mM imidazole before being eluted with a buffer containing 500 mM imidazole. 4-20 % SDS-PAGE identified FAN1-NusA at ~130 kDa and compared to negative controls. **C**. Expression test of FAN1-SUMO and FAN1-MBP from BL21 DE3 PlysS cells grown in 2XYT. Cells were lysed using lysozyme and the soluble fraction (S), and insoluble fraction (P) were analysed by 4-20 % SDS-PAGE with FAN1-SUMO and FAN1-MBP measuring at ~87 and ~117 kDa respectively. **D**. His-spin trap purification of FAN1-MBP. The soluble fraction (S) was loaded onto the column prior to being washed with buffers containing 5 mM imidazole (A1), 40 mM imidazole (A2) and eluted with 500 mM imidazole (B1) 4-20 % SDS-PAGE identified FAN1-MBP at ~117 kDa and compared to negative controls (NC). **E**. His-spin trap purification of FAN1-SUMO and follows the same procedure as part D. 4-20 % SDS-PAGE identified FAN1-SUMO at ~87 kDa and compared to negative controls (NC). The FAN1-NusA construct bands are highlighted by red rectangles (~130 kDa).*

5.3.2 Large-scale expression and purification of FAN1-NusA

As described in section 5.3.1, the FAN1-NusA and FAN1-MBP constructs were over-expressed in *E.coli* and purified using their N-terminal his-tags binding to Ni^{2+} his spin trap columns. The culture volume was increased to 500 mL and the soluble fraction was loaded onto a 5 mL Co^{2+} TALON column. Although cobalt columns have lower affinity for his-tag proteins, they also prevent non-specific binding of protein contaminants to the column, leading to fewer contaminants in the elution fraction. The Co^{2+} IMAC trace showed no peaks in the elution step (Figure 5.6A) and SDS-PAGE analysis confirmed FAN1-NusA remained in the flow-through with overloaded bands at ~130 kDa (FL) (B). The flow-through fractions were pooled and loaded onto a 5 mL Ni^{2+} histrap column displaying high affinity for his-tagged proteins (Figure 5.6C). The use of Co^{2+}-, and Ni^{2+}-charged resins allows for the removal of protein contaminants followed by high affinity binding of his-tagged proteins prior to elution. FAN1-NusA eluted both in the flow-through and elution stages after the addition of 300 mM Imidazole, possibly due to column overloading (Figure 5.6C) and was later confirmed by SDS-PAGE (Figure 5.6D). SDS-PAGE of the eluted sample also showed potential non-specific removal of the NusA fusion protein from FAN1 through overloaded bands at 75 kDa. Cloudy white precipitate was visible in the eluted sample before dialysis into a buffer containing approximately ~100-150 mL sodium chloride in preparation for heparin purification. The remaining NusA tags still attached to FAN1 were removed through the addition of thrombin protease during dialysis (Figure 5.6E).

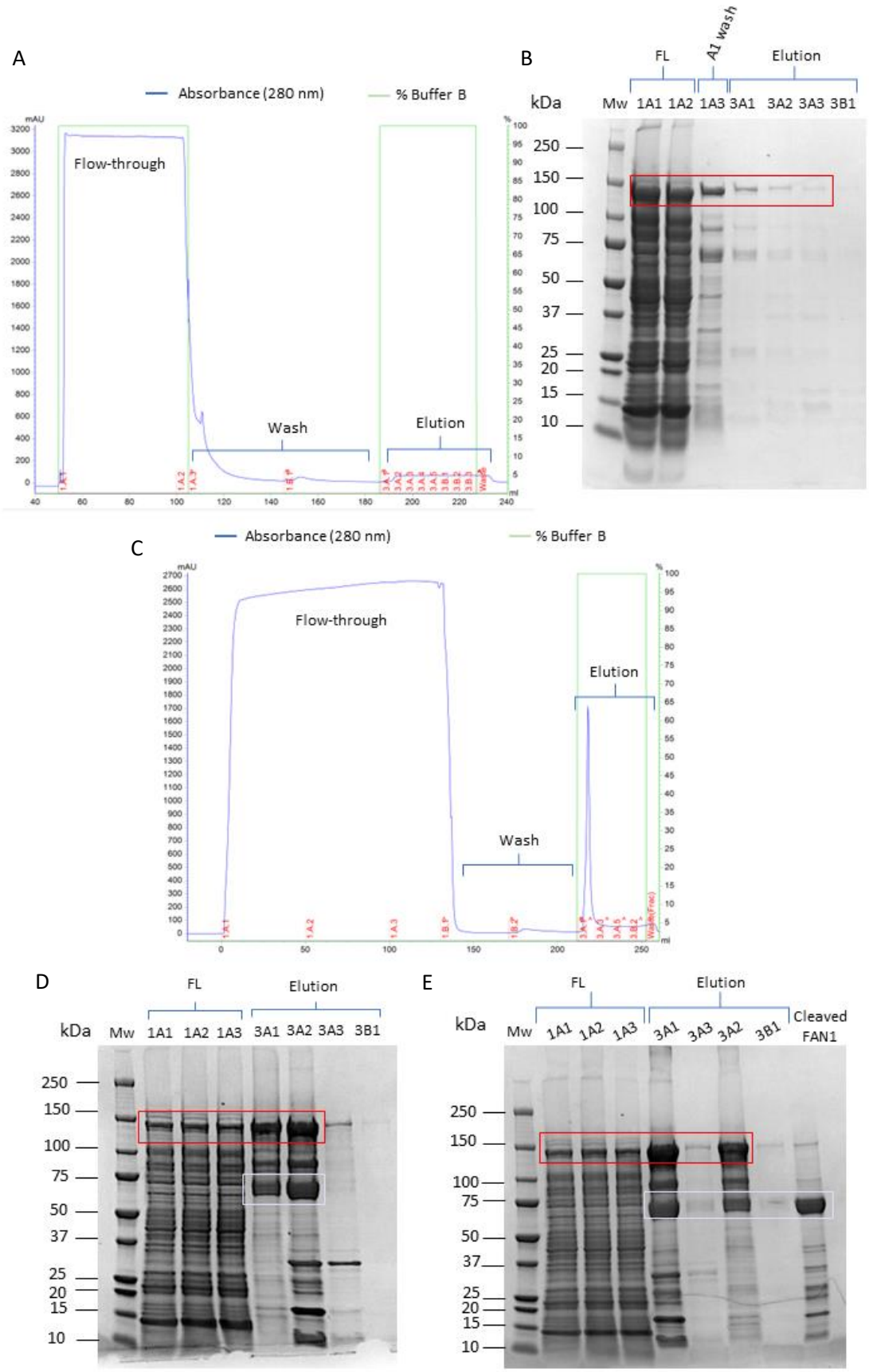

A
Absorbance (280 nm)
% Buffer B
mAU
Flow-through
Wash
Elution
ml
B
FL
A1 wash
Elution
kDa
Mw 1A1 1A2 1A3 3A1 3A2 3A3 3B1
250
150
100
75
50
37
25
20
15
10
C
Absorbance (280 nm)
% Buffer B
mAU
Flow-through
Wash
Elution
ml
D
FL
Elution
kDa
Mw 1A1 1A2 1A3 3A1 3A2 3A3 3B1
250
150
100
75
50
37
25
20
15
10
E
FL
Elution
kDa
Mw 1A1 1A2 1A3 3A1 3A3 3A2 3B1 Cleaved FAN1
250
150
100
75
50
37
25
20
15
10

Figure 5.6 Large-scale purification of FAN1-NusA

A. *Co^{2+} TALON IMAC of a 500 mL culture expressing FAN1-NusA. Unbound proteins remained in the flow-through. The column was washed with buffer containing 5mM, and 20 mM imidazole prior to isocratic eluting with buffer containing 300 mM imidazole.* ***B****. 4-20 % SDS-PAGE of fractions following 5 mL Co^{2+} TALON IMAC of FAN1-NusA. FAN-NusA was found in the flow-through (FL) and A1 wash (5 mM imidazole) with bands at ~130 kDa using 4-20 % SDS-PAGE.* ***C****. Ni^{2+} histrap purification of FAN1-NusA using the buffer compositions as part A.* ***D****. 4-20 % SDS-PAGE of the flow-through (FL), and elution with FAN1-NusA being identified at 130 kDa.* ***E****. 4-20 % SDS-PAGE of Ni^{2+} histrap purification fractions to cleaved FAN1 following overnight dialysis with thrombin. The FAN-NusA construct bands are highlighted by red rectangles (~130 kDa) and the cleaved FAN1 protein is highlighted by light-blue rectangles (~75 kDa).*

Thrombin treated FAN1 samples (referred to as FAN1) were injected onto a Heparin column (Figure 5.7A). The flow-through and elution samples were analysed by SDS-PAGE (Figure 5.7B). An aliquot of the precipitated sample from the Ni^{2+} IMAC elution was also saved and analysed on the same SDS-PAGE gel. From this, higher concentrations of non-specifically cleaved FAN1 were found in the cloudy white precipitate following nickel IMAC purification compared to the heparin elution fractions, shown by the lack of overloaded bands at 75 kDa. Due to precipitation occurring before the fusion protein was removed, it was decided to remove the FAN1-NusA construct from future expression and purification strategies and to progress with large-scale purification of the FAN1-MBP construct.

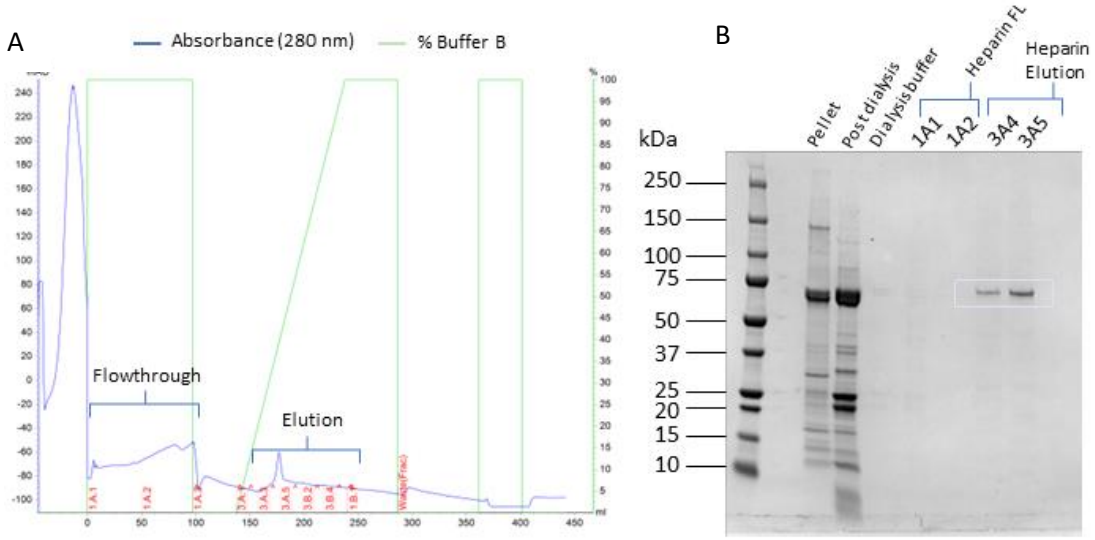


Figure 5.7 Dialysis, NusA cleavage, and Heparin purification of FAN1-NusA

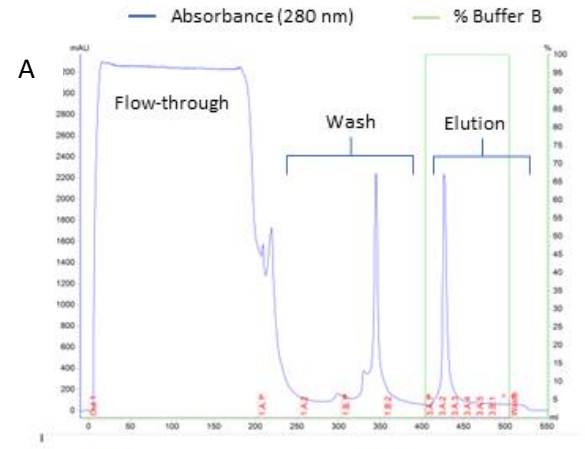
A*. Heparin affinity purification of Cleaved FAN1. The cleaved FAN1 sample was injected onto the 20 mL Heparin column. A gradient of 0.1-1.0 M NaCl was used to elute bound proteins from the column.* ***B****. 4-20 % SDS-PAGE of the cloudy white precipitate, FAN1 following dialysis and Heparin purification. No FAN1 was found in the flow-through (FL) and only faint bands at 75 kDa of FAN1 was found in the elution fractions. The majority of FAN1 was observed in the precipitated sample. The cleaved FAN1 protein is highlighted by light-blue rectangles (~75 kDa).*

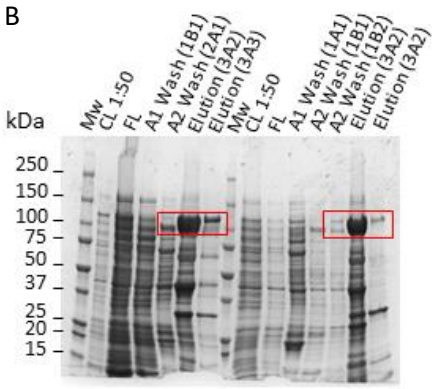
5.3.3 Large-scale expression and purification of the FAN1-MBP construct

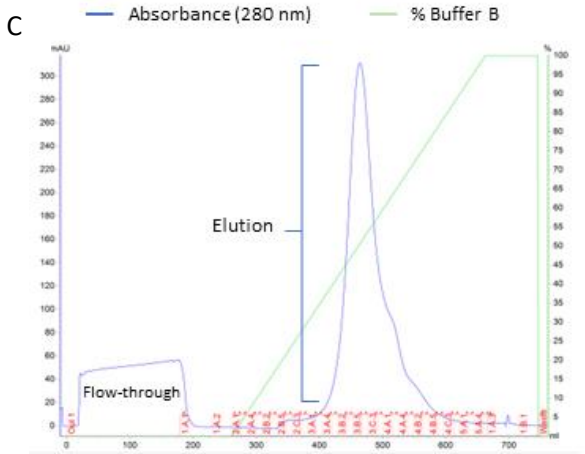
The FAN1-MBP construct was purified to homogeneity after lengthy exploration of multiple purification strategies. Purification methods including IMAC, and heparin chromatography were able to separate FAN1-MBP from impurities, although trace levels of FAN1 did remain in the flow-through fractions prior to removal of the MBP fusion protein by TEV protease [153]. It was believed that FAN1-MBP could be separated from other contaminants by using hydrophobicity-based purification systems such as hydrophobic interaction chromatography (HIC) with buffers containing ammonium sulphate [154]. However, the use of HIC led to cloudy precipitate forming and therefore resulted in more difficulty identifying samples via SDS-PAGE. Following Ni^{2+} IMAC purification, anion exchange chromatography was incorporated to remove DNA contamination and smaller protein contaminants from the sample. This resulted in the FAN1-MBP remaining in the flow-through fraction, although it was found that anion exchange was unable to remove protein contaminants between 10-100 kDa. Desalting columns were preferred for buffer exchange over membrane dialysis because less time was taken to perform the buffer exchange step and thus, reducing the possibility of precipitation. The time required to concentrate and buffer exchange protein samples was reduced from 1-2 days to 30 minutes using the HiPrep 2610 desalting column. Finally, size exclusion chromatography did result in sharp peaks which indicated isolated protein; however, it severely reduced the yield. Therefore, this was only employed if necessary to remove any final remaining contaminants.

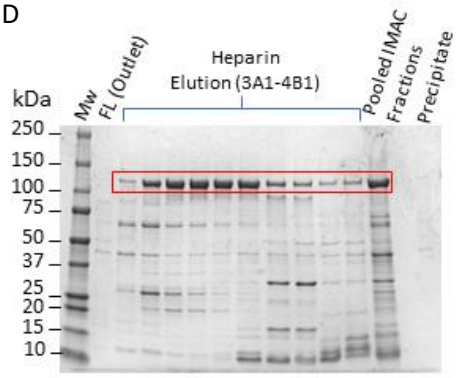
Following initial experiments, 5 litres of cells were grown. The cells were harvested, sonicated, and loaded onto a 20 mL Ni^{2+} IMAC FF nickel column. The chromatogram trace showed peaks in both the 40 mM imidazole wash stage and the elution stage with 300 mM imidazole (Figure 5.8A). SDS-PAGE analysis showed FAN1-MBP remaining predominately in the elution whereas other contaminants remained in the flow-through, and wash stages (Figure 5.8B). Fractions containing FAN1-MBP were slowly diluted 4-fold to reduce the NaCl concentration before being loaded onto a heparin affinity column with a gradient of 0-1.2 M NaCl and a single large peak was observed in the elution step (Figure 5.8C). White precipitate was observed in the elution fractions however SDS-PAGE analysis showed no presence of FAN1-MBP in the resuspended precipitate. SDS-PAGE analysis also revealed FAN1-MBP exclusively in the elution with distinct bands at ~117 kDa but other impurities also remained in the elution with bands between 10-100 kDa (Figure 5.8D). FAN1-

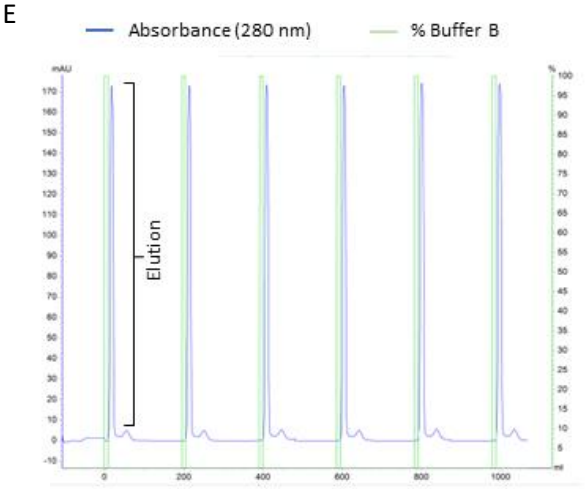
MBP was buffer exchanged by 6 consecutive 2610 desalting purifications (Figure 5.8E) and SDS-PAGE showed FAN1-MBP at ~117 kDa. Following TEV protease treatment, 75 kDa bands were observed followed by the removed MBP/his tag at ~42 kDa (Figure 5.8F). FAN1 was loaded onto the 20 mL Ni^{2+} IMAC FF nickel column to remove the MBP/his-tag. Peaks were observed both in the 40 mM imidazole wash, and 300 mM imidazole elution (Figure 5.8G). FAN1 was expected to elute into the flow-through fraction, however it eluted in the 40 mM imidazole wash with no other visible contaminants with an isolated band at 75 kDa (Figure 5.8H). The 100 mL of sample was buffer exchanged into 2× storage buffer and concentrated using a vivaspin-20 column to a final concentration of 2 µM in 10 mL following 2-fold dilution with 50 % Glycerol.

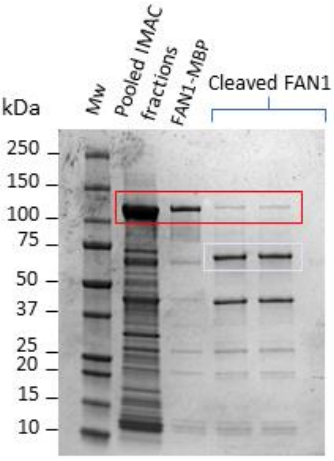

A
Absorbance (280 nm)
% Buffer B
Flow-through
Wash
Elution

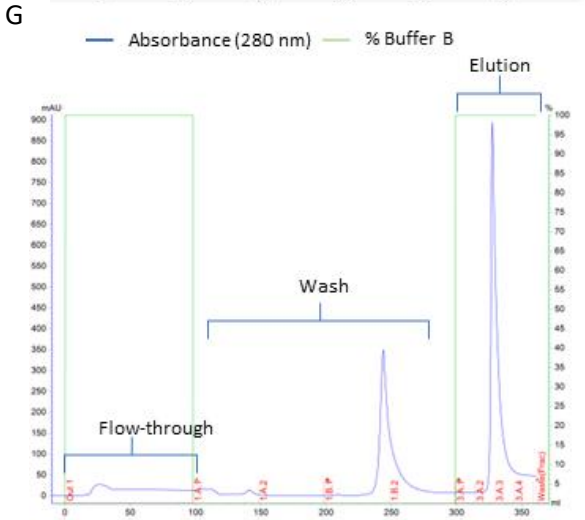

B
kDa
Mw
CL 1:50
FL
A1 Wash (1B1)
A2 Wash (2A1)
Elution (3A2)
Elution (3A3)
Mw
CL 1:50
FL
A1 Wash (1A1)
A2 Wash (1B1)
A2 Wash (1B2)
Elution (3A2)
Elution (3A2)
250
150
100
75
50
37
25
20
15


C
Absorbance (280 nm)
% Buffer B
Elution
Flow-through


D
kDa
Mw
FL (Outlet)
Heparin
Elution (3A1-4B1)
Pooled IMAC
Fractions
Precipitate
250
150
100
75
50
37
25
20
15
10


E
Absorbance (280 nm)
% Buffer B
Elution


F
kDa
Mw
Pooled IMAC fractions
FAN1-MBP
Cleaved FAN1
250
150
100
75
50
37
25
20
15
10


G
Absorbance (280 nm)
% Buffer B
Elution
Wash
Flow-through

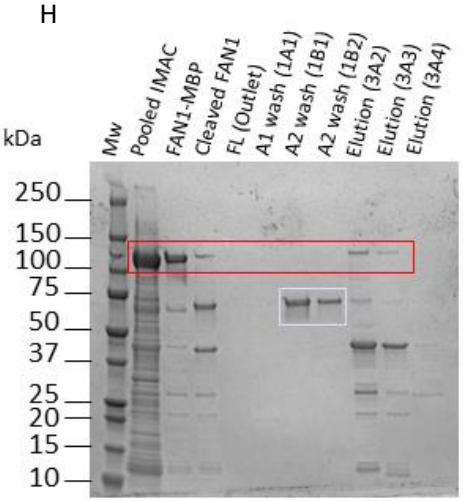

H
kDa
Mw
Pooled IMAC
FAN1-MBP
Cleaved FAN1
FL (Outlet)
A1 wash (1A1)
A2 wash (1B1)
A2 wash (1B2)
Elution (3A2)
Elution (3A3)
Elution (3A4)
250
150
100
75
50
37
25
20
15
10

Figure 5.8 Purification of FAN1-MBP

***A**. Ni^{2+} IMAC FF purification of FAN1-MBP. The soluble fraction was loaded onto the column before being washed with buffers containing 5 mM imidazole (A1) and 40 mM imidazole (A2). Bound proteins were eluted from the column with buffer containing 300 mM imidazole. **B**. 4-20 % SDS-PAGE of fractions collected from the flow-through (FL), wash and elution steps of part A. the clarified lysate (CL) was diluted 1:50 to identify over-expressed FAN-MBP at 117 kDa. Due to a systems failure, a second Ni^{2+} IMAC purification was performed. Both purifications revealed FAN1-MBP through overloaded bands at 117 kDa. **C.** Heparin purification of the elution fractions from diluted 4-fold with buffer to reduce the salt concentration of the sample. Non-bound proteins remained in the flow-through (FL). A gradient of 0.1-1.2 mM NaCl was used to eluted bound proteins from the heparin column. **D**. 4-20 % SDS-PAGE analysis of the heparin purification showing FAN1-MBP residing in the gradient elution through with bands at 117 kDa and no FAN1-MBP observed in the FL or precipitate. **E**. 2610 buffer exchange purification of FAN1-MBP into TEV buffer. The sample was buffer exchanged over 6 purifications shown by a blue line. **F**. TEV protease cleavage of the MBP (37 kDa) fusion protein from FAN1-MBP (117 kDa). **G**. Ni^{2+} IMACC FF purification of the FAN1. The same buffers were used as part A. Cleaved FAN1 eluted in the wash and the amongst other impurities removed during the elution stage. **H**. 4-20 % SDS-PAGE of collected fractions from the second Ni^{2+} IMAC FF purification of the cleaved FAN1. The FAN-MBP construct bands are highlighted by red rectangles (~117 kDa) and the cleaved FAN1 protein is highlighted by light-blue rectangles (~75 kDa).*

In summary, after multiple purification attempts, the FAN1-MBP construct was successfully purified using a combination of Ni^{2+} IMAC, desalting columns and heparin chromatography. In-house purified TEV protease was used to remove the MBP fusion tag from the FAN1 protein. Ni^{2+} IMAC successfully removed the MBP/his-tag from the sample following overnight TEV protease treatment. The yield of FAN1 was lower than expected from this purification. This likely occurred during concentration of the final protein sample using vivaspin-20 columns. Prolonged centrifugation without properly inverting the vivaspin-20 concentrator may have led to FAN1 irreversibly precipitating onto the membrane. Nonetheless, 10 mL of 2 µM of FAN1 was sufficient to perform initial nuclease activity assays to assess whether FAN1 was active.

5.4 DSF screening

The findings presented in section 5.3.2 describes the successful expression and purification of FAN1-MBP after exploring multiple purification strategies albeit low yeild. From this, DSF screening could now be performed with the aim of identifying buffer and additive conditions that increase the thermal stability of FAN1 in solution, this in turn increases the likelihood of FAN1 remaining correctly folded and soluble during kinetic and biophysical assays.

5.4.1 DSF screening to identify buffer and additive conditions.

DSF experiments were performed by mixing 0.1-0.2 µM FAN1 with 1 × SYPRO-Orange dye. These values were obtained from a preliminary screen testing FAN1 between 0.0-1.0 µM and dye concentration of 1-6 × (Figure 5.9A). After screening multiple conditions, FAN1 exhibited higher thermal stability at pH ranges 6.5-7.0 with no significant change to the T_M in the presence of 250 mM NaCl (Figure 5.9B). Individual buffers were then tested against FAN1 in the addition of high ionic strength. These screens showed that the highest T_M value was recorded using PIPES-NaOH pH 7.0 and the T_M increased further upon the addition of 250 mM NaCl. The T_M decreased at low and high pH, regardless of whether 250 mM NaCl was present (Figure 5.9C). Moreover, FAN1 in the presence of PIPES-NaOH pH 7.0 displayed the most uniform melt curve (Figure 5.1A) compared to other conditions where biphasic melting occurred or melt curves were not achieved entirely. The PIPES-NaOH pH 7.0 in the presence of NaCl demonstrated a higher T_M than Tris-HCl pH 7.5. However, it cannot be directly concluded whether the change in buffer, or the change in pH is the reason for the increased thermal stability of FAN1. Focus screens of PIPES-NaOH over the pH range 6.4-7.4 showed a decrease in T_M as the pH decreased. No significant increase in T_M was found whether FAN1 was in the presence of $MgCl_2$ or KCl (Figure 5.9D) compared to PIPES-NaOH in the presence of NaCl in SPG and TBG buffers (Figure 5.9B). Moreover, FAN1 showed an increase in thermal stability upon the addition of the divalent metal Mg^{2+} ions compared to mono-valent metals or a combination of both (Figure 5.9D). The increase in upon the addition of Mg^{2+} ions may stabilise the active site of FAN1 prior to substrate catalysis.

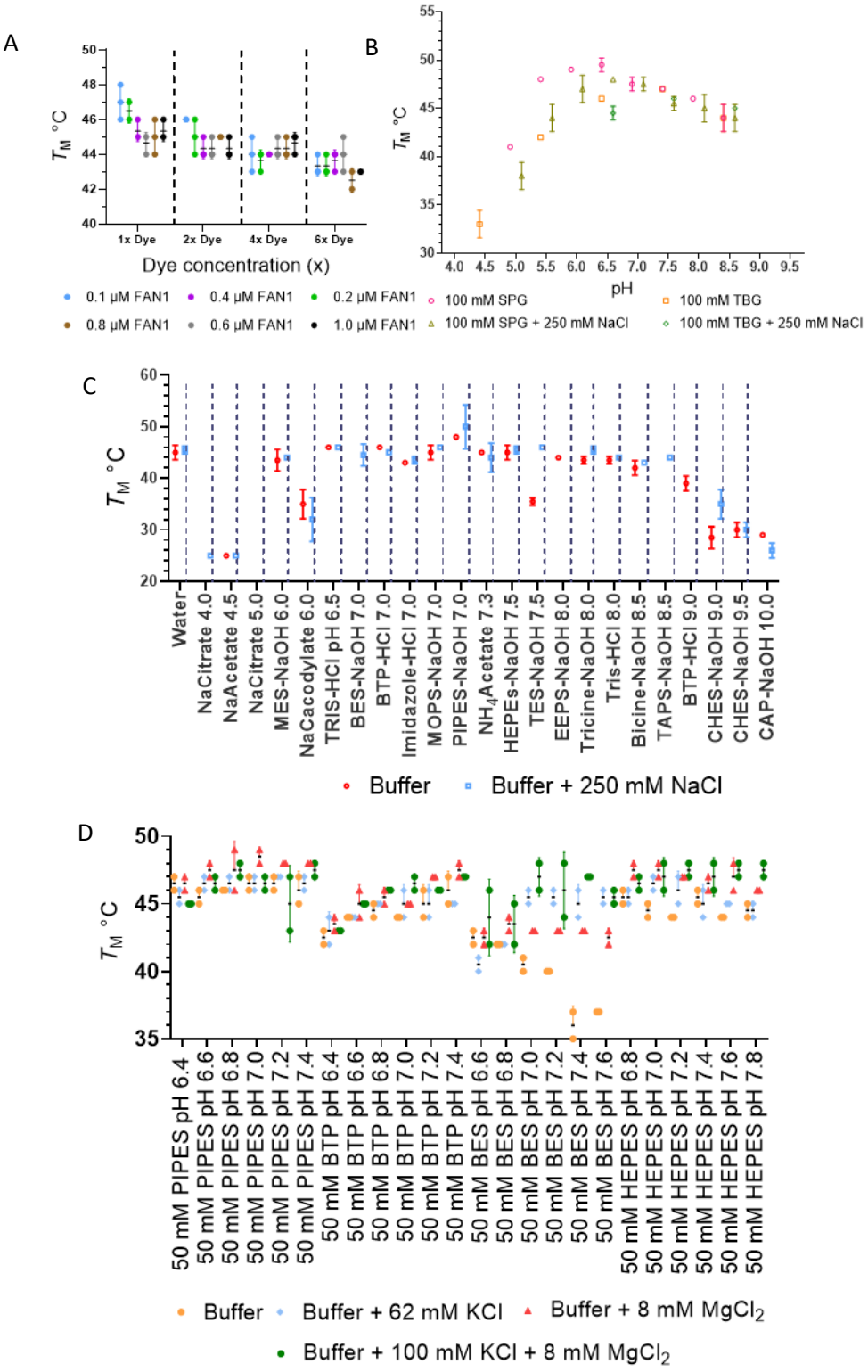


Figure 5.9 Broad-range buffer, individual buffers, and mono/divalent metal DSF screens

***A**. FAN1 and SYPRO-Orange dye concentration optimisation screen performed in triplicate from 0.0-1 µM FAN1 and 1-6 × SYPRO-Orange dye. **B**. DSF screen and pH titration curve of SPG AND TBG buffers with FAN1. **C**. DSF screen of widely available buffers ranging from pH 4.0-10.0 C. DSF focus screen of PIPES, BTP, BES and HEPES at pH ranges 6.4-7.8 in the presence of KCl and $MgCl_2$. **D**. DSF screen testing variable mono and divalent salt conditions using PIPES-NaOH pH 7.0 as the buffering system for FAN1. Mono and dibasic phosphate salt screens controlled for ionic strength with sodium phosphate, potassium phosphate, sodium chloride and potassium chloride at pH ranges 6.4-7.2.*

Interestingly, sodium phosphate mono-basic salts in PIPES-NaOH pH 7.0 resulted in a comparatively higher T_M value (Figure 5.10A) compared to other salt conditions. From this, it was asked whether a sodium phosphate buffer would have similar effect on the T_M of FAN1 compared to PIPES-NaOH pH 7.0. DSF screens controlled for ionic strength showed that phosphate buffers regardless of salt additive and showed a T_M of ~49 °C (Figure 5.10B). Due to PIPES being an expensive buffer compared to sodium phosphate, a sodium phosphate mono basic buffer (pH 7.0) was used to express and purify FAN1 before storing FAN1 in PIPES-NaOH pH 7.0. DSF screens of FAN1 with phosphate buffers controlled for ionic strength presented a T_M value of 49 °C, independent of the type of phosphate buffer or ionic strength controlling salt.

Finally, a RUBIC additive screen and Durham osmolyte screen was used to test several additive conditions. From these experiments, a dramatic change in the T_M of FAN1 was not seen with any of the additives. The presence of amino acids and imidazole did drastically reduce the T_M however these conditions are not pH adjusted and therefore led to a drastic change in pH of the PIPES buffer, resulting in decreased FAN1 stability. From these findings, the additive glycerol was added into the buffer system at 5 % v/v during protein production methods. Furthermore, $CaCl_2$ was also added to the PIPES storage buffer to ensure the FAN1 protein was saturated in non-catalytic metal conditions.

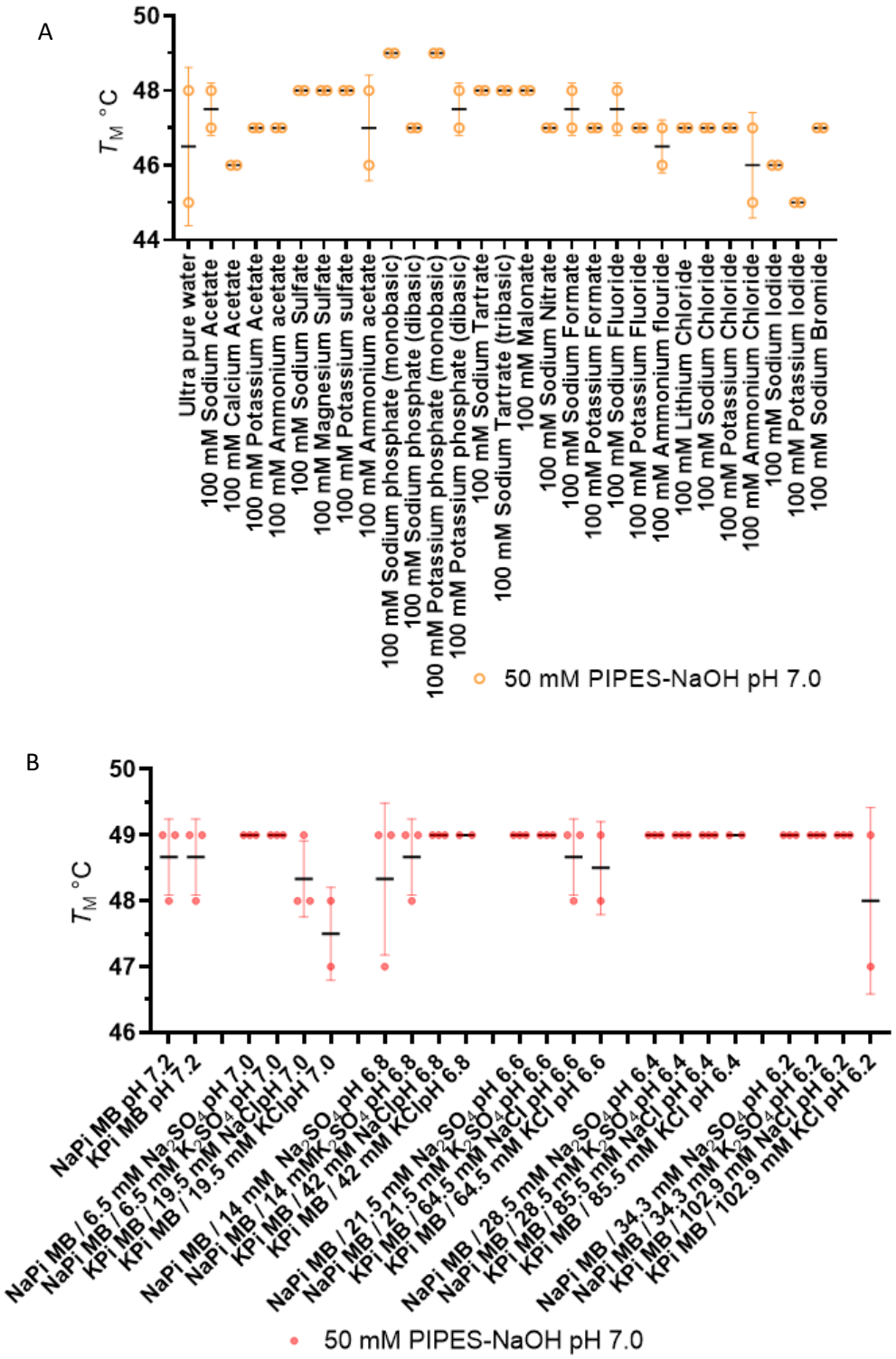


Figure 5.10 DSF screen of other salt additives and phosphate buffers controlled for ionic strength

***A**. Mono, and divalent salt screen of FAN1 in PIPES-NaOH pH 7.0. **B**. Sodium phosphate mono basic DSF screen of FAN1 controlled for ionic strength with sodium sulfate, potassium sulfate, sodium chloride and potassium chloride at pH ranged 6.2-7.2.*

In summary, DSF screening allowed high-throughput screening of buffer and additive conditions. Without DSF, screening individual conditions would prove time consuming and repetitive production of the FAN1-MBP would be required. DSF screening showed that PIPES-NaOH pH 7.0 and sodium phosphate buffers (pH 7.0) accompanied by NaCl increase the T_M of FAN1 and from this, a new strategy was devised where FAN1 was purified in phosphate buffer and stored in PIPES-NaOH. This is due to phosphate buffers being freely available, cheaper, and relatively simple to produce. Buffer exchanging into PIPES buffer was also preferred as phosphate buffers may form complexes with metal ions during the assay. Furthermore, DSF screening also showed that the change in FAN1 T_M was predominantly due to ionic-strength rather than specific mono-valent or divalent salt conditions.

5.5 Revised purification strategy for FAN1-MBP with improved buffer conditions

As previously described, DSF screening identified new buffer components that increase the T_M of FAN1 in the form sodium phosphate monobasic pH 7.0 for purification and PIPES-NaOH pH 7.0 for protein storage. For this final purification method, amylose resin was incorporated into the purification strategy due to the MBP fusion protein's ability to bind amylose. This purification strategy was originally believed to provide limited success due to the low binding capacity of amylose resin, especially for large-scale expression of MBP tagged protein. Furthermore, amylose resin is an expensive purification strategy and therefore, it was opted to remain optimising IMAC and ion exchange purification steps rather than incorporate an entirely new purification into the protocol. Despite the potential disadvantages of amylose resin, it was believed that the MBP tag could be used to purify the FAN1-MBP construct whilst simultaneously keeping FAN1 soluble in solution during purification. Moreover, on-column TEV protease cleavage could remove the MBP fusion tag from FAN1 however stay bound to the resin whilst FAN1 is eluted and thus, reduce the time taken to purify the FAN1 protein. From this, an experiment was devised to express FAN1-MBP and purify by amylose batch resin, followed by on-column TEV protease cleavage of the His-tag and Ni^{2+} IMAC to remove any residual MBP from the sample. If contaminants remained, heparin and size exclusion chromatography would be used to achieve FAN1 homogeneity.

5.5.1 Refined purification of FAN1-MBP

Expression of FAN1-MBP was repeated as described in Section 5.3.2. Following expression and purification via amylose resin, FAN1 was successfully purified, but with the loss of substantial amounts of FAN1 which did not bind to the column (Figure 5.11A). FAN1-MBP eluted into the first flow-through after column equilibration. It was suspected that FAN1-MBP eluted into other flow-through fractions however could not be confirmed due to overloaded bands. Individual bands at 75, and ~ 37 kDa after the addition of TEV protease indicated successful removal of the MBP fusion protein from FAN1 however 100 % MBP cleavage by TEV protease was not achieved, even after incubation for 12 hours at 4 °C. The addition of 10 mM Maltose eluted the MBP fusion tag from the amylose resin. However, the MBP fusion tag also eluted throughout the binding buffer wash stages and TEV protease cleavage steps, implying weak binding affinity between the MBP and the amylose resin. Overall, amylose resin separated a significant number of contaminants from the FAN1 protein and following removal of the MBP fusion tag, no signs of precipitation were observed.

Fractions containing bands at 75 kDa were believed to be the cleaved FAN1 protein and were pooled. The pooled fractions were subsequently loaded onto a Ni^{2+} IMAC column (Figure 5.11B). FAN1 remained bound to the column despite his-tag removal by TEV protease during amylose resin purification. A large peak was obtained in the 20 mM imidazole wash step and another in the 250 mM elution step. The cleaved FAN1 protein was identified as the peak in the wash step by SDS-PAGE with minimal to no contaminants in the sample. The band just above 37 kDa suggested that the cleaved MBP remained bound to the column until prior to elution with 250 mM Imidazole (Figure 5.11C). The wash fraction from the nickel IMAC purification was desalted using the 26/10 SEC column into 2 × storage buffer (Table 13) with a singular peak shown in the chromatogram trace (Figure 5.11D) with SDS-PAGE displaying overloaded FAN1 at 75 kDa (Figure 5.11E). Fractions were pooled and concentrated using vivaspin-20 concentrators to 5 mL and diluted with 50 % Glycerol to a final concentration of 18 µM (Figure 5.11F).

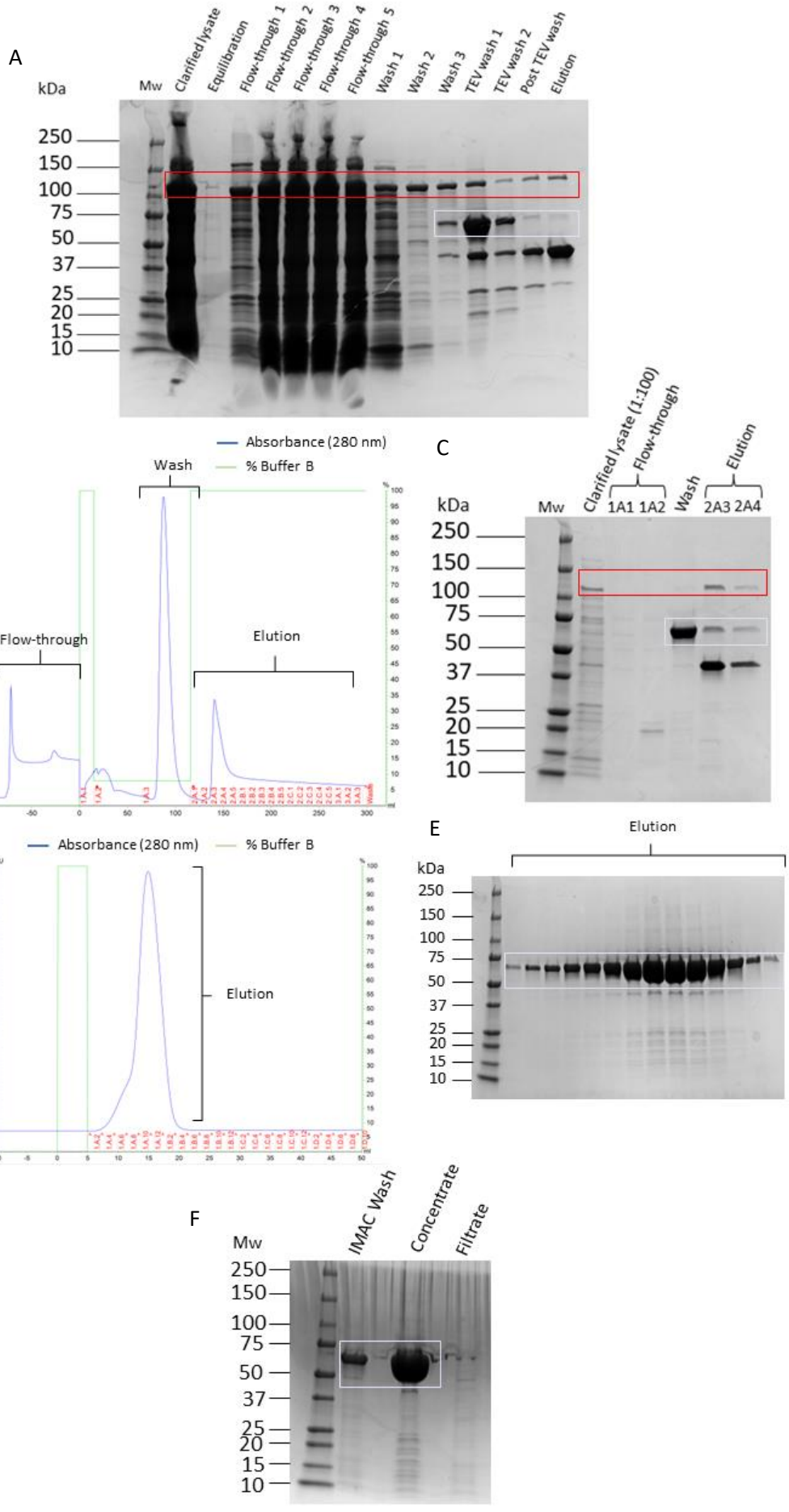

A
kDa
Mw
Clarified lysate
Equilibration
Flow-through 1
Flow-through 2
Flow-through 3
Flow-through 4
Flow-through 5
Wash 1
Wash 2
Wash 3
TEV wash 1
TEV wash 2
Post TEV wash
Elution
250
150
100
75
50
37
25
20
15
10
B
Absorbance (280 nm)
% Buffer B
Wash
Flow-through
Elution
C
kDa
Mw
Clarified lysate (1:100)
Flow-through
1A1 1A2
Wash
Elution
2A3 2A4
250
150
100
75
50
37
25
20
15
10
D
Absorbance (280 nm)
% Buffer B
Elution
E
Elution
kDa
250
150
100
75
50
37
25
20
15
10
F
Mw
IMAC Wash
Concentrate
Filtrate
250
150
100
75
50
37
25
20
15
10

Figure 5.11 revised FAN1-MBP purification method.
***A**. 4-20 % SDS-PAGE analysis of FAN1 purified using amylose resin. The column was equilibrated prior to loading processed supernatant onto the column. The column was washed to remove non-specifically bound or free-floating sample from the column before adding TEV protease. Elution of the sample was performed using the same buffer and 10 mM maltose. **B**. Ni^{2+} IMAC FF of FAN1. The wash steps were performed using 8 % elution buffer (20 mM imidazole) and the remaining contaminants were eluted with 250 mM imidazole. **C**. 4-20 % SDS-PAGE analysis of Flow-through (FL), wash and elution fractions from part B showing FAN1 and the MBP tag at 75, and ~42 kDa respectively. **D**. 2610 desalting into 2 × storage buffer (100 mM PIPES-NaOH pH 7.0, 200 mM KCl, 2 mM $CaCl_2$, 0.02 % NaN_3 and 10 mM DTT. **E**. 4-20 % SDS-PAGE analysis of the 2610 desalting column purification showing FAN1 at 75 kDa. The band highlighted with a light blue rectangle shows band intensity commonly by samples containing ~1.0 µg of protein. **F**. 4-20 % SDS-PAGE analysis of vivaspin-20 centrifugal concentration of the FAN1 protein. The FAN-MBP construct bands are highlighted by red rectangles (~117 kDa) and the cleaved FAN1 protein is highlighted by light-blue rectangles (~75 kDa).*

5.6 Summary

The primary aim of these experiments was to optimise expression and purification of FAN1 to ensure FAN1 stocks remained soluble and stable in solution prior to kinetic and biophysical investigation. In efforts to increase solubility, various fusion proteins were screened during FAN1 production and purification experiments including NusA and SUMO. However, MBP was shown to be the most effective in keeping FAN1 soluble during expression and aided in protein purification. Furthermore, several purification strategies were tested with varying degrees of success. Purifications using amylose resin was shown to be the critical step in removing contaminants from the sample, where other purification methods failed. In addition to this, amylose resin also enabled the removal of the MBP/his-tag from FAN1 using gravity flow funnel resin apparatus. However, despite the success of amylose resins, a significant amount of FAN1 failed to bind the resin, likely due to column overloading. To combat this, significant investment would need to be made in increasing the amylose resin volume if larger cultures of FAN1-MBP are desired. Amylose resin accompanied by Ni^{2+} IMAC was extremely effective in purifying FAN1 to homogeneity. DSF screening also proved useful in identifying buffer conditions that increased the T_M of FAN1 through PIPES and sodium phosphate buffers. Moreover, DSF screening also showed that FAN1 favours lower imidazole environments which were subsequently incorporated into the Ni^{2+} IMAC purification steps with significant improvements to both yield and solubility.

Chapter 6 Preliminary biophysical characterisation of FAN1

5.1 Introduction

Following successful purification of the FAN1-MBP construct (residues 364-1017), assays that monitor the products formed in FAN1 catalysed reactions were conducted to give greater insight into the kinetic properties of the protein. FAN1 investigations of substrate affinity showed that the enzyme prefers to bind to a double flapped DNA structure. Moreover, DNA substrates bearing an 8-nucleotide 3'-flap accompanied by a 1-, 2- or 0-nucleotide 5'-flaps afforded K_d values of 0.4, 1.4 and 1.0 nM respectively whereas substrates without an 8-nucleotide 3'-flap yielded a significantly increased the K_d (Figure 6.1A). Structural studies have also suggested that the phosphorylation status of the 5'-terminus was an important factor during monomeric FAN1-mediated catalysis; the 5'-phosphate enters a basic pocket and makes contacts to the side chains of R706, R952, H742 and L986 (Figure 6.1B). After association of the 5'-phosphate to the basic pocket, the scissile phosphate meets the divalent metal ion active site co-ordinated by D960 and E975 (Figure 6.1C). Double alanine mutations of the basic pocket residues (R706A/R952A) resulted in a 400-fold reduction in K_d (210 nM) for substrates bearing a 1-nucleotide phosphorylated 5'-flap and an 8-nucleotide 3' flap (denoted as DF(p1,8)) (Figure 6.1C). The 5'-phosphate of substrates containing a 2-nucleotide 5'-single flap (SF(p2,0)) or a phosphorylated nick (pNi,0) were also observed in the basic pocket and made contacts with R706, R952, H742 and K986, which further highlights the necessity for a 5'-phosphate [102]. Nuclease activity assays have provided evidence that FAN1 produces multiple products during catalysis. For DF(p1,8), FAN1 follows an N+3 recognition pattern and makes incisions every 3-nucleotides into the processed strand. Although, this model can fluctuate depending on the location of the 5'-phosphate. For example, FAN1 in complex single-flapped substrates with SF(p1,0) and pNi,0 presented N+4 and N+2 models respectively [102].

When two different proteins come together to form a complex, this is known as a heterodimer whereas the combination of two identical proteins are referred to as a homodimer. Recent reports have suggested that FAN1-mediated substrate catalysis also functions as a homodimer (Figure 6.1D). Moreover, it is claimed that dimerisation occurs upon binding of the substrate. Further evidence for substrate-induced homodimerization

comes from the mutations K525E, R526E, and K528E (Figure 6.1D) on α9 that prevented substrate induced FAN1 dimerisation and reduce endonuclease activity [105].

Monomeric and dimeric forms of FAN1 were also observed in bacteria. *Pseudomonas aeruginosa* (Pa) FAN1 was reported to hydrolyse the substrate as a monomer whereas *Vibrio vulnificus* (Vv) was reported to function constitutively as a dimer [155]. Despite this difference, nuclease activity assays revealed no significant difference in substrate processing of 5'-flap strands compared to human FAN1. Similar to human FAN1, extension of the 5'-flap resulted in reduced endonuclease activity; however, it was more severe for the monomeric PaFAN1 compared to VvFAN1 [155]. As few rigorous quantitative studies have been conducted with FAN1, we sought to establish a robust assay that could be used to compare the nuclease activity of FAN1 and its mutants to elucidate structure-function relationships and mechanism. It was also desirable to move away from the use of radioactively labelled substrates and towards the use of fluorescent labels. Therefore, this chapter will focus on identifying that the purified human FAN1 protein from the MBP construct was active and showed similar product formation patterns as previously reported. In addition, the viability of using fluorescent substrates in combination to RP-dHPLC separation to assess FAN1 catalysed reactions will be described.

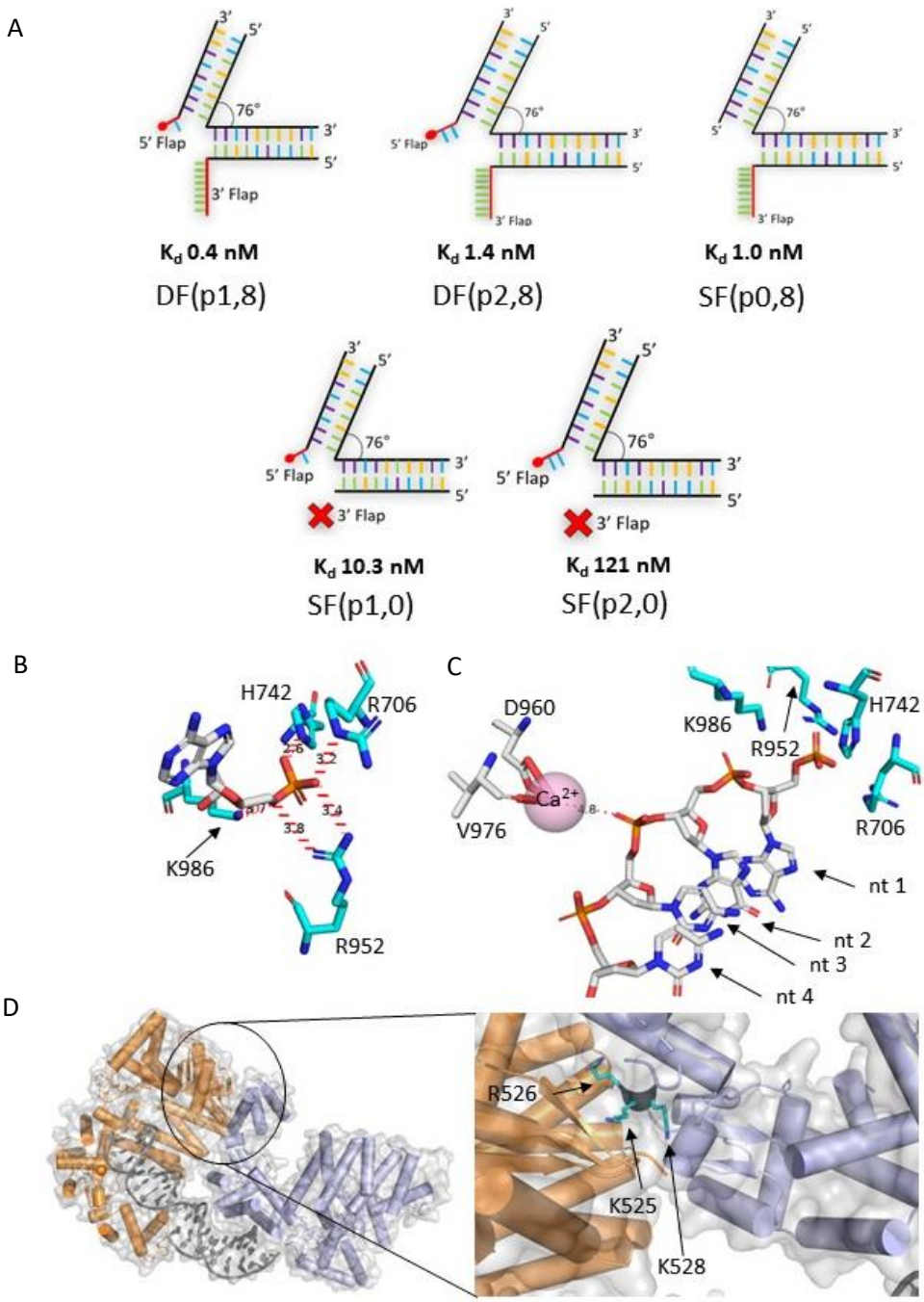


Figure 6.1 FAN1 structure and proposed substrate specificity

***A**. Proposed substrates and assigned K_d values of FAN1 from binding assays using ^{32}P labelled substrates under non-catalytic conditions from the Pavletich group. The K_d values were obtained by fitting a one site binding model* [102]. ***B**. FAN1 5'-phosphate binding pocket. The three non-bridging oxygens of the 5'-phosphate form hydrogen bonds with R706, H742, R952 and K986 (PDB code: 4RIA). **C**. Active site residues D960 and V976 co-ordinating a Ca^{2+} ion towards the phosphate diester of the third nucleotide B3 (Blue). **D**. Dimeric FAN1 in the 'substrate scanning' mode with the FAN1 responsible for catalysis in orange and the other FAN1 monomer (purple) functions in substrate binding and stabilisation. The image also contains a magnified view of α9 on the violet coloured FAN1 where basic residues form electrostatic interactions with the protein-protein interaction domain of the catalytic FAN1 monomer (PDB code: 4REB)* [105].

5.2 FAN1 nuclease assays and HPLC method development

5.2.1 FAN1 nuclease assay with SB5,1,HT2

Following purification of FAN1 using the MBP construct (section 5.3.2), a nuclease activity assay was performed to determine if FAN1 was catalytically active. The enzyme activity was initially observed by separation of substrate and product using RP-dHPLC. Initial FAN1 activity assays were performed under semi-single turnover conditions and a 5'-fluorescently labelled DNA 5'-flap substrate SB5,1,HT2, which possess a 5-nucleotide 5'-flap and a 1-nucleotide 3'-flap (Figure 6.2A). Once initiated, reactions were quenched with 500 mM EDTA and analysed by RP-dHPLC. The reaction was initiated by mixing FAN1 (diluted in 1× RRB) and the SB5,1,HT2 substrate (diluted in 1×FB) to a final concentration of 400 nM and 10 nM respectively and timepoints were sampled between 0.5-256 minutes (Figure 6.2B). At the first data point of 0.5 minutes, 2.5 nM product was observed, and this remained near constant between 0.5-4.0 minutes before sharply increasing to 7 nM product production after 16 minutes. Product formation began to plateau after 32 minutes with 10 nM product formation after 256 minutes. A first order rate was calculated as 0.05 min^{-1} using Akaike information criterion (AIC) comparison. Single exponential models had a square of -4.4 whereas double exponential showed an AIC of 3.4. A difference of -7.8 in AIC scores led to the single exponential model being preferred. However, after reviewing the residual plot, it was clear that the chosen model does not accurately model the data (Figure 6.2C). Fitting to a single exponential model, a Y0 of 1.17 nM was calculated. Considering that zero product was produced at zero minutes followed by 2.5 nM (10 %) product formation at 0.5 minutes, this implies an initial burst of product formation upon initiation of the reaction. The rate of this phase must be at least 5 min^{-1} (2.5 nM/0.5 minutes), some 100 times greater than the measured rate of the slower portion of the reaction (0.05 min^{-1}). The observation of this burst was reproducible (n>3 repeats). The observation of a fast and slow phase of the reaction implies that two different species exist in solution. One of these species may react faster than the other or one may be inactive requiring re-assembly to the active form before the slow reaction occurs. These species could involve the monomeric and dimeric states of the FAN1-substrate complexes that have been observed by X-ray crystallography. Another explanation is that FAN1 dimerisation leads to the initial burst in product formation due to fast binding to the substrate SB5,1HT2. However due to SB5,1HT2 being a suboptimal substrate with

a longer 5'-flap and 5'-phosphate labelled with a 5'-FAM label, the slower phase is then observed which contributes towards catalysis of SB5,1,HT2. However, it must be stated that within this thesis, not enough evidence is afforded to support these claims and therefore further experiments are required to understand the slow and fast phases of this reaction.

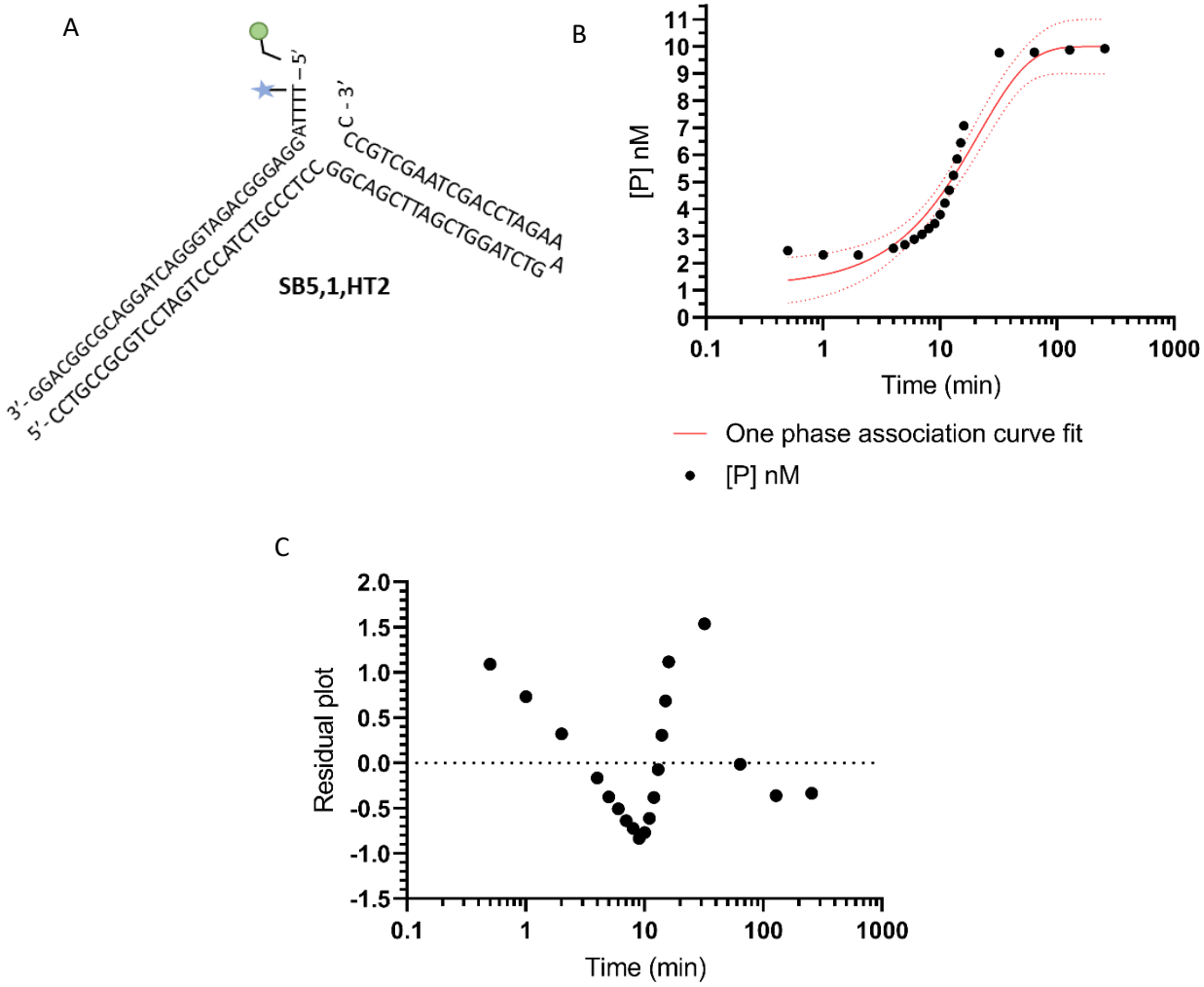
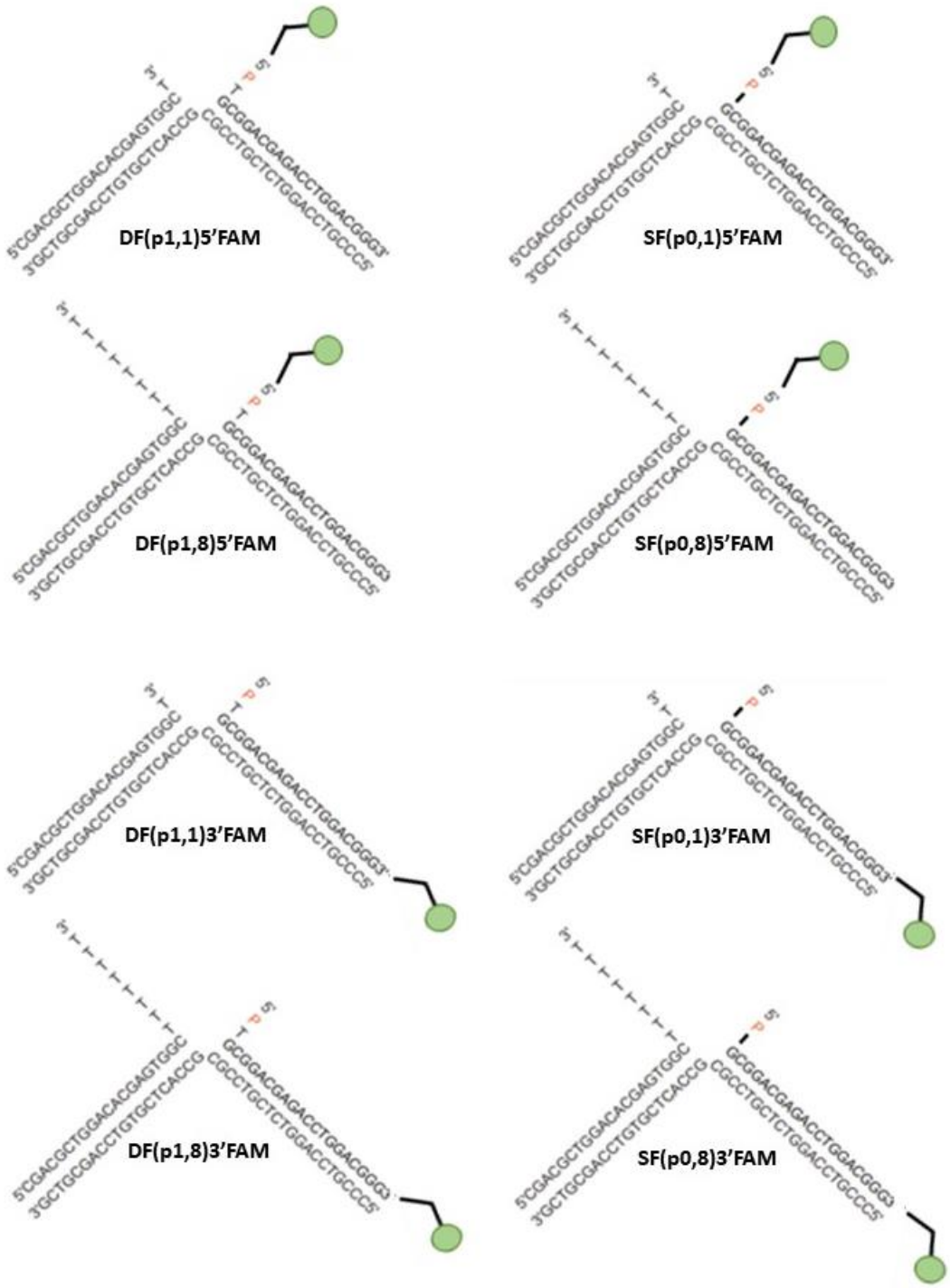


Figure 6.2 FAN1 nuclease assay with refined timepoints

***A**. SB5,1,HT2 substrate used to test FAN1 activity following expression and purification. **B**. Nuclease activity assay with 400 nM FAN1 and 10 nM SB5,1,HT2 at 0.5, 1.0, 2.0, 4.0, 8.0, 16.0, 32.0, 64.0, 128.0 & 256 minutes. Product concentrations were plotted against a log_{10} x-axis (black circles). Non-linear regression with one-phase association was used to plot a fitted curve to the product concentration (Red line) with upper and lower 95 % confidence limits incorporated into the analysis to identify outliers (red dashed line). A reaction rate of 0.05 min^{-1} and a Y0 of 1.17 nM. **C**. Residual plot of the plotted product concentration against a log_{10} x-axis. Plotted data was analysed using GraphPad Prism 9.0.*

5.2.2 FAN1 nuclease assays with DF(p1,8)3'FAM and DF(p1,8)5'FAM

The findings of section 5.2.1 provided evidence of an active FAN1 protein that could hydrolyse oligonucleotide substrates. The rate of reaction was measured using a RP-dHPLC system equipped with a fluorescence detector to separate and quantify products and reactants. However, the substrate used in these preliminary experiments were suggested to be sub-optimal based on binding studies of the FAN1 protein. Excluding ICL-linked substrates, FAN1 exhibited optimal affinity for the DF(p1,8) substrate [102]. Previous studies also implied FAN1 prefers substrates bearing a 5'-phosphate due to its interaction with the basic pocket which potentially offers substrate stabilisation or correct orientation into the active site [102], [105]. Therefore, it was decided to compare the efficiency of reaction using the substrate DF(p1,8) (Figure 6.3). FAN1 nuclease assays with the new substrates were performed using denaturing PAGE. Like RP-dHPLC, the fluorescein label was used to monitor the varying-length products formed during the reaction. Another exploratory aspect was the position of the fluorescent label, which was varied from a 5'FAM to a 3'FAM on the strand that was hydrolysed. It must be noted that experiments performed with the new substrates used a new batch of FAN1 (section 5.5.1) displaying an overall higher purity of sample with optimum buffer and additive conditions.



DF/SF	Double or single flap substrate
p	5' end of the processed strand is phosphorylated
1,	Length of the 5'-flap is 1-nucleotide
,8	Length of the 3'-flap is 8-nucleotides
5'FAM/ 3'FAM	FAM label attached to the 5' or 3' end of the 5'-flap strand

Figure 6.3 New oligonucleotide substrates used in FAN1 nuclease assays.

Oligonucleotide substrate structures used in FAN1 nuclease assays. Double flap and single flap structures are named 'DF' and 'SF' respectively. The length of the 5'-flap is shown by the left number found in the parentheses. The letter 'p' corresponds to the 5'-end of the processed strand is phosphorylated. The 3'-flap length is shown by the right most number in the parentheses. Substrates either contain a 3'FAM, or 5'FAM labelled on the processed strand.

Nuclease assays using the new substrates follows the same experimental configuration as section 5.2.1. Upon mixing of the reactants the final concentration of FAN1 was diluted between 0.4-400 nM and the substrates DF(p1,8)3'FAM or DF(p1,8)5'FAM were diluted to 10 nM final concentration. The results of the nuclease assays showed that the DF(p1,8)3'FAM substrate was consumed at a faster rate compared to the DF(p1,8)5'FAM substrate. Moreover, at 0.4 nM FAN1 with 10 nM DF(p1,8)3'FAM, 3 products were produced consistent with the literature which proposed an N+3 pattern [102]. In contrast, with DF(p1,8)5'FAM at the same concentration no product bands were visible.

Upon mixing of the reactants, the final concentration of FAN1 and DF(p1,8)3'FAM or DF(p1,8)5'FAM were 400 nM, and 10 nM respectively. For DF(p1,8)3'FAM, multiple products were produced, and DF(p1,8)3'FAM was completely consumed by 0.5 minutes. Shorter product lengths were observed at 64 minutes compared to 0.5 minutes (Figure 6.4A). For DF(p1,8)5'FAM, the substrate was consumed by 2-4 minutes (Figure 6.4B). When the FAN1 concentration was reduced to 40 nM, multiple products were still produced for DF(p1,8)3'FAM however the substrate remained present until ~2 minutes (Figure 6.4C). In contrast, an 8-fold increase in time was required to fully consume DF(p1,8)5'FAM at 40 nM FAN1 (Figure 6.4D). At 4 nM FAN1, three product lengths were predominantly observed and a further decreased in product length between 16-64 minutes for DF(p1,8)3'FAM (Figure 6.4E). Product formation for DF(p1,8)5'FAM began between 16-64 minutes (Figure 6.4F). Finally, at 0.4 nM FAN1, three products were produced between 2-64 minutes with DF(p1,8)3'FAM (Figure 6.4G) whereas no product formation occurred with DF(p1,8)5'FAM (Figure 6.4H).

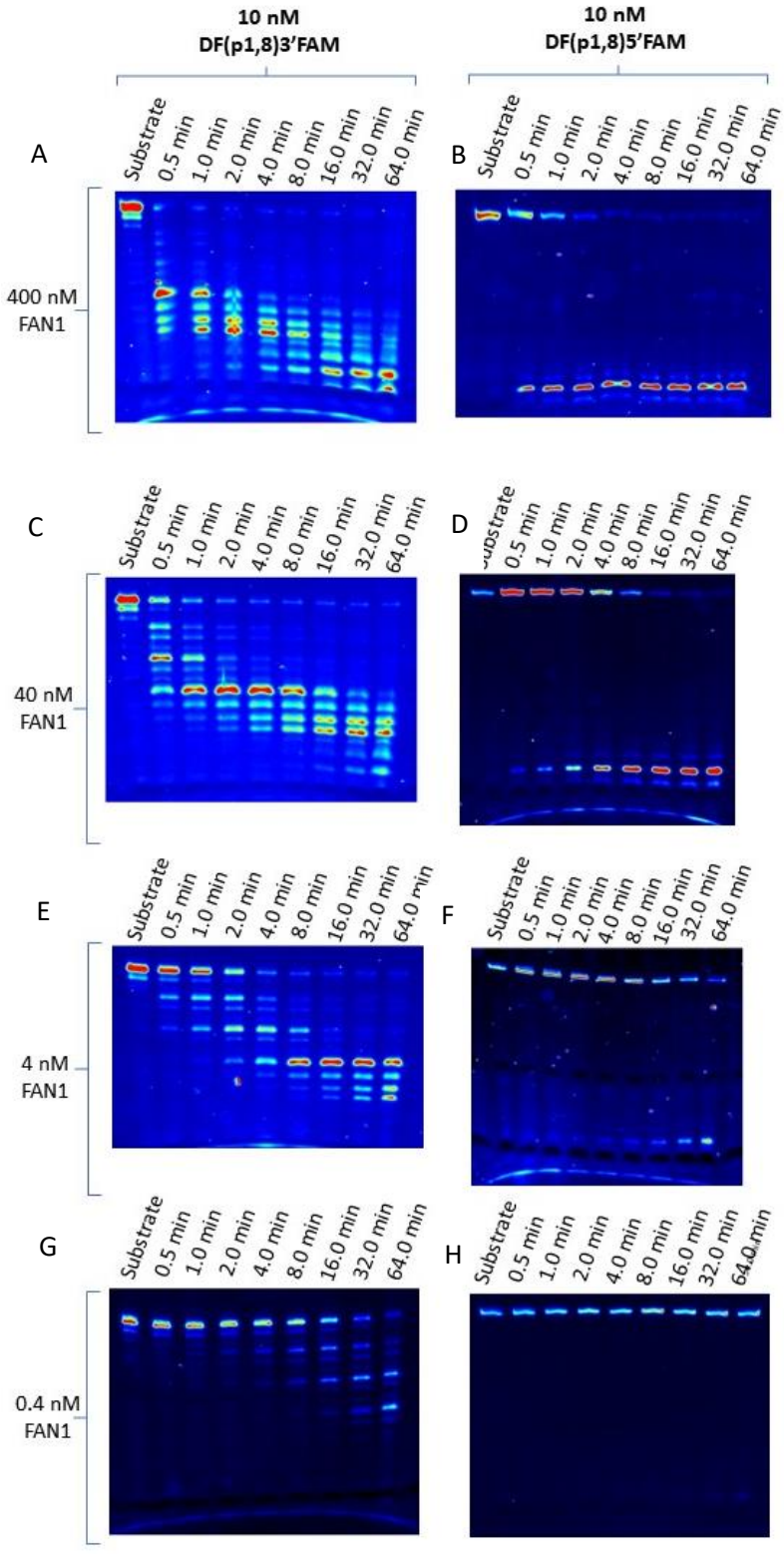


Figure 6.4 Nuclease activity assay of FAN1 with DF(p1,8)3'FAM and DF(p1,8)5'FAM

***A**. Nuclease assay of 400 nM FAN1 with 10 nM DF(p1,8)3'FAM. **B**. Nuclease assay of 400 nM FAN1 with 10 nM DF(p1,8)5'FAM. **C**. Nuclease assay of 40 nM FAN1 with 10 nM DF(p1,8)3'FAM. **D**. Nuclease assay of 40 nM FAN1 with 10 nM DF(p1,8)5'FAM. **E**. Nuclease assay of 4 nM FAN1 with 10 nM DF(p1,8)3'FAM. **F**. Nuclease assay of 4 nM FAN1 with 10 nM DF(p1,8)5'FAM. **G**. Nuclease assay of 0.4 nM FAN1 with 10 nM DF(p1,8)3'FAM. **H**. Nuclease assay of 0.4 nM FAN1 with 10 nM DF(p1,8)5'FAM*

This data provides evidence FAN1 hydrolyses the DF(p1,8)3'FAM substrate at a faster rate compared to DF(p1,8)5'FAM. Furthermore, 5'FAM labelled substrates may prove to be suboptimal when measuring the rate of reaction of FAN1. This provides further confirmation that FAN1 displays preference for double-flapped DNA substrates with a shorter 5'-flap and a longer 3'-flap as described in the literature. The formation of multiple product lengths within individual timepoints gave further information as to how FAN1 functions in the wider role of DNA repair. Rather than producing one specific nucleotide product, FAN1 may be employed to completely process the 5'-end of the substrate until significant nucleotide removal prior to the DNA damaged site is achieved. This may be required for inter-strand cross-link unhooking and removal by XPF-ERCC1, MUS81, and SLX1 [100], [134], [156]. 3'-labelling of the processed strand enables the ability to identify multiple products formed throughout the reaction although it poses issues for the easy monitoring of FAN1 reactions where the first product is recorded by a 5'FAM label and in turn used to calculate the rate of reaction.

To gain preliminary kinetic information, the reaction containing 4.0 nM FAN1 and 10 nM DF(p1,8)5'FAM produced approximately 10 % product after 10 minutes. The bands made visible by the 5'FAM label was quantified and the product concentration was calculated (Figure 6.5A). Simple linear regression yielded an initial rate of 0.11 nmoles/min^{-1} and a Y-intercept value of 0.30 nM (Figure 6.5A). However, the fit should be interpreted with caution considering only 4 datapoints were incorporated into the analysis. Moreover, the fitted plot displayed a wide 95 % confidence limit and residual plot show deviation from the linear regression model (Figure 6.5B).

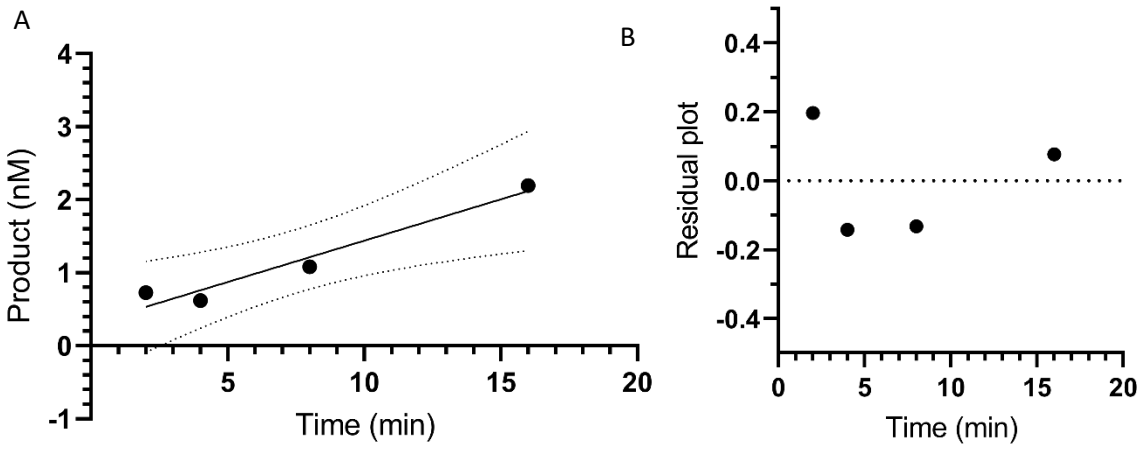


Figure 6.5 Simple linear regression of 4 nM FAN1 and 10 nM DF(p1,8)5'FAM

***A**. Simple linear regression fit of time points 2, 4, 8 & 16 minutes. The line of best fit shown by a black line. Upper and lower 95 % confidence limits were also displayed to show the level of deviation from the linear regression fit. Simple linear regression of the four timepoints calculated an initial rate of 0.11 nmoles/min^{-1}, a k_{cat} of 0.028 min^{-1} and a Y-intercept value of 0.3 nM. **B**. Residual plot of the 2-, 4-, 8- & 16-minute timepoints against time. The dashed line represents deviation of the timepoints from the linear regression fit. The plotted timepoints were analysed using GraphPad Prism 9.0.*

5.2.3 FAN1 nuclease assays with DF(p1,1)3/5'FAM, SF(p0,1)3/5'FAM and SF(p0,8)3/5'FAM substrates

From previous experiments, it was evident that FAN1 activity was far better with substrates where the fluorescein label was on the 3'-terminus of the substrate compared to the 5'-terminus. What was now required was to quantify the difference in product formation of single flapped substrates (SF(p0,8) and (SF(p0,1) and truncated double flaps (DF(p1,1) (Figure 6.3). Although DF(p1,8)5'FAM resulted in a comparatively slower reaction rate, it was unknown whether substrates with reduced 5'-flap length would observe the same kinetic characteristics. Thus, further testing of whether FAM labels on the 5', or 3'-terminus was monitored in the following experiments. Samples were quenched at 16 and 32 minutes using 0.4 nM enzyme to allow comparison of DF(p1,8)3'FAM and DF(p1,8)5'FAM because previous experiments reported three products at these timepoints.

Nuclease assays were performed on substrates SF(p0,8)3'FAM, SF(p0,1)3'FAM and DF(p1,1)3'FAM at a final concentration of 10 nM with 0.4 nM of FAN1 and unexpected results were achieved (Figure 6.6A). 3'FAM labelled substrates produced one product between 16 and 32 minutes irrespective of 5'-flap or 3'-flap length. One product was also produced for substrates lacking a 5'-flap or 3'-flap of 1 or 8- nucleotides (Figure 6.6A.). This was strikingly different to DF(p1,8)3'FAM where 2 shorter length products were also observed between 16-64 minutes (Figure 6.4G). On the other hand, 5'FAM labelled single flap substrates and truncated double flaps displayed limited to no product production between 16 and 32 minutes (Figure 6.6B) This was again attributed to FAN1 exhibiting a slower rate of reaction when exposed to 5'FAM labelled substrates on the 5' terminus of the 5'-flap. The production of one oligonucleotide product from SF(p0,8)3'FAM, SF(p0,1)3'FAM and DF(p1,1)3'FAM with 0.4 nM FAN1 was a peculiar finding as it does not follow the three-product pattern as shown previous experiments. This suggests that FAN1 requires double-flapped DNA structures with a 1-nucleotide 5'-flap and longer 3'-flap to perform repetitive incisions into the substrate. In the wider role of DNA repair, this may act as a failsafe mechanism to ensure FAN1 does not continuously hydrolyse substrates that are not scheduled to be processed by the enzyme. Thus, if FAN1 binds a substrate it is not scheduled to hydrolyse, FAN1 will only produce one nucleolytic incision instead of making multiple hydrolysis events on

the same substrate. No product was observed in the reaction containing SF(p0,8)5'FAM and only faint bands were observed in reactions containing SF(p0,1)5'FAM and DF(p1,1)5'FAM. Again, this demonstrates that the presence of the 5'FAM label reduces FAN1's ability to hydrolyse the substrate.

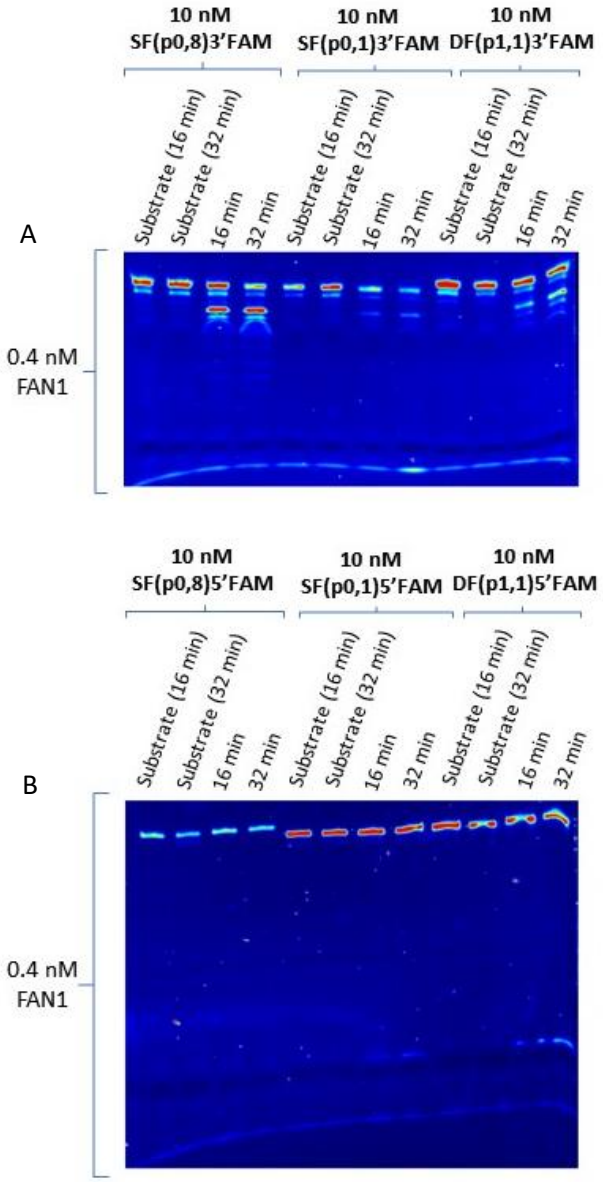


Figure 6.6 FAN1 nuclease assay with SF(p0,8)3/5'FAM, SF(p0,1)3/5'FAM and DF(p1,1)3/5'FAM
***A**. Denaturing PAGE analysis of 0.4 nM FAN1 and the substrates SF(p0,8)3'FAM, SF(p0,1)3'FAM and DF(p1,1)3'FAM. Separate reactions were produced for each timepoint, and negative controls were removed from each reaction before the addition of 0.4 nM FAN1. **B**. Denaturing PAGE analysis of 0.4 nM FAN1 and the substrates SF(p0,8)5'FAM, SF(p0,1)5'FAM and DF(p1,1)5'FAM. Separate reactions were produced for each timepoint, and negative controls were removed from each reaction before the addition of 0.4 nM FAN1.*

The question remains as to how the 5'FAM label (Figure 6.7A) leads to reduced FAN1 nuclease activity. This finding may occur through steric hindrance caused between the bulky FAM label and the 5'-phosphate binding pocket (Figure 6.7B). Because of this, the 5'-phosphate would be unable to make electrostatic, and hydrogen bond contacts with R706, H742, R952 and K986. This in turn may reduce FAN1's ability to achieve a catalytically competent state prior to hydrolysis of the scissile phosphate. Another explanation is that FAN1 possess higher affinity for phosphate monoester groups rather than diesters at the end of the 5'-flap (Figure 6.7C). 5'-flaps attached to a 5'FAM label become a phosphate diester and therefore loses charge which may impact its ability to form electrostatic contacts with the basic residues beyond the active site. However, robust conclusions cannot be made until further experiments determine whether the phosphodiester is required for efficient FAN1-mediated catalysis.

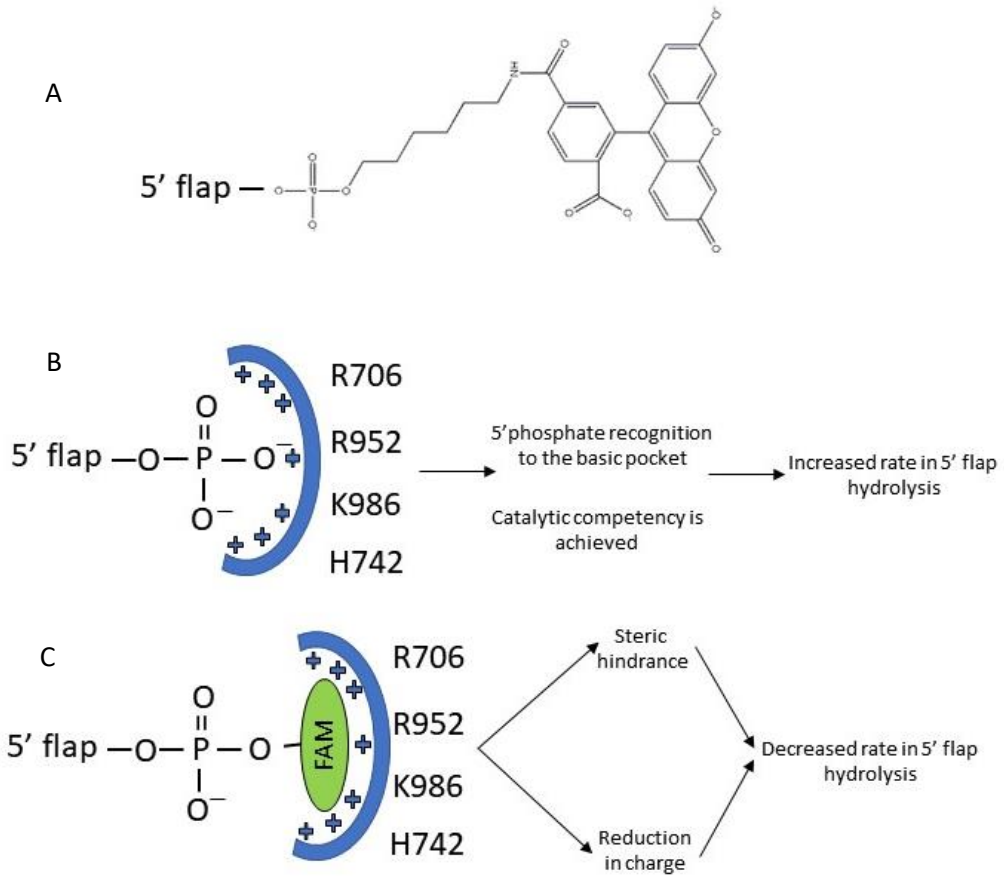


Figure 6.7 FAN1 basic pocket and its role in FAN1 catalysis

***A**. Structure of the 5'FAM label forming a phosphodiester bond to the 5'-flap of the substrate. The structure of the FAM label is a modified image from IDT Integrated DNA Technologies website. **B**. 5'-flap phosphate monoester forming hydrogen bonds and electrostatic interactions with the four basic residues after the active site of FAN1. **C**. 5'-flap substrate with a phosphate diester bond to the 5'FAM label potentially causing steric hindrance or removing charge from the phosphate group of the 5'-flap. Leading to a decrease in FAN1 reaction rate.*

5.2.4 Reverse-phase ion pair HPLC gradient screening of 3'FAM PB1 and DF (1,8) 3'FAM samples

The data presented within this chapter has shown that the use of 5'FAM labelled substrates was not a viable avenue to measure FAN1 activity. A major disadvantage with denaturing UREA gels is their time-consuming preparation. Because of this, it was believed that incorporating RP-dHPLC analysis would be a suitable method to replace denaturing PAGE separation of quenched substrate and product samples. The use of RP-dHPLC would reduce the time taken to analyse cleaved products. Semi-high throughput RP-dHPLC combined with FAM substrates allow easy product quantification to calculate the rate of reaction.

2.5 mM TBAB was tested in the elution buffer as an ion pairing reagent accompanied by the organic modifier, MeCN, for elution. SsDNA oligos were used as molecular weight standards at 11, 13, 15 and 17-nucleotides in length to resemble similar sized FAN1 products following hydrolysis of the substrate DF(p1,8)3'FAM. After several tests, an optimal gradient was achieved whereby all product and product molecular weight standard peaks were separated by RP-dHPLC with a gradient of 28.0-42.0 % MeCN (Figure 6.8A). Despite this, better separation between the non-hydrolysed control sample and the 17-nucleotide standard was never achieved despite extensive efforts at optimisation (Figure 6.8A). 100 mM TEAA was also tested as the ion pairing reagent with a gradient of MeCN. This provided greater separation between the non-hydrolysed substrate peak and the 17-nucleotide product standard; however, it failed to separate the molecular weight standards. A peak was also observed in the 100 % buffer B wash stage containing 25 % MeCN. It is suspected that this may be non-hydrolysed control strands eluting later off the column which raises issues as to whether the gradient could reliably elute all the product, and control samples at the same retention time (Figure 6.8B).

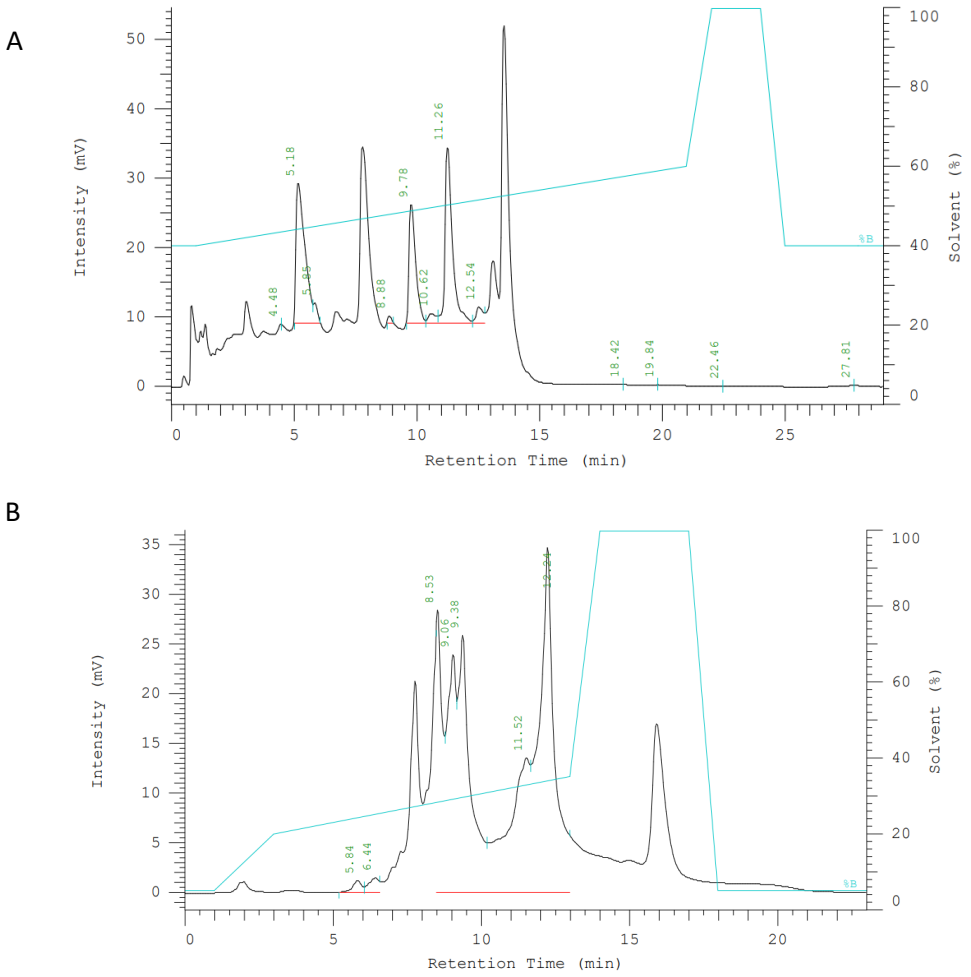


Figure 6.8 RP-HPLC gradient scouting for FAN1 product analysis

***A**. RP-HPLC of molecular weight standards and DF(p1,8)3'FAM. A gradient of 28.0-42.0% acetonitrile with 2.5 mM TBAB was used for the elution gradient before a wash step with 100 % buffer B (70 % MeCN). **B**. RP-dHPLC of molecular weight standards and DF(p1,8)3'FAM. A gradient of 5.0-8.5 % MeCN with 100 mM TBAB was used for the elution gradient before a wash step with 100 % buffer B (25 % MeCN).*

It was believed that FAN1 would be a suitable candidate for RP-dHPLC analysis from its ability to produce products of 3-nucleotides in length that could be routinely resolved from the 20-21 length un-cleaved substrate strand as previously suggested from the previously published N+3 pattern by *Pavletich et al* [102]. FAN1 exhibits similar functionality to FEN1. However, one major difference between the two nucleases is that FAN1 has a slower rate of reaction in the presence of 5'FAM labelled substrates. Because of this, if fluorescence detection is the desired method desired method to visualise FAN1 reactions, the cleaved substrate must contain a 3'FAM label. However, this results in multiple products formed and therefore,

multiple peaks of only 3-nucleotide difference in length. For RP-dHPLC analysis to be a feasible option with future FAN1 nuclease assays, limiting the number of products formed would be useful. Truncation of the 5' and 3'-duplexes may force FAN1 to only produce a single oligonucleotide product. However, this process of shortening the substrate could cause more potential problems including potentially reduced binding affinity of FAN1 to the substrate due to a lack of residue contacts to the 5' and 3'-duplexes. A final point regarding gradient testing is the length of time taken to perform the elution gradient. If longer gradients were able to provide greater peak separation, it also poses a significant issue in terms of MeCN required to perform semi-high throughput nuclease activity assays of FAN1. Specifically, the assay becomes more expensive due to more MeCN required to perform the assay. Moreover, the assay will produce more organic solvent waste, which is less environmentally sustainable. As a result of these issues, it was desirable to reduce the total time of the RP-dHPLC analysis for each sample and find optimum gradient percentages in the timescale of 10-12 minutes to reduce time taken to analyse a 96-well plate of quenched reactions and reduce the volume of MeCN waste which needs to be disposed of. In the end, it would show that if RP-dHPLC analysis of FAN1 products were to be continued in the future, longer elution gradient time may need to be factored into the experimental design.

5.2 Summary

To summarise, the conclusions of the combined experiments determining FAN1 nuclease activity in the presence of varying flap length and position of the FAM label suggest that FAN1 displays increased rate of reaction for substrates containing a 3'FAM label compared to identical substrates possessing a 5'FAM label. Moreover, DF(p1,8)3'FAM produces multiple products between 0.4-400 nM enzyme concentration whereas 3'FAM labelled substrates in the absence of a 5'-flap, truncated 3'-flap of 1-nucleotide or a shortened double flap with 0.4 nM FAN1 produce one product. Overall, this data suggests similarities to the conclusions of the literature where FAN1 prefers shorter 5'-flaps and longer 3'-flaps. Furthermore, three products were produced at 0.4 nM FAN1 alluding to the proposed N+3 incision pattern shown in the literature. However, without molecular weight markers to identify product length, this remains partially enigmatic with regards to our nuclease assays. The RP-dHPLC experiments were performed to offer a more convenient alternative to denaturing PAGE gels to monitor FAN1 reaction. However, further gradient optimisation is required to achieve optimal/reproducible substrate and product traces. Furthermore, a new RP-dHPLC column is required to ensure high resolution peaks during analysis, as samples containing BSA and FAN1 will lead to peak broadening because of column degradation. For monitoring the rates of reaction using 3'FAM labelled substrates, it is critical that products are reproducibly separated from the substrate. The separation of differing lengths products from one another is not essential but were this achieved, important information as to the characteristics of the reaction would be obtained. Ion-exchange HPLC analysis is an alternative route for substrate and product separation due to separation based on relative charge of the phosphodiester backbone of the bound sample. This would also eliminate the requirement for large volumes of MeCN. However, this still requires extensive gradient optimisation and it is unknown whether ion-exchange can provide single-nucleotide of FAN1-catalysed products compared to RP-dHPLC [157].

Chapter 7 Conclusions and Future work

7.1 Kinetic and biophysical studies of FEN1 active site and H2TH mutants

7.1.1 Chapter 3 conclusions and Future work

The 5'-flap endonuclease FEN1, is an extensively researched enzyme because of its essential role in DNA replication and repair. The purpose of chapter 3 was to replicate the findings from *Song et al* in which they implied conformational change was rate-limiting for the FEN1 mutant D34A reaction. D34 is a component of the FEN1 active site that is believed to interact directly with one of the catalytic Mg^{2+} ions. *Song et al* suggested that in the presence of Mg^{2+} ions, a burst in 2-AP fluorescence was observed between 0-30 seconds, despite an extremely slow reaction of substrate that required a much greater timescale [68]. This was starkly different to previous kinetic reports where enzyme-product release was believed to be rate-limiting throughout the FEN1-DNA reaction mechanism. Through our own 2-AP assays using the more conservative D34N mutant, the presence of a burst was not recorded, suggesting opposite findings to *Song et al*. Furthermore, single turnover D34N experiments demonstrate varying length 5'-flap products are formed in the D34N-DNA reaction and may be the cause of a lag phase observed in the 2-AP assay. This underscores the importance of a D34-Mg^{2+} contact and its role in positioning the 5'-flap for reproducible hydrolysis of the +1-1 scissile phosphate. However, the findings of chapter 3 do not directly conclude that D34A does not alter the rate-limiting characteristics of FEN1. To fully unearth the mechanistic properties of these active site mutants, stopped-flow experiments must be undertaken before accurate, and reproducible comparisons between our datasets and that of *Song et al*.

7.1.2 Chapter 4 conclusions and future work

Considerable research has been performed on the FEN1 protein, especially in domains with high residue conservation amongst the FEN superfamily including the helical gateway, active site and 3'-flap binding site (Figure 7.1). However, limited information was available surrounding the H2TH domain and its role during FEN1-mediated substrate processing. Reports by *Tainer et al* and *Beese et al* have proposed that the H2TH domain acts to anchor the substrate and form a track for the phosphodiester backbone to bind to [137], [140]. Therefore, in chapter 4, kinetic and ECCD binding studies were performed to characterise nuclease activity and DNA conformational change at the scissile phosphate of mutants of the H2TH motif, R239A, K244A and R245A. Ultimately, FEN1 nuclease activity was sustained throughout all the H2TH mutants with the catalytic turnover within proximity of the WT protein. Only R245A showed the highest change in K_M compared to the WT protein which aligns with previous literature as to how R245 provides electrostatic interactions between the template backbone of the substrate, and the H2TH domain. Although nuclease activity remained, the catalytic efficiency of the mutants was partially impaired compared to WT, implicating the H2TH domain as a necessary domain for optimal substrate orientation, incision, or binding. Kinetic experiments also displayed similar k_{STmax} rates compared to WT; the single turnover rates further proved enzyme-product release was rate-limiting for all three H2TH mutants. ECCD assays using tandem 2-AP labelled substrates showed that DNA conformational change at the +1-1/-1-2 positions of the 5'-strand was only partially affected by the H2TH mutants, and that AST of the scissile phosphate could still feasibly be achieved albeit its less efficient manner with all three H2TH mutants. Ultimately, these findings suggest that R239, K244 & R245 do not have individual roles during catalysis, but instead, the three residues may work together to form a basic environment to provide optimal anchoring of the template strand to the H2TH domain and its associated K^+ ion.

To further explore the nature of the H2TH basic residues, double and triple residue mutants would offer greater understanding as to how complete lack of basic environment changes the kinetic parameters accompanied by potentially reduced DNA conformational changes towards the active site. Furthermore, the incorporation of microscale thermophoresis (MST) or smFRET would allow the calculation of the binding equilibrium constant K_d towards the H2TH based mutants and the double flapped substrates which is lacking

in the literature. Finally, optimal binding between the H2TH domain and the template strand on the 5'-duplex may be facilitated primarily through the K^+ binding sight and is only supplemented by the electrostatic properties of basic residues. Therefore, Repeating the kinetic and ECCD experiments in the presence of other monovalent metals would elucidate whether binding and nuclease activity is dependent on the presence of metal during the reaction.

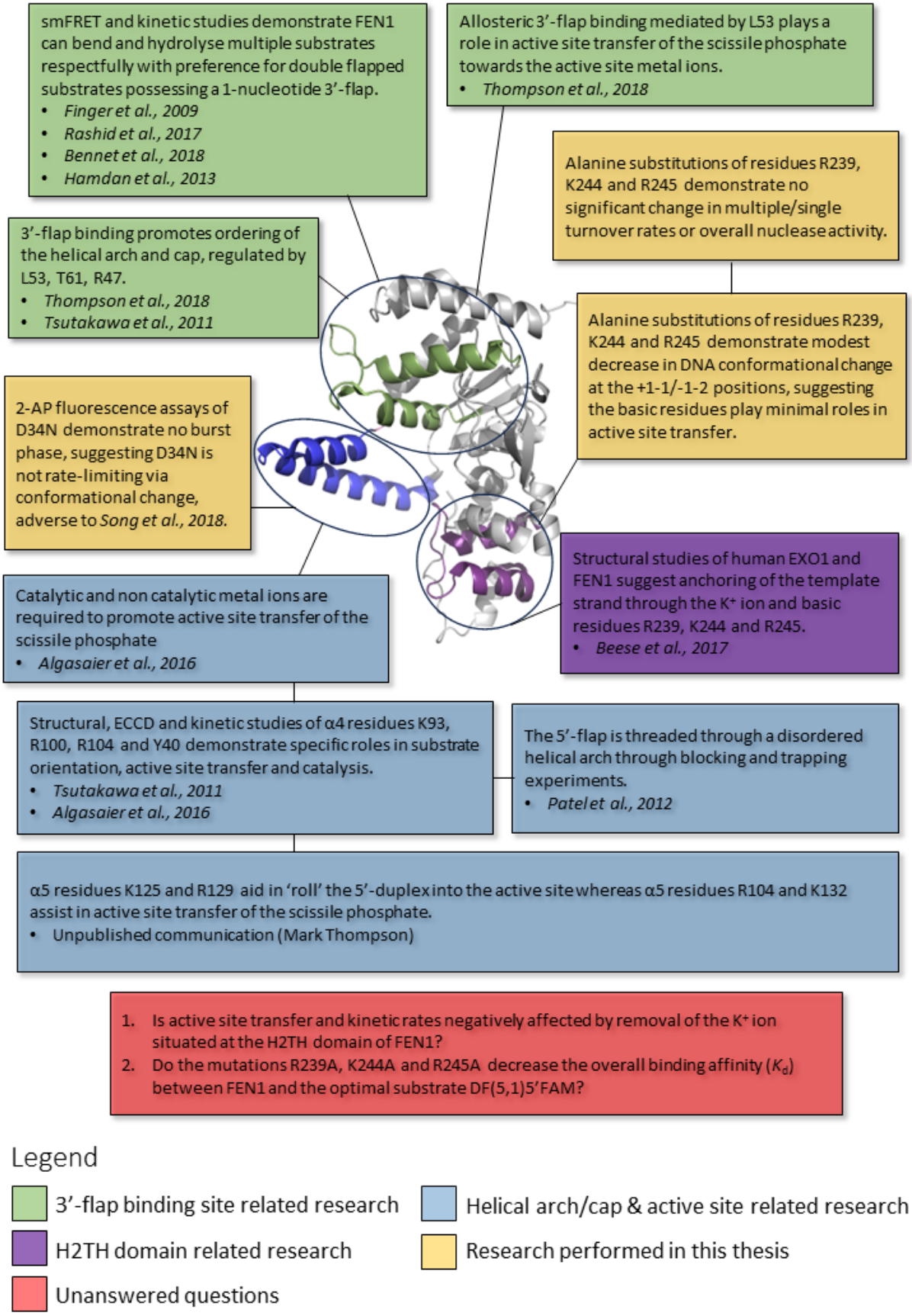


Figure 7.1 Summarising schematic of FEN1-related research

Schematic demonstrating previous FEN1-related research colour-coded by individual structural components of FEN1 followed by research performed in this thesis and remaining unanswered questions

7.2 Production and preliminary characterisation of FAN1

7.2.1 Chapter 5 and chapter 6 conclusions and future work

The latter chapters of this thesis focus on the production and characterisation of FAN1, a 5'-nuclease involved in multiple DNA repair pathways. Multiple reports have presented various mechanisms as to how FAN1 operates during substrate processing (Figure 7.2). Reports have suggested that FAN1 shows preference for double flapped substrates bearing shorter 5'-flaps and longer 3'-flaps. It is also debated whether FAN1 operates as a homodimer or monomer. The relatively ambiguous nature surrounding FAN1 function made it clear that an in-depth kinetic exploration of the FAN1 protein would shed light on the conformational preferability accompanied by a greater insight into the FAN1 reaction mechanism. Due to FAN1 being a new protein to the Grasby group, chapter 5 explored multiple production strategies to ensure FAN1 remained stable in solution. DSF screening suggested conditions that increased the thermal stability of FAN1 compared to the original TRIS-HCl buffer, and the incorporation of an MBP fusion protein to the FAN1 sequence both increased solubility, and aided purification of the FAN1-MBP construct via amylose resin batch purification. Continued optimisation efforts may include upscaling production to achieve a higher concentration of FAN1 with minimal loss in yield during each purification step. Furthermore, stability studies would offer more information as to the half-life of FAN1 over prolonged storage time. This could be further supplemented with DSF experiments testing the protein at varying times during storage and assess the change in T_M compared to nuclease activity.

Preliminary FAN1 characterisation and assay development began in chapter 6. These assays showed a burst followed by a slow phase with the SB5,1,HT2 substrate whereas the rate of reaction significantly increased upon shortening the 5'-flap and elongation of the 3'-flap with a 3'FAM label on the hydrolysed strand. Depending on enzyme concentration, nuclease assays also displayed slower or complete abolishment of product formation when FAN1 was mixed with 5'FAM labelled substrates over 64 minutes with truncated double flaps or elongated 3'-single flaps. This possibility demonstrates that FAN1 activity is hindered by the bulky 5'FAM label or catalytic competency being severely depleted because of limited access of the 5'-phosphate at the basic pocket. The nuclease assays also showed three products formed with DF(p1,8)3'FAM as shown in the literature, although the product length cannot be determined without repeating the assays

with molecular markers. However, this three-product incision pattern was not observed with truncated double flaps or 3'-single flaps due to only 1 product was formed. Overall, it can be concluded from chapter 6 that FAN1 prefers double flapped substrates with shorter 5'-flaps and longer 3'-flaps and structure of the substrate may dictate the number of products formed.

Considering FAN1s preference for 3'FAM labelled substrates, further method development is required to optimise HPLC analysis of the FAN1-DNA reaction as the result of multiple products being visualised by the fluorescence detector. However, FRET assays could offer isolation of one specific hydrolysis event due to the loss in FRET signal once the 5'-flap is removed. Additionally, substantial contacts are made between R752, H718, R710 and R679 and the template strand proximal to the active site. A high density of contacts could imply that this interface is important for active site transfer of the scissile phosphate and could be explored further through site directed mutagenesis and kinetic studies. Finally, it remains unknown of the level of internal conformational changes that occur during the reaction and the employment of NMR may offer more information as to the different structural states of the FAN1 protein throughout the reaction.

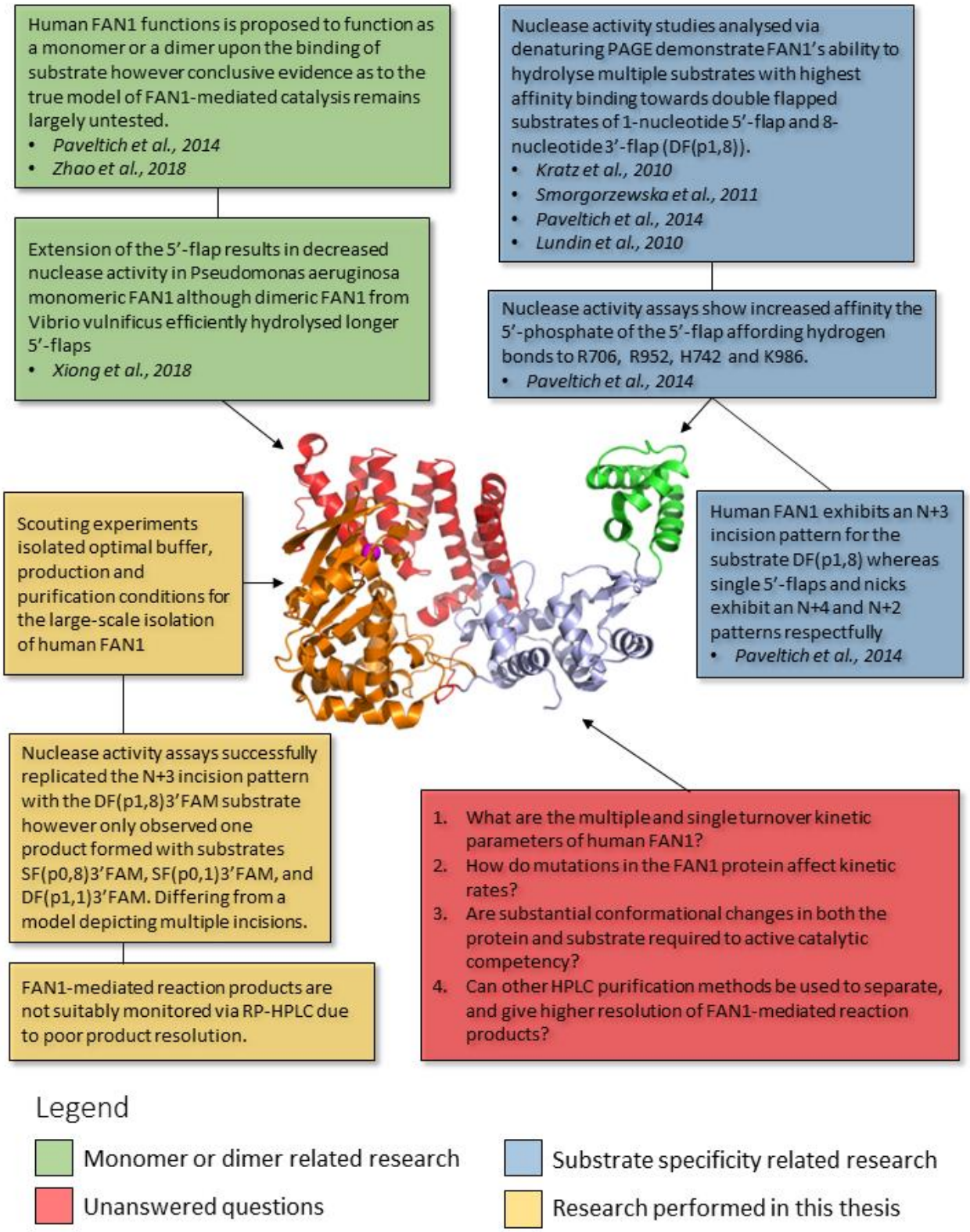


Figure 7.2 Summarising schematic of FAN1-related research

Schematic demonstrating previous FA1N-related research colour coded by substrate specificity and proposed models of catalysis. This schematic also includes research performed within this thesis and unanswered questions.

References

[1] G. Laflamme *et al.*, "Structural maintenance of chromosome (SMC) proteins link microtubule stability to genome integrity," *J. Biol. Chem.*, vol. 289, no. 40, pp. 27418–27431, 2014.

[2] A. Travers and G. Muskhelishvili, "DNA structure and function," *FEBS J.*, vol. 282, no. 12, pp. 2279–2295, 2015.

[3] H. Subramanian and R. A. Gatenby, "Evolutionary advantage of anti-parallel strand orientation of duplex DNA," *Sci. Rep.*, vol. 10, no. 1, pp. 1–18, 2020.

[4] A. Rich and S. Zhang, "Z-DNA: the long road to biological function," vol. 4, no. July, pp. 566–572, 2003.

[5] J. Watson and F. Crick, "Molecular structure of nucleic acids," *Nature*, vol. 171, no. 4356, pp. 737–738, 1953.

[6] N. Chatterjee and G. C. Walker, "Mechanisms of DNA damage, repair, and mutagenesis," *Environ. Mol. Mutagen.*, vol. 58, no. 5, pp. 235–263, 2017.

[7] J. Malinge and M. Leng, "Interstrand cross-links of cisplatin induce striking distortions in DNA," *Elsevier*, vol. 77, pp. 23–29, 1999.

[8] P. D. Lawley and D. H. Phillips, "DNA adducts from chemotherapeutic agents," *Mutat. Res. - Fundam. Mol. Mech. Mutagen.*, vol. 355, no. 1–2, pp. 13–40, 1996.

[9] T. Nishino and K. Morikawa, "Structure and function of nucleases in DNA repair: Shape, grip and blade of the DNA scissors," *Oncogene*, vol. 21, no. 58 REV. ISS. 8, pp. 9022–9032, 2002.

[10] G. K. Schroeder, C. Lad, P. Wyman, N. H. Williams, and R. Wolfenden, "The time required for water attack at the phosphorus atom of simple phosphodiesters and of DNA," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 103, no. 11, pp. 4052–4055, 2006.

[11] S. Mikkola, T. Lönnberg, and H. Lönnberg, "Phosphodiester models for cleavage of nucleic acids," *Beilstein J. Org. Chem.*, vol. 14, pp. 803–837, 2018.

[12] T. A. Steitz, "DNA polymerases: Structural diversity and common mechanisms," *J. Biol. Chem.*, vol.

274, no. 25, pp. 17395–17398, 1999.

[13] L. S. Beese and T. A. Steitz, “Structural basis for the 3’- 5’ exonuclease activity of Escherichia coli DNA polymerase I: A two metal ion mechanism,” *EMBO J.*, vol. 10, no. 1, pp. 25–33, 1991.

[14] Y. Sun, J. Li, and L. Xia, “Precise genome modification via sequence-specific nucleases-mediated gene targeting for crop improvement,” *Front. Plant Sci.*, vol. 7, pp. 1–13, 2016.

[15] M. S. Zaher *et al.*, “Missed cleavage opportunities by FEN1 lead to Okazaki fragment maturation via the long-flap pathway,” *Nucleic Acids Res.*, vol. 46, no. 6, pp. 2956–2974, 2018.

[16] E. M. Goellner *et al.*, “Exonuclease 1-dependent and independent mismatch repair,” *DNA Repair (Amst).*, vol. 1, no. 32, pp. 24–32, 2016.

[17] E. Bolderson *et al.*, “Phosphorylation of Exo1 modulates homologous recombination repair of DNA double-strand breaks,” *Nucleic Acids Res.*, vol. 38, no. 6, pp. 1821–1831, 2010.

[18] U. Rass *et al.*, “Mechanism of Holliday junction resolution by the human GEN1 protein,” *Genes Dev.*, vol. 24, no. 14, pp. 1559–1569, 2010.

[19] A. O’Donovan, A. A. Davies, J. G. Moggs, S. C. West, and R. D. Wood, “XPG endonuclease makes the 3’ Incision in human DNA nucleotide excision repair,” *Nature*, vol. 371, no. 6496, pp. 432–435, 1994.

[20] S. E. Tsutakawa *et al.*, “Human flap endonuclease structures, DNA double-base flipping, and a unified understanding of the FEN1 superfamily,” *Cell*, vol. 145, no. 2, pp. 198–211, 2011.

[21] M. Kucherlapati *et al.*, “Haploinsufficiency of Flap endonuclease (Fen1) leads to rapid tumor progression,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 99, no. 15, pp. 9924–9929, 2002.

[22] Y. Zhang *et al.*, “Upregulation of FEN1 is associated with the tumor progression and prognosis of hepatocellular carcinoma,” *Dis. Markers*, vol. 2020, pp. 1–17, 2020.

[23] Y. Matsuzaki, N. Adachi, and H. Koyama, “Vertebrate cells lacking FEN-1 endonuclease are viable but hypersensitive to methylating agents and,” vol. 30, no. 14, pp. 3273–3277, 2002.

[24] A. Serra-Cardona and Z. Zhang, “Replication-coupled nucleosome assembly as a passage of epigenetic

information and cell identity Albert," *Trends Biochem. Sci.*, vol. 43, no. 2, pp. 136–148, 2018.

[25] I. Bruck, P. Perez-Arnaiz, M. K. Colbert, and D. L. Kaplan, "Insights into the initiation of eukaryotic DNA replication," *Nucleus*, vol. 6, no. 6, pp. 449–454, 2015.

[26] M. O. Donnell, L. Langston, and B. Stillman, "Principles and Concepts of DNA Replication in Bacteria, Archaea, and Eukarya," *Perspect. Biol.*, vol. 5, pp. 1–13, 2013.

[27] P. M. J. Burgers, "Polymerase dynamics at the eukaryotic DNA replication fork," *J. Biol. Chem.*, vol. 284, no. 7, pp. 4041–4045, 2009.

[28] L. Zheng, H. Dai, J. Qiu, Q. Huang, and B. Shen, "Disruption of the FEN-1/PCNA Interaction Results in DNA Replication Defects, Pulmonary Hypoplasia, Pancytopenia, and Newborn Lethality in Mice," *Mol. Cell. Biol.*, vol. 27, no. 8, pp. 3176–3186, 2007.

[29] J. Fuchs, A. Cheblal, and S. M. Gasser, "Underappreciated Roles of DNA Polymerase δ in Replication Stress Survival," *Trends Genet.*, vol. 37, no. 5, pp. 476–487, 2021.

[30] P. Burgers, "It's all about flaps: Dna2 and checkpoint activation," *Cell Cycle*, vol. 10, no. 15, pp. 2417–2418, 2011.

[31] N. R. Jena, "DNA damage by reactive species: Mechanisms, mutation and repair," *J. Biosci.*, vol. 37, no. 3, pp. 503–517, 2012.

[32] J. Cadet and J. Richard Wagner, "DNA base damage by reactive oxygen species, oxidizing agents, and UV radiation," *Cold Spring Harb. Perspect. Biol.*, vol. 5, no. 2, pp. 1–18, 2013.

[33] R. Hakem, "DNA-damage repair; the good, the bad, and the ugly," *EMBO J.*, vol. 27, no. 4, pp. 589–605, 2008.

[34] L. Y. Li, Y. Di Guan, X. S. Chen, J. M. Yang, and Y. Cheng, "DNA Repair Pathways in Cancer Therapy and Resistance," *Front. Pharmacol.*, vol. 11, no. February, pp. 1–13, 2021.

[35] H. E. Krokan and M. Bjoras, "Base Excision Repair," *Cold Spring Harb Perspect Biol.*, vol. 5, no. 4, pp. 1–22, 2013.

[36] T. H. Lee and T. H. Kang, “DNA oxidation and excision repair pathways,” *Int. J. Mol. Sci.*, vol. 20, no. 23, 2019.

[37] Y.-J. Kim and D. M. Wilson III, “Overview of Base Excision Repair Biochemistry,” *Curr. Mol. Pharmacol.*, vol. 5, no. 1, pp. 3–13, 2011.

[38] S. S. Wallace, D. L. Murphy, and J. B. Sweasy, “Base Excision Repair and Cancer,” *Cancer Lett.*, vol. 23, no. 1, pp. 1–7, 2012.

[39] Y. Lin, W. A. Beard, D. D. Shock, R. Prasad, E. W. Hou, and S. H. Wilson, “DNA polymerase β and flap endonuclease 1 enzymatic specificities sustain DNA synthesis for long patch base excision repair,” *J. Biol. Chem.*, vol. 280, no. 5, pp. 3665–3674, 2005.

[40] S. E. Tsutakawa *et al.*, “Phosphate steering by Flap Endonuclease 1 promotes 5′-flap specificity and incision to prevent genome instability,” *Nat. Commun.*, vol. 8, no. May, pp. 1–14, 2017.

[41] S. Sakurai *et al.*, “Structural basis for recruitment of human flap endonuclease 1 to PCNA,” *EMBO J.*, vol. 24, no. 4, pp. 683–693, 2005.

[42] B. R. Chapados *et al.*, “Structural Basis for FEN-1 Substrate Specificity and PCNA-Mediated Activation in DNA Replication and Repair,” *Cell*, vol. 116, no. 1, pp. 39–50, 2004.

[43] F. Storici, G. Henneke, E. Ferrari, D. A. Gordenin, U. Hübscher, and M. A. Resnick, “The flexible loop of human FEN1 endonuclease is required for flap cleavage during DNA replication and repair,” *EMBO J.*, vol. 21, no. 21, pp. 5930–5942, 2002.

[44] D. Finger *et al.*, “The Wonders of Flap Endonucleases: Structure, Function, Mechanism and Regulation,” *Subcell. Biochem.*, vol. 62, pp. 259–79, 2012.

[45] S. E. Tsutakawa *et al.*, “Human flap endonuclease structures, DNA double-base flipping, and a unified understanding of the FEN1 superfamily,” *Cell*, vol. 145, no. 2, pp. 198–211, 2011.

[46] L. D. Finger *et al.*, “The 3′-flap pocket of human flap endonuclease 1 is critical for substrate binding and catalysis,” *J. Biol. Chem.*, vol. 284, no. 33, pp. 22184–22194, 2009.

[47] E. Friedrich-heineken and U. Hu, “The FEN1 extrahelical 3’-flap pocket is conserved from archaea to human and regulates DNA substrate specifcity,” *Nucleic Acids Res.*, vol. 32, no. 8, pp. 2520–2528, 2004.

[48] M. J. Thompson, V. J. B. Gotham, B. Ciani, and J. A. Grasby, “A conserved loop-wedge motif moderates reaction site search and recognition by FEN1,” *Nucleic Acids Res.*, vol. 46, no. 15, pp. 7858–7872, 2018.

[49] B. Dalhus *et al.*, “Structural insight into repair of alkylated DNA by a new superfamily of DNA glycosylases comprising HEAT-like repeats,” *Nucleic Acids Res.*, vol. 35, no. 7, pp. 2451–2459, 2007.

[50] Y. Shi, H. W. Hellinga, and L. S. Beese, “Interplay of catalysis, fidelity, threading, and processivity in the exo-and endonucleolytic reactions of human exonuclease i,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 114, no. 23, pp. 6010–6015, 2017.

[51] B. Song, S. M. Hamdan, and M. M. Hingorani, “Positioning the 5-flap junction in the active site controls the rate of flap endonuclease-1- catalyzed DNA cleavage,” *J. Biol. Chem.*, vol. 293, no. 13, pp. 4792–4804, 2018.

[52] T. Ceska, J. Sayers, G. Stier, and D. Suck, “A helical arch allowing single-stranded DNA to thread through TS 5’ -exonuclease,” *Nature*, vol. 382, no. July, pp. 90–93, 1996.

[53] C. J. Barnes, A. F. Wahl, B. Shen, M. S. Park, and R. A. Bambara, “Mechanism of tracking and cleavage of adduct-damaged DNA substrates by the mammalian 5’- to 3’-exonuclease/endonuclease RAD2 homologue 1 or flap endonuclease 1,” *J. Biol. Chem.*, vol. 271, no. 47, pp. 29624–29631, 1996.

[54] C. J. Bornarth, T. A. Ranalli, L. A. Henricksen, A. F. Wahl, and R. A. Bambara, “Effect of flap modifications on human FEN1 cleavage,” *Biochemistry*, vol. 38, no. 40, pp. 13347–13354, 1999.

[55] J. W. Gloor, L. Balakrishnan, and R. A. Bambara, “Flap endonuclease 1 mechanism analysis indicates flap base binding prior to threading,” *J. Biol. Chem.*, vol. 285, no. 45, pp. 34922–34931, 2010.

[56] N. Patel *et al.*, “Flap endonucleases pass 5’-flaps through a flexible arch using a disorder-thread-order mechanism to confer specificity for free 5’-ends,” vol. 40, no. 10, pp. 4507–4519, 2012.

[57] J. Grasby and L. Finger, “Unpairing and gating substrate duplex ends enforces FEN superfamily specificity,” *Trends Biochem. Sci.*, vol. 37, no. 2, pp. 74–84, 2012.

[58] I. Frouin, A. Montecucco, S. Spadari, and G. Maga, “DNA replication: A complex matter,” *EMBO Rep.*, vol. 4, no. 7, pp. 666–670, 2003.

[59] and K. C. M. Mhatre V. Ho, Ji-Ann Lee and 2013 Dien et al., “FEN1 Functions in Long Patch Base Excision Repair Under Conditions of Oxidative Stress in Vertebrate Cells,” *Mol. Cancer Res.*, vol. 23, no. 1, pp. 1–7, 2008.

[60] M. A. Sobhy, M. Tehseen, M. Takahashi, A. Bralić, A. De Biasio, and S. M. Hamdan, “Implementing fluorescence enhancement, quenching, and FRET for investigating flap endonuclease 1 enzymatic reaction at the single-molecule level,” *Comput. Struct. Biotechnol. J.*, vol. 19, pp. 4456–4471, 2021.

[61] M. A. Sobhy, L. I. Joudeh, X. Huang, M. Takahashi, and S. M. Hamdan, “Report Sequential and Multistep Substrate Interrogation Provides the Scaffold for Specificity in Human Flap Endonuclease 1,” *CellReports*, vol. 3, no. 6, pp. 1785–1794, 2013.

[62] R. B. Sekar and A. Periasamy, “Fluorescence resonance energy transfer (FRET) microscopy imaging of live cell protein localizations,” *J. Cell Biol.*, vol. 160, no. 5, pp. 629–633, 2001.

[63] D. K. Sasmal, L. E. Pulido, S. Kasal, and J. Huang, “Single-molecule fluorescence resonance energy transfer in molecular biology,” *Nanoscale*, vol. 8, no. 48, pp. 19928–19944, 2016.

[64] T. Ha, T. Enderle, D. F. Ogletree, D. S. Chemla, P. R. Selvin, and S. Weiss, “Probing the interaction between two single molecules: fluorescence resonance energy transfer between a single donor and a single acceptor.,” *Proc. Natl. Acad. Sci.*, vol. 93, pp. 6264–6268, 1996.

[65] T. Förster, “Transfer Mechanisms of Electronic Excitation,” *Radiat. Res. SUPPLEMEN*, vol. 2, pp. 326–339, 1959.

[66] F. Rashid *et al.*, “Single-molecule FRET unveils induced-fit mechanism for substrate selectivity in flap endonuclease 1,” *Elife*, vol. 6, pp. 1–23, 2017.

[67] S. I. Algasaier *et al.*, “DNA and protein requirements for substrate conformational changes necessary for human flap endonuclease-1-catalyzed reaction,” *J. Biol. Chem.*, vol. 291, no. 15, pp. 8258–8268, 2016.

[68] B. Song, S. M. Hamdan, and M. M. Hingorani, “Positioning the 5-flap junction in the active site controls the rate of flap endonuclease-1- catalyzed DNA cleavage,” *J. Biol. Chem.*, vol. 293, no. 13, pp. 4792–4804, 2018.

[69] S. I. Algasaier *et al.*, “DNA and protein requirements for substrate conformational changes necessary for human flap endonuclease-1-catalyzed reaction,” *J. Biol. Chem.*, vol. 291, no. 15, pp. 8258–8268, 2016.

[70] E. Friedrich-Heineken, G. Henneke, E. Ferrari, and U. Hübscher, “The acetylatable lysines of human Fen1 are important for endo- and exonuclease activities,” *J. Mol. Biol.*, vol. 328, no. 1, pp. 73–84, 2003.

[71] N. P. Johnson, W. A. Baase, and P. H. Von Hippel, “Low-energy circular dichroism of 2-aminopurine dinucleotide as a probe of local conformation of DNA and RNA,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 101, no. 10, pp. 3426–3431, 2004.

[72] I. A. Bennet *et al.*, “Regional conformational flexibility couples substrate specificity and scissile phosphate diester selectivity in human flap endonuclease 1,” *Nucleic Acids Res.*, vol. 46, no. 11, pp. 5618–5633, 2018.

[73] L. Balakrishnan and R. Bambara, “Flap Endonuclease 1,” *Annu. Rev. Biochem.*, vol. 2, no. 82, pp. 119–138, 2013.

[74] H. Xu *et al.*, “Structural basis of 5 flap recognition and protein–protein interactions of human flap endonuclease 1,” *Nucleic Acids Res.*, vol. 46, no. 21, pp. 11315–11325, 2018.

[75] L. D. Finger *et al.*, “Observation of unpaired substrate DNA in the flap endonuclease-1 active site,” *Nucleic Acids Res.*, vol. 41, no. 21, pp. 9839–9847, 2013.

[76] L. Zheng, J. Jia, L. D. Finger, Z. Guo, C. Zer, and B. Shen, “Functional regulation of FEN1 nuclease and its link to cancer,” *Nucleic Acids Res.*, vol. 39, no. 3, pp. 781–794, 2011.

[77] R. Liu, J. Qiu, L. D. Finger, L. Zheng, and B. Shen, “The DNA-protein interaction modes of FEN-1 with gap substrates and their implication in preventing duplication mutations,” *Nucleic Acids Res.*, vol. 34, no. 6, pp. 1772–1784, 2006.

[78] W. Yang, “Nucleases: Diversity of Structure, Function and Mechanism,” *Physiol. Behav.*, vol. 176, no. 1, pp. 139–148, 2011.

[79] A. Balian and F. J. Hernandez, “Nucleases as molecular targets for cancer diagnosis,” *Biomark. Res.*, vol. 9, no. 1, pp. 1–16, 2021.

[80] W. Wang, “Emergence of a DNA-damage response network consisting of Fanconi anaemia and BRCA proteins,” *Nat. Rev. Genet.*, vol. 8, pp. 735–748, 2007.

[81] J. Zhang and J. C. Walter, “Mechanism and regulation of incisions during DNA interstrand cross-link repair,” *DNA Repair (Amst).*, vol. 19, pp. 135–142, 2014.

[82] M. Ishiai, K. Sato, J. Tomida, H. Kitao, and H. Kurumizaka, “Activation of the FA pathway mediated by phosphorylation and ubiquitination,” *Mutat Res Fund Mol Mech Mutagen*, vol. 803–805, no. April, pp. 89–95, 2017.

[83] M. Ishiai, “Regulation of the Fanconi Anemia DNA Repair Pathway by Phosphorylation and Monoubiquintination,” *Genes (Basel).*, vol. 12, pp. 1763–1771, 2021.

[84] H. Jin and Y. Cho, “Structural and functional relationships of FAN1,” *DNA Repair (Amst).*, vol. 56, no. June, pp. 135–143, 2017.

[85] L. Qiuzhen, D. Kata, G. Tick, and H. Lajos, “Coordinated Cut and Bypass : Replication of Interstrand Crosslink-Containing DNA,” *Front. Cell Dev. Biol.*, vol. 9, no. June, pp. 1–8, 2021.

[86] W. S. Hoogenboom, R. A. C. M. Boonen, and P. Knipscheer, “The role of SLX4 and its associated nucleases in DNA interstrand crosslink repair,” *Nucleic Acids Res.*, vol. 47, no. 5, pp. 2377–2388, 2019.

[87] K. Kirschner and D. W. Melton, “Multiple Roles of the ERCC1-XPF Endonuclease in DNA Repair and Resistance to Anticancer Drugs,” vol. 3232, pp. 3223–3232, 2010.

[88] K. Kratz *et al.*, "Deficiency of FANCD2-Associated Nuclease KIAA1018/FAN1 Sensitizes Cells to Interstrand Crosslinking Agents," *Cell*, vol. 142, no. 1, pp. 77–88, 2010.

[89] W. et al Zhou, "FAN1 mutations cause karyomegalic interstitial nephritis, linking chronic kidney failure to defective DNA damage repair," *Nat. Genet.*, vol. 44, no. 8, pp. 910–915, 2013.

[90] S. Thongthip *et al.*, "Fan1 deficiency results in DNA interstrand cross-link repair defects, enhanced tissue karyomegaly, and organ dysfunction," *Genes Dev.*, vol. 30, no. 6, pp. 645–659, 2016.

[91] A. Porro *et al.*, "FAN1-MLH1 interaction affects repair of DNA interstrand cross-links and slipped-CAG/CTG repeats," *Sci. Adv.*, vol. 7, no. 31, pp. 1–13, 2021.

[92] R. Goold *et al.*, "FAN1 controls mismatch repair complex assembly via MLH1 retention to stabilize CAG repeat expansion in Huntington's disease," *Cell Rep.*, vol. 36, no. 9, 2021.

[93] B. Bonilla, S. Hengel, M. Grundy, and K. Bernstein, "RAD51 Gene Family Structur and Function," *Annu. Rev. Genet.*, vol. 23, no. 54, pp. 25–46, 2020.

[94] A. L. Deshmukh *et al.*, "FAN1, a DNA Repair Nuclease, as a Modifier of Repeat Expansion Disorders," *J. Huntingtons. Dis.*, vol. 10, no. 1, pp. 95–122, 2021.

[95] S. Bhandari *et al.*, "Karyomegalic nephropathy: An uncommon cause of progressive renal failure," *Nephrol. Dial. Transplant.*, vol. 17, no. 11, pp. 1914–1920, 2002.

[96] C. Lachaud *et al.*, "Karyomegalic interstitial nephritis and DNA damageinduced polyploidy in fan1 nuclease-defective knock-in mice," *Genes Dev.*, vol. 30, no. 6, pp. 639–644, 2016.

[97] R. Goold *et al.*, "FAN1 modifies Huntington's disease progression by stabilizing the expanded HTT CAG repeat," *Hum. Mol. Genet.*, vol. 28, no. 4, pp. 650–661, 2019.

[98] X.-N. Zhao and K. Usdin, "FAN1 protects against repeat expansions in a Fragile X mouse model," *DNA Repair (Amst).*, vol. 69, pp. 1–5, 2019.

[99] A. Forsingdal, K. Fejgin, V. Nielsen, T. Werge, and J. Nielsen, "15Q13.3 Homozygous Knockout Mouse Model Display Epilepsy-, Autism-and Schizophrenia-Related Phenotypes," *Transl. Psychiatry*, vol. 6,

no. 7, pp. e860-9, 2016.

[100] A. Smogorzewska *et al.*, “A genetic screen identifies FAN1, a Fanconi anemia associated nuclease necessary for DNA interstrand crosslink repair,” vol. 39, no. 1, pp. 36–47, 2011.

[101] S. Pennell *et al.*, “FAN1 activity on asymmetric repair intermediates is mediated by an atypical monomeric virus-type replication-repair nuclease domain,” *Cell Rep.*, vol. 8, no. 1, pp. 84–93, 2014.

[102] R. Wang *et al.*, “Mechanism of DNA Interstrand Crosslink Processing by Repair Nuclease FAN1,” *Science (80-.).*, vol. 346, no. 6213, pp. 1127–1130, 2015.

[103] J. M. Daley, W. A. Gaines, Y. Kwon, and P. Sung, “Regulation of DNA Pairing in Homologous Recombination,” *Perspect. Biol.*, vol. 6, no. 11, pp. 1–15, 2014.

[104] H. Jin, U. Roy, G. Lee, and Y. Cho, “Structural mechanism of DNA interstrand cross-link unhooking by the bacterial FAN1 nuclease,” vol. 293, pp. 6482–6496, 2018.

[105] Q. Zhao, X. Xue, S. Longerich, P. Sung, and Y. Xiong, “Structural insights into 5' flap DNA unwinding and incision by the human FAN1 dimer,” *Nat. Commun.*, vol. 5, pp. 1–9, 2014.

[106] G. H. Gwon *et al.*, “Crystal structure of a fanconi anemiaassociated nuclease homolog bound to 5’ flap DNA: Basis of interstrand cross-link repair by FAN1,” *Genes Dev.*, vol. 28, no. 20, pp. 2276–2290, 2014.

[107] C. Lundin *et al.*, “Identification of KIAA1018 / FAN1 , a DNA Repair Nuclease Recruited to DNA Damage by Monoubiquitinated FANCD2,” vol. 1, pp. 65–76, 2010.

[108] J. Kosinski, M. Feder, and J. M. Bujnicki, “The PD-(D/E)XK superfamily revisited: Identification of new members among proteins involved in DNA metabolism and functional predictions for domains of (hitherto) unknown function,” *BMC Bioinformatics*, vol. 6, pp. 1–13, 2005.

[109] T. Nishino, K. Komori, D. Tsuchiya, Y. Ishino, and K. Morikawa, “Crystal structure of the archaeal Holliday junction resolvase Hjc and implications for DNA recognition,” *Structure*, vol. 9, no. 3, pp. 197–204, 2001.

[110] C. Sissi and M. Palumbo, “Effects of magnesium and related divalent metal ions in topoisomerase

structure and function," *Nucleic Acids Res.*, vol. 37, no. 3, pp. 702–711, 2009.

[111] D. T. Dang, "Molecular Approaches to Protein Dimerization: Opportunities for Supramolecular Chemistry," *Front. Chem.*, vol. 10, no. February, pp. 1–12, 2022.

[112] T. Rao, S. Longerich, W. Zhao, H. Aihara, P. Sung, and Y. Xiong, "Importance of homo-dimerization of Fanconi-associated nuclease 1 in DNA flap cleavage," *DNA Repair (Amst).*, vol. 64, no. February, pp. 53–58, 2018.

[113] G. H. Gwon *et al.*, "Crystal structures of the structure-selective nuclease Mus81-Eme1 bound to flap DNA substrates," *EMBO J.*, vol. 33, no. 9, pp. 1061–1072, 2014.

[114] R. S. Williams *et al.*, "Mre11 Dimers Coordinate DNA End Bridging and Nuclease Processing in Double-Strand-Break Repair," *Cell*, vol. 135, no. 1, pp. 97–109, 2008.

[115] J. Orans *et al.*, "Structures of human exonuclease 1 DNA complexes suggest a unified mechanism for nuclease family," *Cell*, vol. 145, no. 2, pp. 212–223, 2011.

[116] H. I. Kao, L. A. Henricksen, Y. Liu, and R. A. Bambara, "Cleavage specificity of Saccharomyces cerevisiae flap endonuclease 1 suggests a double-flap structure as the cellular substrate," *J. Biol. Chem.*, vol. 277, no. 17, pp. 14379–14389, 2002.

[117] L. He *et al.*, "FEN1 promotes tumor progression and confers cisplatin resistance in non-small-cell lung cancer," *Mol. Oncol.*, vol. 11, no. 6, pp. 640–654, 2017.

[118] X. Xin *et al.*, "Inhibition of FEN1 Increases Arsenic Trioxide-Induced ROS Accumulation and Cell Death: Novel Therapeutic Potential for Triple Negative Breast Cancer," *Front. Oncol.*, vol. 10, no. April, pp. 1–11, 2020.

[119] C. Sun, Y. Li, E. A. Yates, and D. G. Fernig, "SimpleDSFviewer: A tool to analyze and view differential scanning fluorimetry data for characterizing protein thermal stability and interactions," *Protein Sci.*, vol. 29, no. 1, pp. 19–27, 2020.

[120] J. J. Harrington and M. R. Lieber, "The characterization of a mammalian DNA structure specific

endonuclease," *EMBO J.*, vol. 13, no. 5, pp. 1235–1246, 1994.

[121] Y. Liu, H.-I. Kao, and R. A. Bambara, "Flap Endonuclease 1: A Central Component of DNA Metabolism," *Annu. Rev. Biochem.*, vol. 73, no. 1, pp. 589–615, 2004.

[122] T. D. Craggs *et al.*, "Substrate conformational dynamics facilitate structure-specific recognition of gapped DNA by DNA polymerase," *Nucleic Acids Res.*, vol. 47, no. 20, pp. 10788–10800, 2019.

[123] I. A. Bennet *et al.*, "Regional conformational flexibility couples substrate specificity and scissile phosphate diester selectivity in human flap endonuclease 1," *PLoS Comput. Biol.*, vol. 4, no. 9, 2008.

[124] A. Sassa, W. A. Beard, D. D. Shock, and S. H. Wilson, "Steady-state , Pre-steady-state , and Single-turnover Kinetic Measurement for DNA Glycosylase Activity," no. August, pp. 1–9, 2013.

[125] J. R. Lorsch, "Practical steady-state enzyme kinetics," *Methods in Enzymology*, vol. 536. pp. 3–15, 2014.

[126] L. D. Finger *et al.*, "The 3'-Flap Pocket of Human Flap Endonuclease 1 Is Critical for Substrate Binding and Catalysis," vol. 284, no. 33, pp. 22184–22194, 2009.

[127] J. C. Exell *et al.*, "Cellular Active N-Hydroxyurea FEN1 Inhibitors Block Substrate Entry to the Active Site," *Nat. Chem. Biol.*, vol. 12, no. 10, pp. 815–821, 2016.

[128] A. C. Jones and R. K. Neely, "2-aminopurine as a fluorescent probe of DNA conformation and the DNA-enzyme interface," *Q. Rev. Biophys.*, vol. 48, no. 2, pp. 244–279, 2015.

[129] P. Rai, T. D. Cole, E. Thompson, D. P. Millar, and S. Linn, "Steady-state and time-resolved fluorescence studies indicate an unusual conformation of 2-aminopurine within ATAT and TATA duplex DNA sequences," *Nucleic Acids Res.*, vol. 31, no. 9, pp. 2323–2332, 2003.

[130] L. Ma, X. Wu, G. G. Wilson, A. C. Jones, and D. T. F. Dryden, "Time-resolved fluorescence of 2-aminopurine in DNA duplexes in the presence of the EcoP15I Type III restriction-modification enzyme," *Biochem. Biophys. Res. Commun.*, vol. 449, no. 1, pp. 120–125, 2014.

[131] S. E. Tsutakawa *et al.*, "Human Flap Endonuclease Structures, DNA Double Base Flipping and a Unified

Understanding of the FEN1 Superfamily," *Cell*, vol. 145, no. 2, pp. 198–211, 2011.

[132] D. J. Hosfield, C. D. Mol, B. Shen, and J. A. Tainer, "Structure of the DNA Repair and Replication Endonuclease and Exonuclease FEN-1 : Coupling DNA and PCNA Binding to FEN-1 Activity," *Cell Press*, vol. 95, pp. 135–146, 1998.

[133] S. Shah, P. Dunten, A. Stiteler, C. K. Park, and N. C. Horton, "Structure and Specificity of FEN-1 from Methanopyrus kandleri," *Proteins*, vol. 83, no. 1, pp. 1–7, 2015.

[134] P. M. Dehé and P. H. L. Gaillard, "Control of structure-specific endonucleases to maintain genome stability," *Nat. Rev. Mol. Cell Biol.*, vol. 18, no. 5, pp. 315–330, 2017.

[135] A. A. Umar, A. Abdulhamid, A. I. Bagudo, and I. M. Magami, "Biochemical characterization of human exonuclease1 (hExo1)," *J. King Saud Univ. - Sci.*, vol. 32, no. 2, pp. 1659–1663, 2020.

[136] A. Cornish-Bowden, "One hundred years of Michaelis–Menten kinetics," *Perspect. Sci.*, vol. 4, pp. 3–9, 2015.

[137] J. Qiu, R. Liu, B. R. Chapados, M. Sherman, J. A. Tainer, and B. Shen, "Interaction interface of human flap endonuclease-1 with its DNA substrates," *J. Biol. Chem.*, vol. 279, no. 23, pp. 24394–24402, 2004.

[138] H. Pelletier, M. R. Sawaya, W. Wolfle, S. H. Wilson, and J. Kraut, "Crystal structures of human DNA polymerase β complexed with DNA: Implications for catalytic mechanism, processivity, and fidelity," *Biochemistry*, vol. 35, no. 39, pp. 12742–12761, 1996.

[139] S. I. Algasaier, L. D. Finger, I. A. Bennet, and J. A. Grasby, "Flap Endonuclease 1 Mutations A159V and E160D Cause Genomic Instability by Slowing Reaction on Double-Flap Substrates," *Biochemistry*, vol. 57, no. 50, pp. 6838–6847, 2018.

[140] Y. Shi, H. W. Hellinga, and L. S. Beese, "Interplay of catalysis, fidelity, threading, and processivity in the exo-and endonucleolytic reactions of human exonuclease I," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 114, no. 23, pp. 6010–6015, 2017.

[141] J. R. Widom, N. P. Johnson, P. H. Von Hippel, and A. H. Marcus, "Solution conformation of 2-

aminopurine dinucleotide determined by ultraviolet two-dimensional fluorescence spectroscopy," *New J. Phys.*, vol. 15, 2013.

[142] S. Raran-Kurussi and D. Waugh, "Unrelated solubility-enhancing fusion partners MBP and NusA utilize a similar mode of action," *Biotechnol. Bioeng.*, vol. 111, no. 12, pp. 2407–2411, 2014.

[143] A. S. Pina, C. Lowe, and A. C. Roque, "Challenges and opportunities in the purification of recombinant tagged proteins," *Biotechnol. Adv.*, vol. 32, no. 2014, pp. 366–381, 2013.

[144] R. J. Giannone and A. B. Dykstra, *Protein Affinity Tags*. New York: Humana Press, 2014.

[145] D. S. Waugh, "An overview of enzymatic reagents for the removal of affinity tags," *Protein Expr. Purif.*, vol. 80, no. January, pp. 283–293, 2011.

[146] T. Rao, S. Longerich, W. Zhao, H. Aihara, P. Sung, and Y. Xiong, "Importance of homo-dimerization of Fanconi-associated nuclease 1 in DNA flap cleavage," *DNA Repair (Amst).*, vol. 64, no. December 2017, pp. 53–58, 2018.

[147] X. Shi and D. Jarvis, "Protein N-Glycosylation in the Baculovirus-Insect Cell System," *Curr. Drug Targets*, vol. 8, no. 10, pp. 1116–1125, 2007.

[148] I. M. Muñoz *et al.*, "Coordination of Structure-Specific Nucleases by Human SLX4/BTBD12 Is Required for DNA Repair," *Mol. Cell*, vol. 35, no. 1, pp. 116–127, 2009.

[149] X. Xue *et al.*, "Production of Authentic SARS-CoV Mpro with Enhanced Activity: Application as a Novel Tag-cleavage Endopeptidase for Protein Overproduction," *J. Mol. Biol.*, vol. 366, no. 3, pp. 965–975, 2007.

[150] S. Boivin, S. Kozak, and R. Meijers, "Optimization of protein purification and characterization using Thermofluor screens," *Protein Expr. Purif.*, vol. 91, no. 2, pp. 192–206, 2013.

[151] K. Huynh and C. Partch, "Current Protocols in Protein Science: Analysis of protein stability and ligand interactions by thermal shift assay," *Curr. Protoc. Protein Sci.*, vol. 79, pp. 28.9.1-28.9.14, 2016.

[152] S. Yarmoluk, V. Kovalska, and K. Volkova, "Optimized Dyes for Protein and Nucleic Acid Detection,"

Adavanced Fluoresc. Reporters Chem. Biol. III, vol. 133, pp. 161–199, 2011.

[153] S. Raran-Kurussi, S. Cherry, D. Zhang, and D. Waugh, "Removal of Affinity Tags with TEV Protease," *Methods Mol. Biol.*, vol. 1586, pp. 33–43, 2017.

[154] J. T. McCue, *Theory and Use of Hydrophobic Interaction Chromatography in Protein Purification Applications*, 1st ed., vol. 463, no. C. Elsevier Inc., 2009.

[155] T. Rao, S. Longerich, W. Zhao, H. Aihara, P. Sung, and Y. Xiong, "Importance of homo-dimerization of Fanconi-associated nuclease 1 in DNA flap cleavage," *DNA Repair (Amst).*, vol. 64, pp. 53–58, 2018.

[156] C. L. Chang, H. N. Shou, and J. C. Juang, "DNA interstrand crosslink repair and cancer," *Int. J. Control Theory Appl.*, vol. 6, no. 1, pp. 1–14, 2013.

[157] F. R. Tsay *et al.*, "Generic anion-exchange chromatography method for analytical and preparative separation of nucleotides in the development and manufacture of drug substances," *J. Chromatogr. A*, vol. 1587, pp. 129–135, 2019.

Chapter 8 Appendices

8.1 pET28b WT human FEN1 plasmid

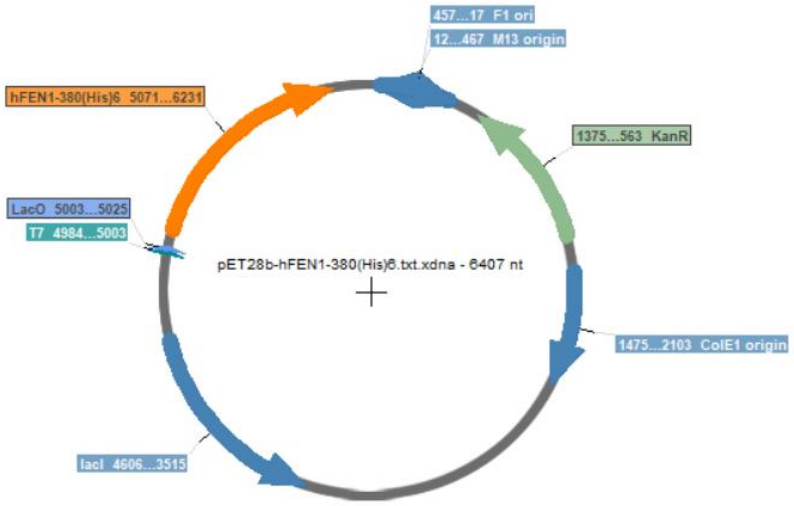


Figure 8.1 WT human FEN1 plasmid construct
Plasmid construct of WT FEN1 (human). The FAN1 gene and DNA sequence encoding the his-tag is shown in orange; Kanamycin resistance gene shown in green.

8.2 Agarose gel electrophoresis of FEN1 and FAN1 constructs

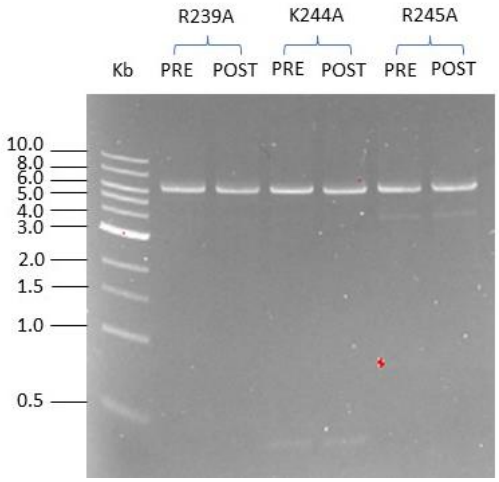


Figure 8.2 H2TH plasmid production and concentration analysis
Agarose gel electrophoresis of H2TH mutants pre and post DPN1 digestion prior to sequencing. A 1 Kb molecular maker was loaded in the left most lane within the gel. Gels were stained with SybrSafe stain.

8.3 WT FEN1 and H2TH mutant Michaelis Menton graphs

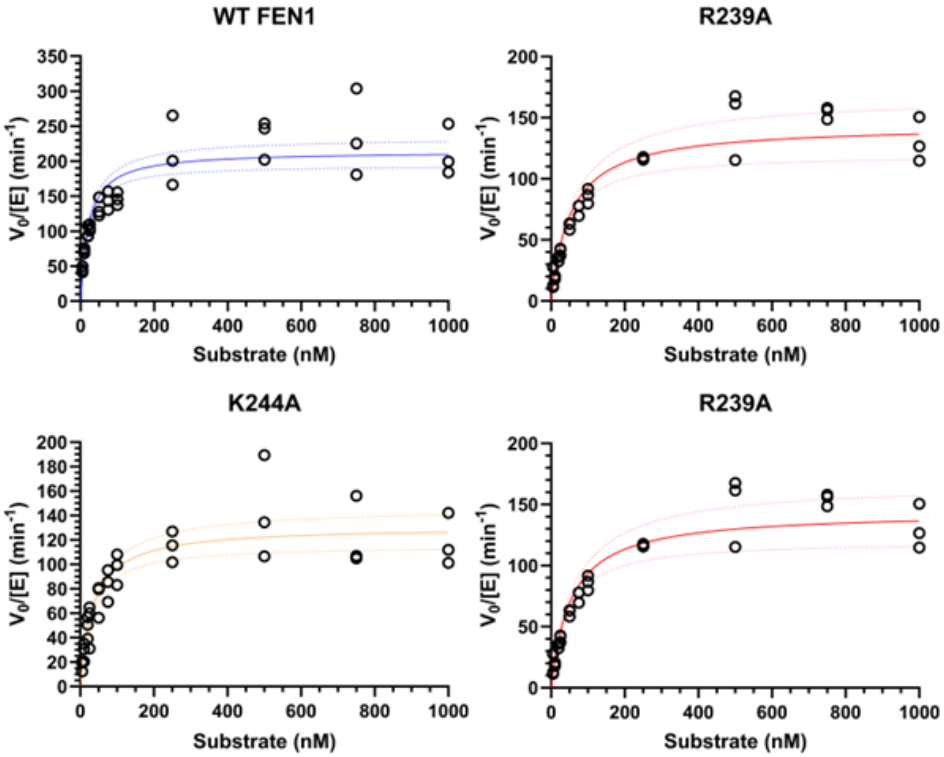


Figure 8.3 WT FEN1 and H2TH mutant Michaelis Menton graphs
Michaelis Menton graphs of WT FEN1, R239A, K244A and R245A displayed using an x-axis (n=3)

8.4 FEN1 and FAN1 sequencing

8.4.1 H2TH FEN1 mutant construct sequencing

```
Alignment of Sequence_1:  [hFEN1_mRNA.xdna] with  Sequence_2: [pET_R239A_F-T7.seq]

Similarity : 1087/1263 (86.06 %)

                                                    M  G  I  Q  G
Seq_1  1     ---------------------------------------------atgggaattcaaggc  15
                                                         ||||||||||||||||
Seq_2  1     CCCCTCTAGAATAATTTTTGTTTAACTTTAAGAAGGAGATATACCATGGGAATTCAAGGC  60
             P  L  *  N  N  F  C  L  T  L  R  R  R  Y  T  M  G  I  Q  G


             L  A  K  L  I  A  D  V  A  P  S  A  I  R  E  N  D  I  K  S
Seq_1  16    ctggccaaactaattgctgatgtggcccccagtgccatccgggagaatgacatcaagagc  75
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  61    CTGGCCAAACTAATTGCTGATGTGGCCCCCAGTGCCATCCGGGAGAATGACATCAAGAGC  120
             L  A  K  L  I  A  D  V  A  P  S  A  I  R  E  N  D  I  K  S


             Y  F  G  R  K  V  A  I  D  A  S  M  S  I  Y  Q  F  L  I  A
Seq_1  76    tactttggccgtaaggtggccattgatgcctctatgagcatttatcagttcctgattgct  135
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  121   TACTTTGGCCGTAAGGTGGCCATTGATGCCTCTATGAGCATTTATCAGTTCCTGATTGCT  180
             Y  F  G  R  K  V  A  I  D  A  S  M  S  I  Y  Q  F  L  I  A


             V  R  Q  G  G  D  V  L  Q  N  E  E  G  E  T  T  S  H  L  M
Seq_1  136   gttcgccagggtggggatgtgctgcagaatgaggagggtgagaccaccagccacctgatg  195
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  181   GTTCGCCAGGGTGGGGATGTGCTGCAGAATGAGGAGGGTGAGACCACCAGCCACCTGATG  240
             V  R  Q  G  G  D  V  L  Q  N  E  E  G  E  T  T  S  H  L  M


             G  M  F  Y  R  T  I  R  M  M  E  N  G  I  K  P  V  Y  V  F
Seq_1  196   ggcatgttctaccgcaccattcgcatgatggagaacggcatcaagcccgtgtatgtcttt  255
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  241   GGCATGTTCTACCGCACCATTCGCATGATGGAGAACGGCATCAAGCCCGTGTATGTCTTT  300
             G  M  F  Y  R  T  I  R  M  M  E  N  G  I  K  P  V  Y  V  F
```

```
                D  G  K  P  P  Q  L  K  S  G  E  L  A  K  R  S  E  R  R  A
Seq_1  256   gatggcaagccgccacagctcaagtcaggcgagctggccaaacgcagtgagcggcgggct  315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  301   GATGGCAAGCCGCCACAGCTCAAGTCAGGCGAGCTGGCCAAACGCAGTGAGCGGCGGGCT  360
                D  G  K  P  P  Q  L  K  S  G  E  L  A  K  R  S  E  R  R  A

                E  A  E  K  Q  L  Q  Q  A  Q  A  A  G  A  E  Q  E  V  E  K
Seq_1  316   gaggcagagaagcagctgcagcaggctcaggctgctggggccgagcaggaggtggaaaaa  375
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  361   GAGGCAGAGAAGCAGCTGCAGCAGGCTCAGGCTGCTGGGGCCGAGCAGGAGGTGGAAAAA  420
                E  A  E  K  Q  L  Q  Q  A  Q  A  A  G  A  E  Q  E  V  E  K

                F  T  K  R  L  V  K  V  T  K  Q  H  N  D  E  C  K  H  L  L
Seq_1  376   ttcactaagcggctggtgaaggtcactaagcagcacaatgatgagtgcaaacatctgctg  435
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  421   TTCACTAAGCGGCTGGTGAAGGTCACTAAGCAGCACAATGATGAGTGCAAACATCTGCTG  480
                F  T  K  R  L  V  K  V  T  K  Q  H  N  D  E  C  K  H  L  L

                S  L  M  G  I  P  Y  L  D  A  P  S  E  A  E  A  S  C  A  A
Seq_1  436   agcctcatgggcatcccttatcttgatgcacccagtgaggcagaggccagctgtgctgcc  495
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  481   AGCCTCATGGGCATCCCTTATCTTGATGCACCCAGTGAGGCAGAGGCCAGCTGTGCTGCC  540
                S  L  M  G  I  P  Y  L  D  A  P  S  E  A  E  A  S  C  A  A

                L  V  K  A  G  K  V  Y  A  A  A  T  E  D  M  D  C  L  T  F
Seq_1  496   ctggtgaaggctggcaaagtctatgctgcggctaccgaggacatggactgcctcaccttc  555
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  541   CTGGTGAAGGCTGGCAAAGTCTATGCTGCGGCTACCGAGGACATGGACTGCCTCACCTTC  600
                L  V  K  A  G  K  V  Y  A  A  A  T  E  D  M  D  C  L  T  F

                G  S  P  V  L  M  R  H  L  T  A  S  E  A  K  K  L  P  I  Q
Seq_1  556   ggcagccctgtgctaatgcgacacctgactgccagtgaagccaaaaagctgccaatccag  615
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  601   GGCAGCCCTGTGCTAATGCGACACCTGACTGCCAGTGAAGCCAAAAAGCTGCCAATCCAG  660
                G  S  P  V  L  M  R  H  L  T  A  S  E  A  K  K  L  P  I  Q
```

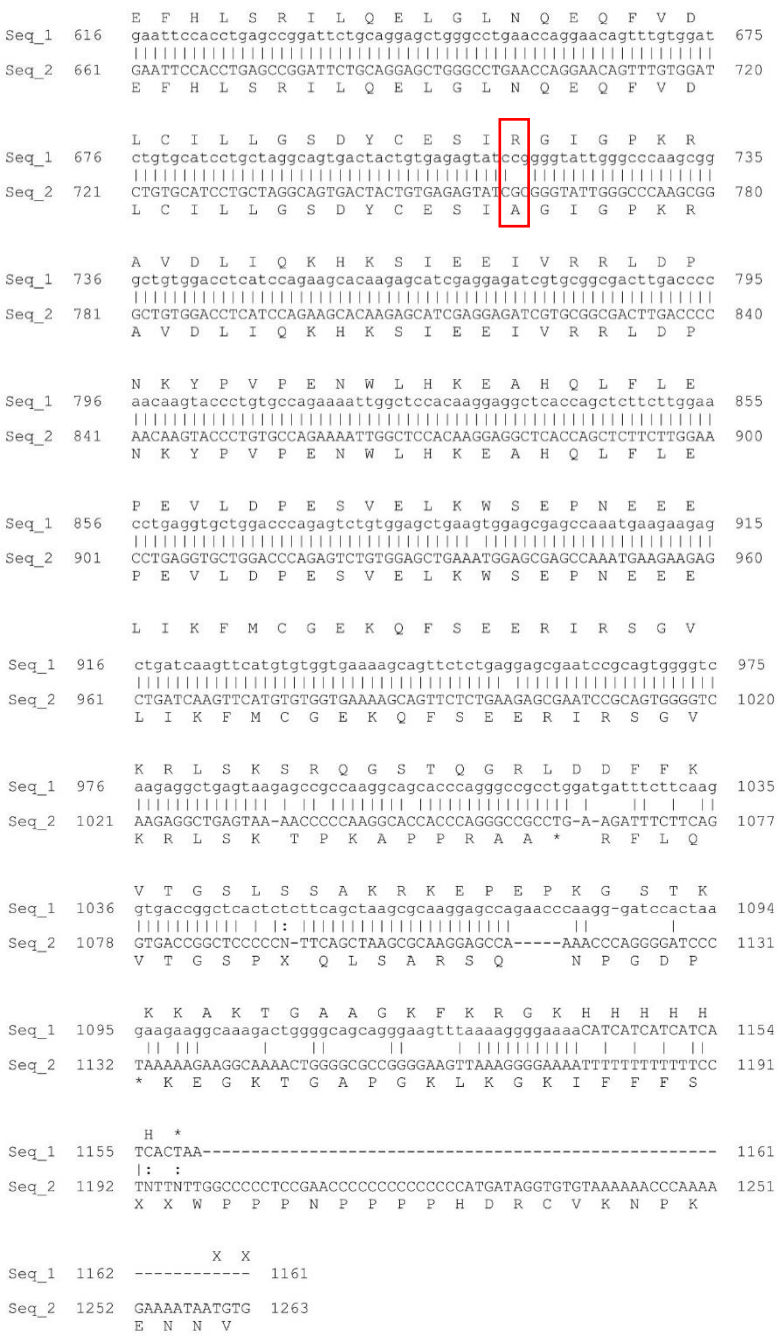


Figure 8.4 R239A forward primer sequencing

Forward primer sequencing of R239A with the alanine substitution highlighted by a red rectangle.

Alignment of Sequence_1: [hFEN1_mRNA.xdna] with Sequence_2: [pET_R239A_R-T7-Term.seq]

Similarity : 1028/1199 (85.74 %)

```
            M  G  I  Q  G  L  A  K  L  I  A  D   V  A  P  S  A  I  R  E
Seq_1  1     atgggaattcaaggcctggccaaactaattgctgat-gtggcccccagtgccatccgggga  59
                                 |  |   | ::|::: || :  | |   ||    ||| ||
Seq_2  1199  ------------------G-CCCAANNANNNATGTNNATTGGT-CCC-CCCCAGCCCTTC  1161
                               A   Q  X  X  M  X  I  G   P   P  Q  P  F

             N  D  I  K  S  Y  F  G  R  K  V  A  I  D  A  S  M  S  I  Y
Seq_1  60    gaatgacatcaagagctactttggccgtaaggtggccattgatgcctctatgagcattta  119
                 |  |          :::|:   | | :  ||        |    || |  |    |
Seq_2  1160  CGGGGGAAAGCCTCCTAGNNNTNTTTGGACNTAGGTGGCCCTTAATCCTTTTTGGGCCTT  1101
             R  G  K  A  S  *  X  X  L  D  X  G  G  P  *  S  F  L  G  L

             Q  F  L  I  A  V  R  Q  G  G  D  V  L  Q  N  E  E  G  E  T
Seq_1  120   tcagttcctgattgctgttcgccagggtggggatgtgctgcagaatgaggagggtgagac  179
             |                 ||  ||   |  |||   |    |  :|    |  |||| |:|
Seq_2  1100  TTTCAGTTCCGGATTACTTTTCCCCAGGTGGGGAATTGCTCCCNAATGAGGAGGTGGGNC  1041
             F  Q  F  R  I  T  F  P  Q  V  G  N  C  S  X  M  R  R  W  X

             T  S  H  L  M  G  M  F  Y  R  T  I  R  M  M  E  N  G  I  K
Seq_1  180   caccagccacctgatgggcatgttctaccgcaccattcgcatgatggagaacggcatcaa  239
             | |||:|| |||||||||||||| | |||| || |||||||||||||||||| ||||
Seq_2  1040  CCCCANCCCCCTGATGGGCATGTTTTCCGCCCC-TTCGCATGATGGAGAACGGCCTCAA  982
             P  X  P  P  D  G  H  V  F  P  P  L   R  M  M  E  N  G  L  K

             P  V  Y  V  F  D  G  K  P  P  Q  L  K  S  G  E  L  A  K  R
Seq_1  240   gcccgtgtatgtctttgatggcaagccgccacagctcaagtcaggcgagctggccaaacg  299
             |||||| ||||| |||||||||||||||||||| |||| ||||||||||||||||||||
Seq_2  981   GCCCGTTTATGTTTTTGATGGCAAGCCGCCACAGTTCAATTCAGGCGAGCTGGCCAAACG  922
              P  V  Y  V  F  D  G  K  P  P  Q  F  N  S  G  E  L  A  K  R

             S  E  R  R  A  E  A  E  K  Q  L  Q  Q  A  Q  A  A  G  A  E
Seq_1  300   cagtgagcggcgggctgaggcagagaagcagctgcagcaggctcaggctgctggggccga  359
             || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  921   CATTGAGCGGCGGGCTGAGGCAGAGAAGCAGCTGCAGCAGGCTCAGGCTGCTGGGGCCGA  862
              I  E  R  R  A  E  A  E  K  Q  L  Q  Q  A  Q  A  A  G  A  E

             Q  E  V  E  K  F  T  K  R  L  V  K  V  T  K  Q  H  N  D  E
Seq_1  360   gcaggaggtggaaaaattcactaagcggctggtgaaggtcactaagcagcacaatgatga  419
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  861   GCAGGAGGTGGAAAAATTCACTAAGCGGCTGGTGAAGGTCACTAAGCAGCACAATGATGA  802
             Q  E  V  E  K  F  T  K  R  L  V  K  V  T  K  Q  H  N  D  E

             C  K  H  L  L  S  L  M  G  I  P  Y  L  D  A  P  S  E  A  E
Seq_1  420   gtgcaaacatctgctgagcctcatgggcatcccttatcttgatgcacccagtgaggcaga  479
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  801   GTGCAAACATCTGCTGAGCCTCATGGGCATCCCTTATCTTGATGCACCCAGTGAGGCAGA  742
             C  K  H  L  L  S  L  M  G  I  P  Y  L  D  A  P  S  E  A  E

             A  S  C  A  A  L  V  K  A  G  K  V  Y  A  A  A  T  E  D  M
Seq_1  480   ggccagctgtgctgccctggtgaaggctggcaaagtctatgctgcggctaccgaggacat  539
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  741   GGCCAGCTGTGCTGCCCTGGTGAAGGCTGGCAAAGTCTATGCTGCGGCTACCGAGGACAT  682
             A  S  C  A  A  L  V  K  A  G  K  V  Y  A  A  A  T  E  D  M

             D  C  L  T  F  G  S  P  V  L  M  R  H  L  T  A  S  E  A  K
Seq_1  540   ggactgcctcaccttcggcagccctgtgctaatgcgacacctgactgccagtgaagccaa  599
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  681   GGACTGCCTCACCTTCGGCAGCCCTGTGCTAATGCGACACCTGACTGCCAGTGAAGCCAA  622
             D  C  L  T  F  G  S  P  V  L  M  R  H  L  T  A  S  E  A  K

             K  L  P  I  Q  E  F  H  L  S  R  I  L  Q  E  L  G  L  N  Q
Seq_1  600   aaagctgccaatccaggaattccacctgagccggattctgcaggagctgggcctgaacca  659
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  621   AAAGCTGCCAATCCAGGAATTCCACCTGAGCCGGATTCTGCAGGAGCTGGGCCTGAACCA  562
             K  L  P  I  Q  E  F  H  L  S  R  I  L  Q  E  L  G  L  N  Q
```

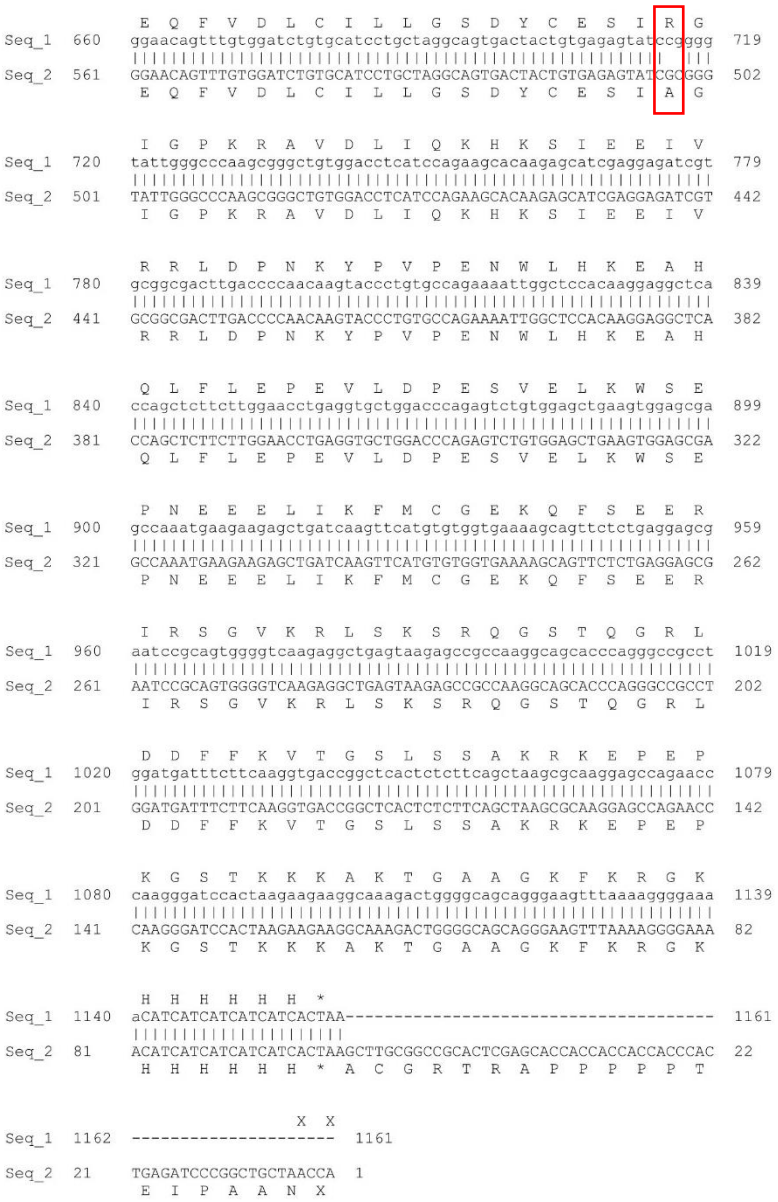


Figure 8.5 R239A reverse primer sequencing

Reverse primer sequencing of R239A with the alanine substitution highlighted by a red rectangle.

```
Alignment of Sequence_1:  [hFEN1_mRNA.xdna] with  Sequence_2: [K244A-F-T7.seq]

Similarity : 1100/1278 (86.07 %)

                                                    M  G  I  Q  G
Seq_1  1     ------------------------------------------------atgggaattcaagg  14
                                                             ||||||||||||||
Seq_2  1     TCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGGAATTCAAGG  60
              P  L  *  K  *  F  C  L  T  L  R  R  R  Y  T  M  G  I  Q  G

              L  A  K  L  I  A  D  V  A  P  S  A  I  R  E  N  D  I  K  S
Seq_1  15    cctggccaaactaattgctgatgtggcccccagtgccatccgggagaatgacatcaagag  74
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  61    CCTGGCCAAACTAATTGCTGATGTGGCCCCCAGTGCCATCCGGGAGAATGACATCAAGAG  120
              L  A  K  L  I  A  D  V  A  P  S  A  I  R  E  N  D  I  K  S

              Y  F  G  R  K  V  A  I  D  A  S  M  S  I  Y  Q  F  L  I  A
Seq_1  75    ctactttggccgtaaggtggccattgatgcctctatgagcatttatcagttcctgattgc  134
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  121   CTACTTTGGCCGTAAGGTGGCCATTGATGCCTCTATGAGCATTTATCAGTTCCTGATTGC  180
              Y  F  G  R  K  V  A  I  D  A  S  M  S  I  Y  Q  F  L  I  A

              V  R  Q  G  G  D  V  L  Q  N  E  E  G  E  T  T  S  H  L  M
Seq_1  135   tgttcgccagggtggggatgtgctgcagaatgaggagggtgagaccaccagccacctgat  194
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  181   TGTTCGCCAGGGTGGGGATGTGCTGCAGAATGAGGAGGGTGAGACCACCAGCCACCTGAT  240
              V  R  Q  G  G  D  V  L  Q  N  E  E  G  E  T  T  S  H  L  M

              G  M  F  Y  R  T  I  R  M  M  E  N  G  I  K  P  V  Y  V  F
Seq_1  195   gggcatgtttctaccgcaccattcgcatgatggagaacggcatcaagcccgtgtatgtctt  254
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  241   GGGCATGTTCTACCGCACCATTCGCATGATGGAGAACGGCATCAAGCCCGTGTATGTCTT  300
              G  M  F  Y  R  T  I  R  M  M  E  N  G  I  K  P  V  Y  V  F

              D  G  K  P  P  Q  L  K  S  G  E  L  A  K  R  S  E  R  R  A
Seq_1  255   tgatggcaagccgccacagctcaagtcaggcgagctggccaaacgcagtgagcggcgggc  314
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  301   TGATGGCAAGCCGCCACAGCTCAAGTCAGGCGAGCTGGCCAAACGCAGTGAGCGGCGGGC  360
              D  G  K  P  P  Q  L  K  S  G  E  L  A  K  R  S  E  R  R  A

              E  A  E  K  Q  L  Q  Q  A  Q  A  A  G  A  E  Q  E  V  E  K
Seq_1  315   tgaggcagagaagcagctgcagcaggctcaggctgctggggccgagcaggaggtggaaaa  374
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  361   TGAGGCAGAGAAGCAGCTGCAGCAGGCTCAGGCTGCTGGGGCCGAGCAGGAGGTGGAAAA  420
              E  A  E  K  Q  L  Q  Q  A  Q  A  A  G  A  E  Q  E  V  E  K

              F  T  K  R  L  V  K  V  T  K  Q  H  N  D  E  C  K  H  L  L
Seq_1  375   attcactaagcggctggtgaaggtcactaagcagcacaatgatgagtgcaaacatctgct  434
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  421   ATTCACTAAGCGGCTGGTGAAGGTCACTAAGCAGCACAATGATGAGTGCAAACATCTGCT  480
              F  T  K  R  L  V  K  V  T  K  Q  H  N  D  E  C  K  H  L  L

              S  L  M  G  I  P  Y  L  D  A  P  S  E  A  E  A  S  C  A  A
Seq_1  435   gagcctcatgggcatcccttatcttgatgcacccagtgaggcagaggccagctgtgctgc  494
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  481   GAGCCTCATGGGCATCCCTTATCTTGATGCACCCAGTGAGGCAGAGGCCAGCTGTGCTGC  540
              S  L  M  G  I  P  Y  L  D  A  P  S  E  A  E  A  S  C  A  A

              L  V  K  A  G  K  V  Y  A  A  A  T  E  D  M  D  C  L  T  F
Seq_1  495   cctggtgaaggctggcaaagtctatgctgcggctaccgaggacatggactgcctcacctt  554
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  541   CCTGGTGAAGGCTGGCAAAGTCTATGCTGCGGCTACCGAGGACATGGACTGCCTCACCTT  600
              L  V  K  A  G  K  V  Y  A  A  A  T  E  D  M  D  C  L  T  F

              G  S  P  V  L  M  R  H  L  T  A  S  E  A  K  K  L  P  I  Q
Seq_1  555   cggcagccctgtgctaatgcgacacctgactgccagtgaagccaaaaagctgccaatcca  614
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  601   CGGCAGCCCTGTGCTAATGCGACACCTGACTGCCAGTGAAGCCAAAAAGCTGCCAATCCA  660
              G  S  P  V  L  M  R  H  L  T  A  S  E  A  K  K  L  P  I  Q
```

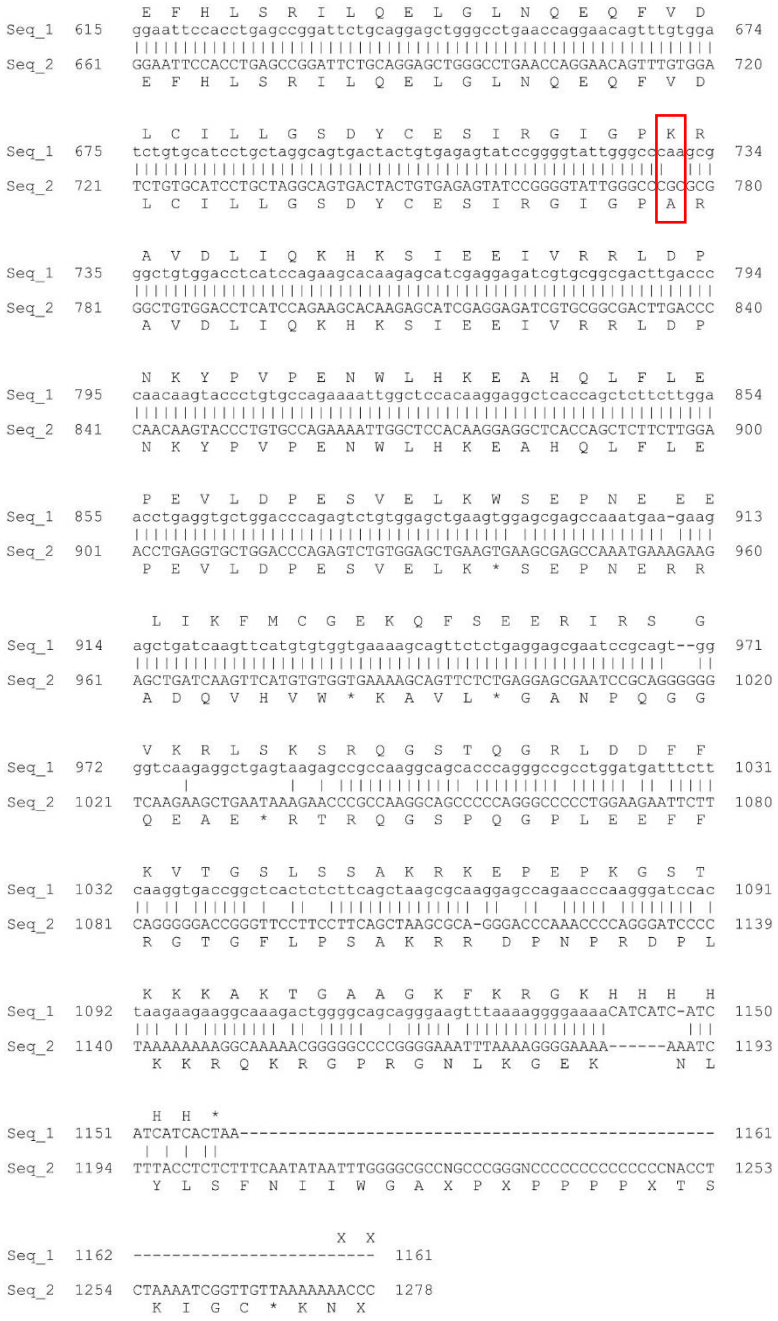


Figure 8.6 K244A forward primer sequencing

Forward primer sequencing of K244A with the alanine substitution highlighted by a red rectangle.

```
Alignment of Sequence_1:  [hFEN1_mRNA.xdna] with  Sequence_2: [K244A-R-T7-Term.seq]

Similarity : 1012/1161 (87.17 %)

            M  G  I  Q  G  L  A  K  L  I  A  D  V  A  P  S  A  I  R  E
Seq_1  1     atgggaattcaaggcctggccaaactaattgctgatgtggcccccagtgccatccgggag  60

Seq_2  1127  ------------------------------------------------------------  1128


            N  D  I  K  S  Y  F  G  R  K  V  A  I  D  A  S  M  S     I  Y
Seq_1  61    aatgacatcaagagctactttggccgtaaggtggccattgatgcctctatgag--cattt  118
                                                                 |  |  | |||
Seq_2  1127  ----------------------------------------------------GGGCCCTTTT  1118
                                                                 G  P  F  F


              Q  F  L  I  A  V  R  Q  G  G  D  V  L  Q  N  E  E  G  E  T
Seq_1  119   atcagttcctgattgctgttcgccagggtggggatgtgctgcagaatgaggagggtgaga  178
                 |  |     ||  | | | ||||||||||  | :  |  | |  | || ||  | |
Seq_2  1117  TCAGTTCCTGATTTCTTTCCCCCAGGGTGGGATTTNCTTCAAAATGGGGGGGGTGGGGC  1058
               S  S  *  F  L  F  P  Q  G  G  I  X  F  K  M  G  G  V  G  P


              T  S  H  L  M  G  M  F  Y  R  T     I  R  M  M  E  N  G  I
Seq_1  179   ccaccagccacctgatgggcatgttctaccgcac-cattcgcatgatggagaacggcatc  237
              || |||||| |||||||||||||||||||| | | ||||||||||||||||||||||||
Seq_2  1057  CCCCCAGCCCCCTGATGGGCATGTTCTACCGC-CCCCTTCGCATGATGGAGAACGGCATC  999
                 P  S  P  L  M  G  M  F  Y  R     P  L  R  M  M  E  N  G  I


            K  P  V  Y  V  F  D  G  K  P  P  Q  L  K  S  G  E  L  A  K
Seq_1  238   aagcccgtgtatgtctttgatggcaagccgccacagctcaagtcaggcgagctggccaaa  297
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  998   AAGCCCGTGTATGTCTTTGATGGCAAGCCGCCACAGCTCAAGTCAGGCGAGCTGGCCAAA  939
              K  P  V  Y  V  F  D  G  K  P  P  Q  L  K  S  G  E  L  A  K


            R  S  E  R  R  A  E  A  E  K  Q  L  Q  Q  A  Q  A  A  G  A
Seq_1  298   cgcagtgagcggcgggctgaggcagagaagcagctgcagcaggctcaggctgctggggcc  357
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  938   CGCAGTGAGCGGCGGGCTGAGGCAGAGAAGCAGCTGCAGCAGGCTCAGGCTGCTGGGGCC  879
              R  S  E  R  R  A  E  A  E  K  Q  L  Q  Q  A  Q  A  A  G  A


            E  Q  E  V  E  K  F  T  K  R  L  V  K  V  T  K  Q  H  N  D
Seq_1  358   gagcaggaggtggaaaaattcactaagcggctggtgaaggtcactaagcagcacaatgat  417
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  878   GAGCAGGAGGTGGAAAAATTCACTAAGCGGCTGGTGAAGGTCACTAAGCAGCACAATGAT  819
              E  Q  E  V  E  K  F  T  K  R  L  V  K  V  T  K  Q  H  N  D


            E  C  K  H  L  L  S  L  M  G  I  P  Y  L  D  A  P  S  E  A
Seq_1  418   gagtgcaaacatctgctgagcctcatgggcatcccttatcttgatgcacccagtgaggca  477
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  818   GAGTGCAAACATCTGCTGAGCCTCATGGGCATCCCTTATCTTGATGCACCCAGTGAGGCA  759
              E  C  K  H  L  L  S  L  M  G  I  P  Y  L  D  A  P  S  E  A


            E  A  S  C  A  A  L  V  K  A  G  K  V  Y  A  A  A  T  E  D
Seq_1  478   gaggccagctgtgctgccctggtgaaggctggcaaagtctatgctgcggctaccgaggac  537
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  758   GAGGCCAGCTGTGCTGCCCTGGTGAAGGCTGGCAAAGTCTATGCTGCGGCTACCGAGGAC  699
              E  A  S  C  A  A  L  V  K  A  G  K  V  Y  A  A  A  T  E  D


            M  D  C  L  T  F  G  S  P  V  L  M  R  H  L  T  A  S  E  A
Seq_1  538   atggactgcctcaccttcggcagccctgtgctaatgcgacacctgactgccagtgaagcc  597
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  698   ATGGACTGCCTCACCTTCGGCAGCCCTGTGCTAATGCGACACCTGACTGCCAGTGAAGCC  639
              M  D  C  L  T  F  G  S  P  V  L  M  R  H  L  T  A  S  E  A


            K  K  L  P  I  Q  E  F  H  L  S  R  I  L  Q  E  L  G  L  N
Seq_1  598   aaaaagctgccaatccaggaattccacctgagccggattctgcaggagctgggcctgaac  657
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  638   AAAAAGCTGCCAATCCAGGAATTCCACCTGAGCCGGATTCTGCAGGAGCTGGGCCTGAAC  579
              K  K  L  P  I  Q  E  F  H  L  S  R  I  L  Q  E  L  G  L  N
```

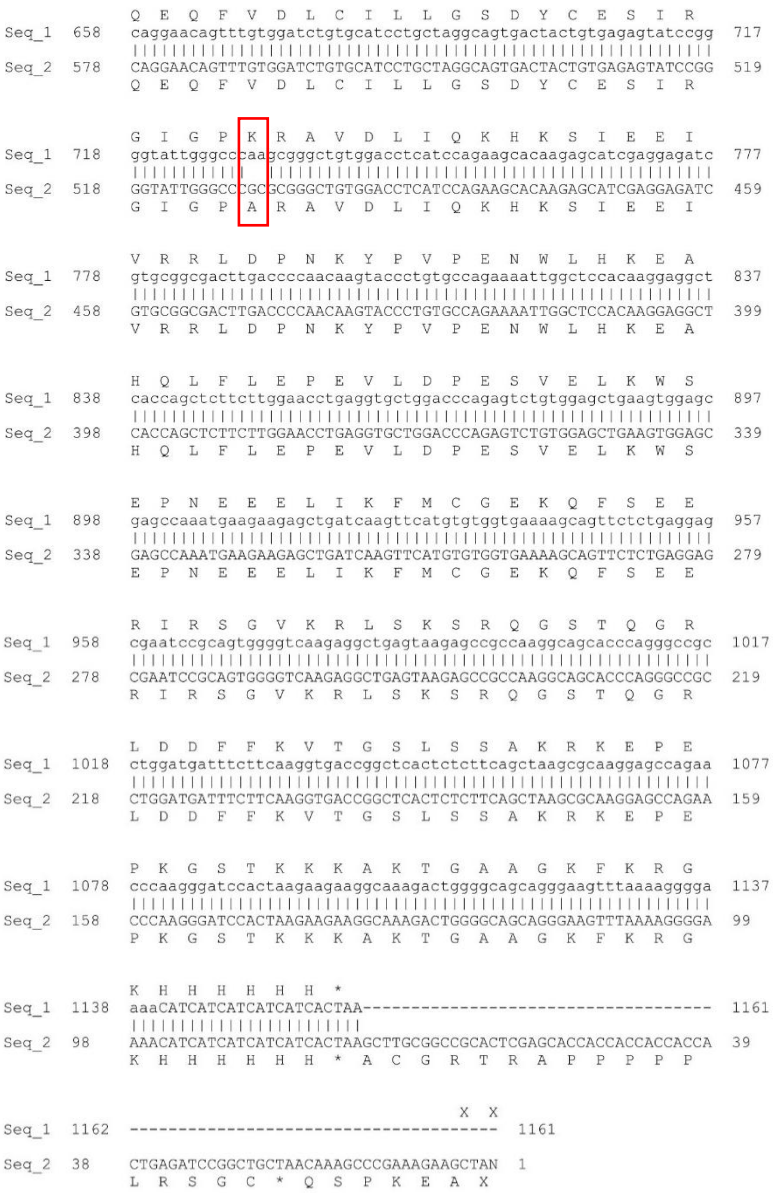


Figure 8.7 K244A reverse primer sequencing

Reverse primer sequencing of K244A with the alanine substitution highlighted by a red rectangle.

```
Alignment of Sequence_1:  [hFEN1_mRNA.xdna] with  Sequence_2: [pET_R245A_F-T7.s

Similarity : 1012/1161 (87.17 %)

                                                                 M  G
Seq_1  1     ------------------------------------------------------atggg  5
                                                                  |||||
Seq_2  1     NNNGGGGAATTCCCCTCTAGAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGG  60
              X  G  N  S  P  L  E  *  F  C  L  T  L  R  R  R  Y  T  M  G

              I  Q  G  L  A  K  L  I  A  D  V  A  P  S  A  I  R  E  N  D
Seq_1  6     aattcaaggcctggccaaactaattgctgatgtggcccccagtgccatccgggagaatga  65
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  61    AATTCAAGGCCTGGCCAAACTAATTGCTGATGTGGCCCCCAGTGCCATCCGGGAGAATGA  120
              I  Q  G  L  A  K  L  I  A  D  V  A  P  S  A  I  R  E  N  D

              I  K  S  Y  F  G  R  K  V  A  I  D  A  S  M  S  I  Y  Q  F
Seq_1  66    catcaagagctactttggccgtaaggtggccattgatgcctctatgagcatttatcagtt  125
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  121   CATCAAGAGCTACTTTGGCCGTAAGGTGGCCATTGATGCCTCTATGAGCATTTATCAGTT  180
              I  K  S  Y  F  G  R  K  V  A  I  D  A  S  M  S  I  Y  Q  F

              L  I  A  V  R  Q  G  G  D  V  L  Q  N  E  E  G  E  T  T  S
Seq_1  126   cctgattgctgttcgccagggtggggatgtgctgcagaatgaggagggtgagaccaccag  185
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  181   CCTGATTGCTGTTCGCCAGGGTGGGGATGTGCTGCAGAATGAGGAGGGTGAGACCACCAG  240
              L  I  A  V  R  Q  G  G  D  V  L  Q  N  E  E  G  E  T  T  S

              H  L  M  G  M  F  Y  R  T  I  R  M  M  E  N  G  I  K  P  V
Seq_1  186   ccacctgatgggcatgttctaccgcaccattcgcatgatggagaacggcatcaagcccgt  245
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  241   CCACCTGATGGGCATGTTCTACCGCACCATTCGCATGATGGAGAACGGCATCAAGCCCGT  300
              H  L  M  G  M  F  Y  R  T  I  R  M  M  E  N  G  I  K  P  V

              Y  V  F  D  G  K  P  P  Q  L  K  S  G  E  L  A  K  R  S  E
Seq_1  246   gtatgtctttgatggcaagccgccacagctcaagtcaggcgagctggccaaacgcagtga  305
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  301   GTATGTCTTTGATGGCAAGCCGCCACAGCTCAAGTCAGGCGAGCTGGCCAAACGCAGTGA  360
              Y  V  F  D  G  K  P  P  Q  L  K  S  G  E  L  A  K  R  S  E

              R  R  A  E  A  E  K  Q  L  Q  Q  A  Q  A  A  G  A  E  Q  E
Seq_1  306   gcggcgggctgaggcagagaagcagctgcagcaggctcaggctgctggggccgagcagga  365
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  361   GCGGCGGGCTGAGGCAGAGAAGCAGCTGCAGCAGGCTCAGGCTGCTGGGGCCGAGCAGGA  420
              R  R  A  E  A  E  K  Q  L  Q  Q  A  Q  A  A  G  A  E  Q  E

              V  E  K  F  T  K  R  L  V  K  V  T  K  Q  H  N  D  E  C  K
Seq_1  366   ggtggaaaaattcactaagcggctggtgaaggtcactaagcagcacaatgatgagtgcaa  425
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  421   GGTGGAAAAATTCACTAAGCGGCTGGTGAAGGTCACTAAGCAGCACAATGATGAGTGCAA  480
              V  E  K  F  T  K  R  L  V  K  V  T  K  Q  H  N  D  E  C  K

              H  L  L  S  L  M  G  I  P  Y  L  D  A  P  S  E  A  E  A  S
Seq_1  426   acatctgctgagcctcatgggcatcccttatcttgatgcacccagtgaggcagaggccag  485
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  481   ACATCTGCTGAGCCTCATGGGCATCCCTTATCTTGATGCACCCAGTGAGGCAGAGGCCAG  540
              H  L  L  S  L  M  G  I  P  Y  L  D  A  P  S  E  A  E  A  S

              C  A  A  L  V  K  A  G  K  V  Y  A  A  A  T  E  D  M  D  C
Seq_1  486   ctgtgctgccctggtgaaggctggcaaagtctatgctgcggctaccgaggacatggactg  545
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  541   CTGTGCTGCCCTGGTGAAGGCTGGCAAAGTCTATGCTGCGGCTACCGAGGACATGGACTG  600
              C  A  A  L  V  K  A  G  K  V  Y  A  A  A  T  E  D  M  D  C

              L  T  F  G  S  P  V  L  M  R  H  L  T  A  S  E  A  K  K  L
Seq_1  546   cctcaccttcggcagccctgtgctaatgcgacacctgactgccagtgaagccaaaaagct  605
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  601   CCTCACCTTCGGCAGCCCTGTGCTAATGCGACACCTGACTGCCAGTGAAGCCAAAAAGCT  660
              L  T  F  G  S  P  V  L  M  R  H  L  T  A  S  E  A  K  K  L
```

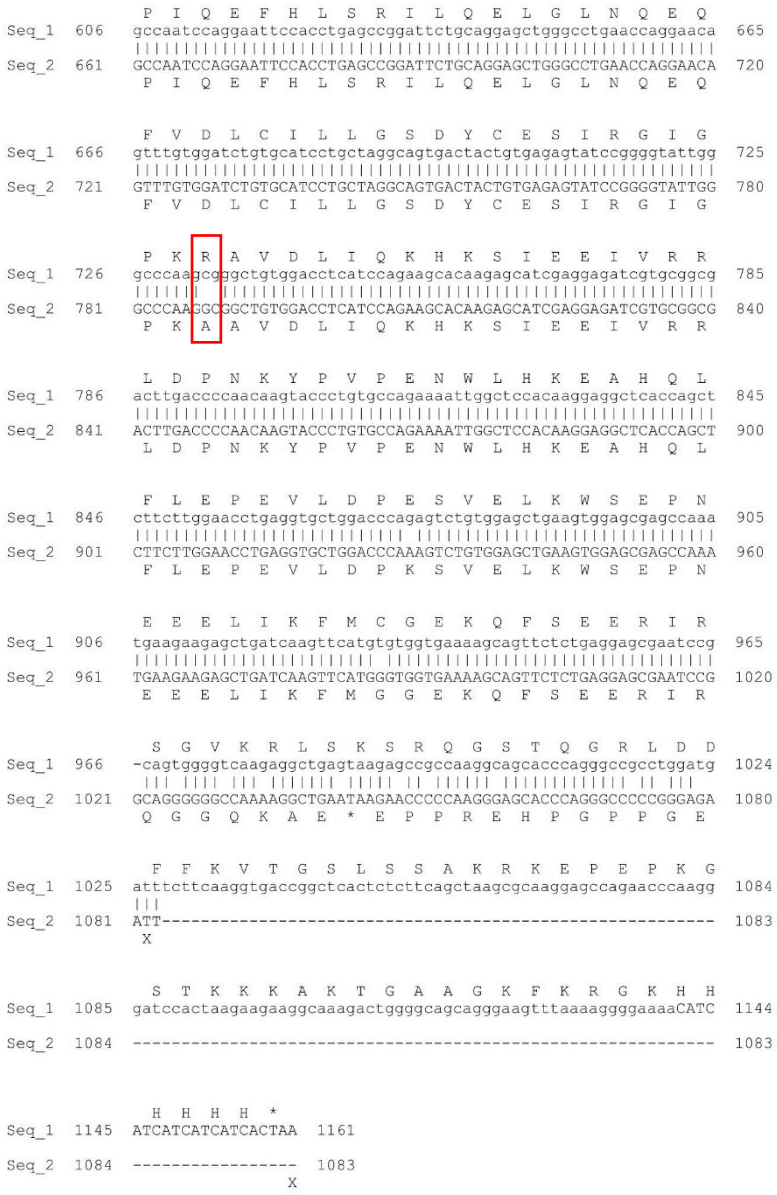


Figure 8.8 R245A forward primer sequencing

Forward primer sequencing of R245A with the alanine substitution highlighted by a red rectangle.

```
Alignment of Sequence_1:  [hFEN1_mRNA.xdna] with  Sequence_2: [pET_R245A_R-T7-Term.seq]

Similarity : 1060/1313 (80.73 %)

Seq_1  1     ------------------------------------------------------------  0

Seq_2  1313  CTCAAGGGGGGGGTTGGGGGGGAACACTTCCCCCCTAAAAAAATTTTTTTTNNNTAAAAA  1254
              S  R  G  G  L  G  G  N  T  S  P  L  K  K  F  F  X  X  K  K


                           M  G  I   Q  G  L  A   K   L  I    A  D
Seq_1  1     ------------------atgggaatt-caaggcctggcc-a-aactaa--ttgctgat  36
                               ||   ||   |  || :      :  | |     ||  :
Seq_2  1253  GGGGAATNNNNCCGGGGAAATT-AAAGGGC-CGGGNAAAAANNTTATTTGTGGTGGGNCC  1196
              G  N  X  X  G  E  I   K  G  P   X  K  X  X  I  C  G  G  X


             V  A   P  S  A  I  R  E  N  D  I  K  S  Y  F  G  R  K  V  A
Seq_1  37    gtgg-ccccagtgccatccgggagaatgacatcaagagctactttggccgtaaggtggc  95
                ::|||  | |       | |     |      ::|      |      ||   |
Seq_2  1195  CCCCNNNCCCCTTCCGGGGAAAAACACCCCAAAAGTTTNNTTTGCGCGTAAGGGGGGCCC  1136
             P  X  X  P  S  G  E  K  H  P  K  S  X  X  C  A  *  G  G  P


              I  D  A  S  M  S  I  Y  Q  F  L  I  A  V  R  Q  G  G  D  V
Seq_1  96    cattgatgcctctatgagcatttatcagttcctgattgctgttcgccagggtggggatgt  155
                 |     |          |||     | || ||||||| ||| ||||| ||||||
Seq_2  1135  TGAAGCCCTTTAGGGCCATTTTTCCAGTTCCCGGATTGCTTTTCCCCAGGGGGGGGATT-  1077
             *  S  P  L  G  P  F  F  Q  F  P  D  C  F  S  P  G  G  G  L


              L  Q  N  E  E  G  E  T  T  S  H  L  M  G  M  F  Y  R  T  I
Seq_1  156   gctgcagaatgaggagggtgagaccaccagccacctgatgggcatgttctaccgcaccat  215
             |:|||||||||||||||||||||| ||||| |||||||||||||||||||||| || |
Seq_2  1076  GNTGCAGAATGAGGAGGGTGAGACCCCCAGCCCCCTGATGGGCATGTTCTACCGCCCCCT  1017
              X  Q  N  E  E  G  E  T  P  S  P  L  M  G  M  F  Y  R  P  L


              R  M  M  E  N  G  I  K  P  V  Y  V  F  D  G  K  P  P  Q  L
Seq_1  216   tcgcatgatggagaacggcatcaagcccgtgtatgtctttgatggcaagccgccacagct  275
             |||: ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  1016  TCGNCTGATGGAGAACGGCATCAAGCCCGTGTATGTCTTTGATGGCAAGCCGCCACAGCT  957
              X  L  M  E  N  G  I  K  P  V  Y  V  F  D  G  K  P  P  Q  L


              K  S  G  E  L  A  K  R  S  E  R  R  A  E  A  E  K  Q  L  Q
Seq_1  276   caagtcaggcgagctggccaaacgcagtgagcggcgggctgaggcagagaagcagctgca  335
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  956   CAAGTCAGGCGAGCTGGCCAAACGCAGTGAGCGGCGGGCTGAGGCAGAGAAGCAGCTGCA  897
              K  S  G  E  L  A  K  R  S  E  R  R  A  E  A  E  K  Q  L  Q


              Q  A  Q  A  A  G  A  E  Q  E  V  E  K  F  T  K  R  L  V  K
Seq_1  336   gcaggctcaggctgctggggccgagcaggaggtggaaaaattcactaagcggctggtgaa  395
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  896   GCAGGCTCAGGCTGCTGGGGCCGAGCAGGAGGTGGAAAAATTCACTAAGCGGCTGGTGAA  837
              Q  A  Q  A  A  G  A  E  Q  E  V  E  K  F  T  K  R  L  V  K


              V  T  K  Q  H  N  D  E  C  K  H  L  L  S  L  M  G  I  P  Y
Seq_1  396   ggtcactaagcagcacaatgatgagtgcaaacatctgctgagcctcatgggcatcccttа  455
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  836   GGTCACTAAGCAGCACAATGATGAGTGCAAACATCTGCTGAGCCTCATGGGCATCCCTTA  777
              V  T  K  Q  H  N  D  E  C  K  H  L  L  S  L  M  G  I  P  Y


              L  D  A  P  S  E  A  E  A  S  C  A  A  L  V  K  A  G  K  V
Seq_1  456   tcttgatgcacccagtgaggcagaggccagctgtgctgccctggtgaaggctggcaaagt  515
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  776   TCTTGATGCACCCAGTGAGGCAGAGGCCAGCTGTGCTGCCCTGGTGAAGGCTGGCAAAGT  717
              L  D  A  P  S  E  A  E  A  S  C  A  A  L  V  K  A  G  K  V


              Y  A  A  A  T  E  D  M  D  C  L  T  F  G  S  P  V  L  M  R
Seq_1  516   ctatgctgcggctaccgaggacatggactgcctcaccttcggcagccctgtgctaatgcg  575
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  716   CTATGCTGCGGCTACCGAGGACATGGACTGCCTCACCTTCGGCAGCCCTGTGCTAATGCG  657
              Y  A  A  A  T  E  D  M  D  C  L  T  F  G  S  P  V  L  M  R
```

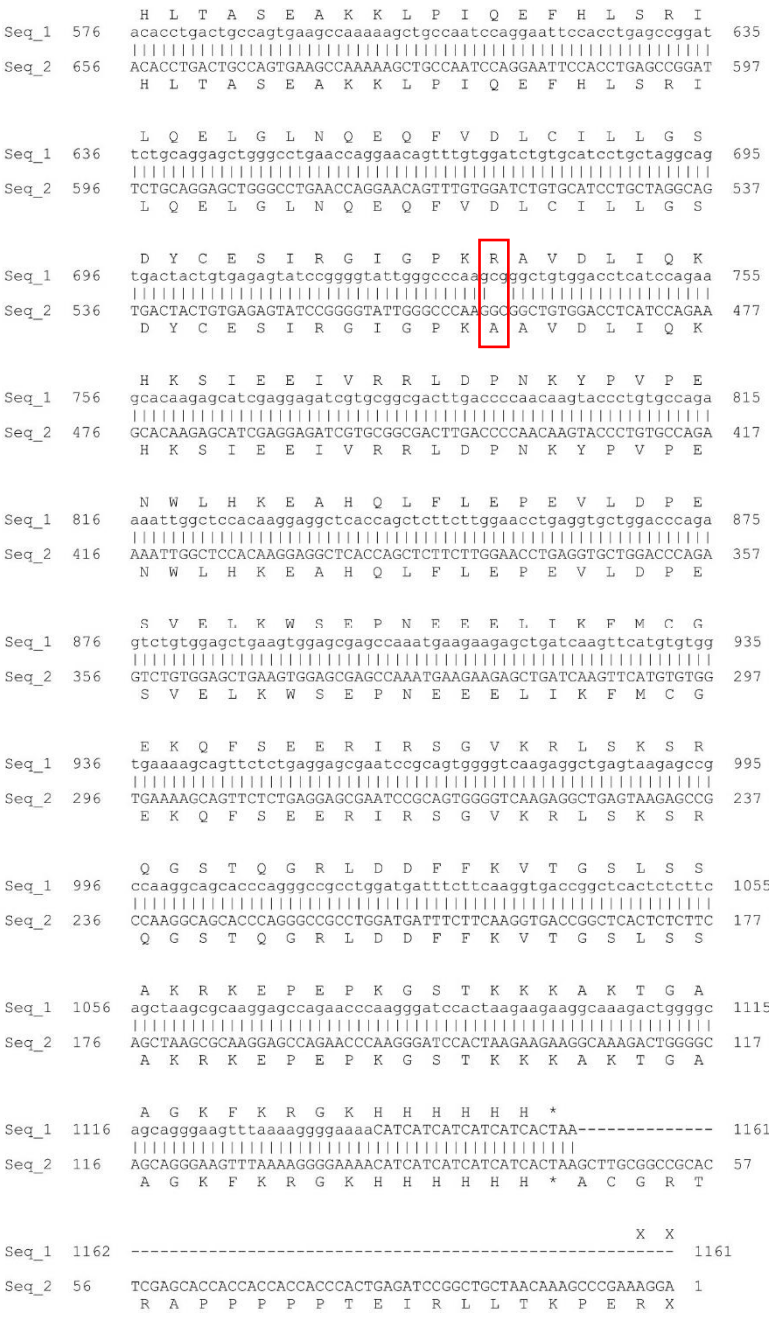


Figure 8.9 R245A reverse primer sequencing

Forward primer sequencing of R245A with the alanine substitution highlighted by a red rectangle.

8.4.2 FAN1-SUMO construct sequencing

```
Alignment of Sequence_1:  [pE-SUMOpro Amp_hFAN1-373-1017.txt.xdna] with  Sequence_2: [pESumo-hFAN1-373-1017_9-
T7[1].seq]

Similarity : 1014/7673 (13.22 %)

                I  S  I  P  R  N  *  Y  D  S  L  *  G  N  C  E  R  I  T  I
Seq_1  1     agatctcgatcccgcgaaattaatacgactcactatagggaattgtgagcggataacaa  60
                                                    ::::::::::
Seq_2  1     ----------------------------------------------NNNNNNNNNNN  11
                                                     X  X  X  X

                P  L  *  K  *  F  C  L  T  L  R  R  R  Y  T  M  G  H  H  H
Seq_1  61    ttccctctagaaataattttgtttaactttaagaaggagatataccatggtcatcacc  120
              | :::|||||  ||||||||||||||||||||||||||||||||||||||||||||
Seq_2  12    -T-NNNTCTAG-AATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGGTCATCACC  68
                X  L  E   *  F  C  L  T  L  R  R  R  Y  T  M  G  H  H  H

                H  H  H  G  S  L  Q  D  S  E  V  N  Q  E  A  K  P  E  V  K
Seq_1  121   atcatcatcacgggtccctgcaggactcagaagtcaatcaagaagctaagccagaggtca  180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  69    ATCATCATCACGGGTCCCTGCAGGACTCAGAAGTCAATCAAGAAGCTAAGCCAGAGGTCA  128
                H  H  H  G  S  L  Q  D  S  E  V  N  Q  E  A  K  P  E  V  K

                P  E  V  K  P  E  T  H  I  N  L  K  V  S  D  G  S  S  E  I
Seq_1  181   agccagaagtcaagcctgagactcacatcaatttaaaggtgtccgatggatcttcagaga  240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  129   AGCCAGAAGTCAAGCCTGAGACTCACATCAATTTAAAGGTGTCCGATGGATCTTCAGAGA  188
                P  E  V  K  P  E  T  H  I  N  L  K  V  S  D  G  S  S  E  I

                F  F  K  I  K  K  T  T  P  L  R  R  L  M  E  A  F  A  K  R
Seq_1  241   tcttcttcaagatcaaaaagaccactcctttaagaaggctgatggaagcgttcgctaaaa  300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  189   TCTTCTTCAAGATCAAAAAGACCACTCCTTTAAGAAGGCTGATGGAAGCGTTCGCTAAAA  248
                F  F  K  I  K  K  T  T  P  L  R  R  L  M  E  A  F  A  K  R

                Q  G  K  E  M  D  S  L  R  F  L  Y  D  G  I  R  I  Q  A  D
Seq_1  301   gacagggtaaggaaatggactccttaagattcttgtacgacggtattagaattcaagctg  360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  249   GACAGGGTAAGGAAATGGACTCCTTAAGATTCTTGTACGACGGTATTAGAATTCAAGCTG  308
                Q  G  K  E  M  D  S  L  R  F  L  Y  D  G  I  R  I  Q  A  D

                Q  A  P  E  D  L  D  M  E  D  N  D  I  I  E  A  H  R  E  Q
Seq_1  361   atcaggcccctgaagatttggacatggaggataacgatattattgaggctcaccgcgaac  420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  309   ATCAGGCCCCTGAAGATTTGGACATGGAGGATAACGATATTATTGAGGCTCACCGCGAAC  368
                Q  A  P  E  D  L  D  M  E  D  N  D  I  I  E  A  H  R  E  Q

                I  G  G  P  Y  Y  L  R  S  F  L  V  V  L  K  T  V  L  E  N
Seq_1  421   agattggaggtCCTTACTACCTTCGGAGTTTCCttgtggtgctgaaaaccgtacttgaga  480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  369   AGATTGGAGGTCCTTACTACCTTCGGAGTTTCCTTGTGGTGCTGAAAACCGTACTTGAGA  428
                I  G  G  P  Y  Y  L  R  S  F  L  V  V  L  K  T  V  L  E  N

                E  D  D  M  L  L  F  D  E  Q  E  K  G  I  V  T  K  F  Y  Q
Seq_1  481   atgaagatgatatgttgctctttgatgagcaggagaagggaattgtaactaaattttatc  540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  429   ATGAAGATGATATGTTGCTCTTTGATGAGCAGGAGAAGGGAATTGTAACTAAATTTTATC  488
                E  D  D  M  L  L  F  D  E  Q  E  K  G  I  V  T  K  F  Y  Q

                L  S  A  T  G  Q  K  L  Y  V  R  L  F  Q  R  K  L  S  W  I
Seq_1  541   agttatcagctactggtcagaagttatatgtaaggctctttcaacgtaaattaagctgga  600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  489   AGTTATCAGCTACTGGTCAGAAGTTATATGTAAGGCTCTTTCAACGTAAATTAAGCTGGA  548
                L  S  A  T  G  Q  K  L  Y  V  R  L  F  Q  R  K  L  S  W  I

                K  M  T  K  L  E  Y  E  E  I  A  L  D  L  T  P  V  I  E  E
Seq_1  601   ttaagatgaccaaattagagtatgaagagattgccttagacttaacacctgtgattgaag  660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  549   TTAAGATGACCAAATTAGAGTATGAAGAGATTGCCTTAGACTTAACACCTGTGATTGAAG  608
                K  M  T  K  L  E  Y  E  E  I  A  L  D  L  T  P  V  I  E  E
```

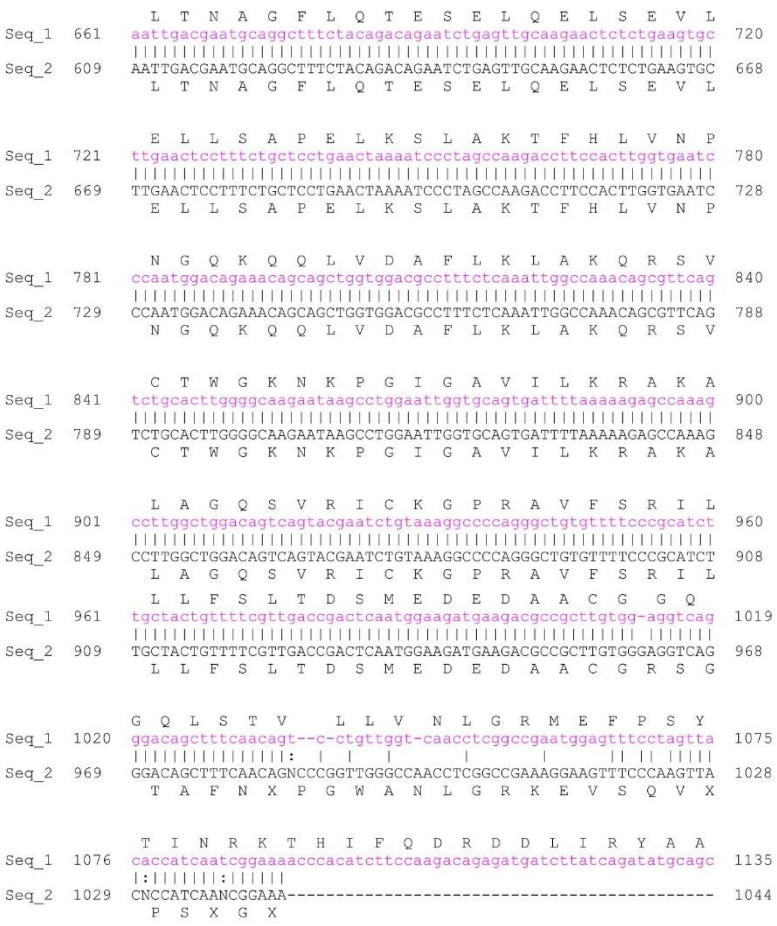
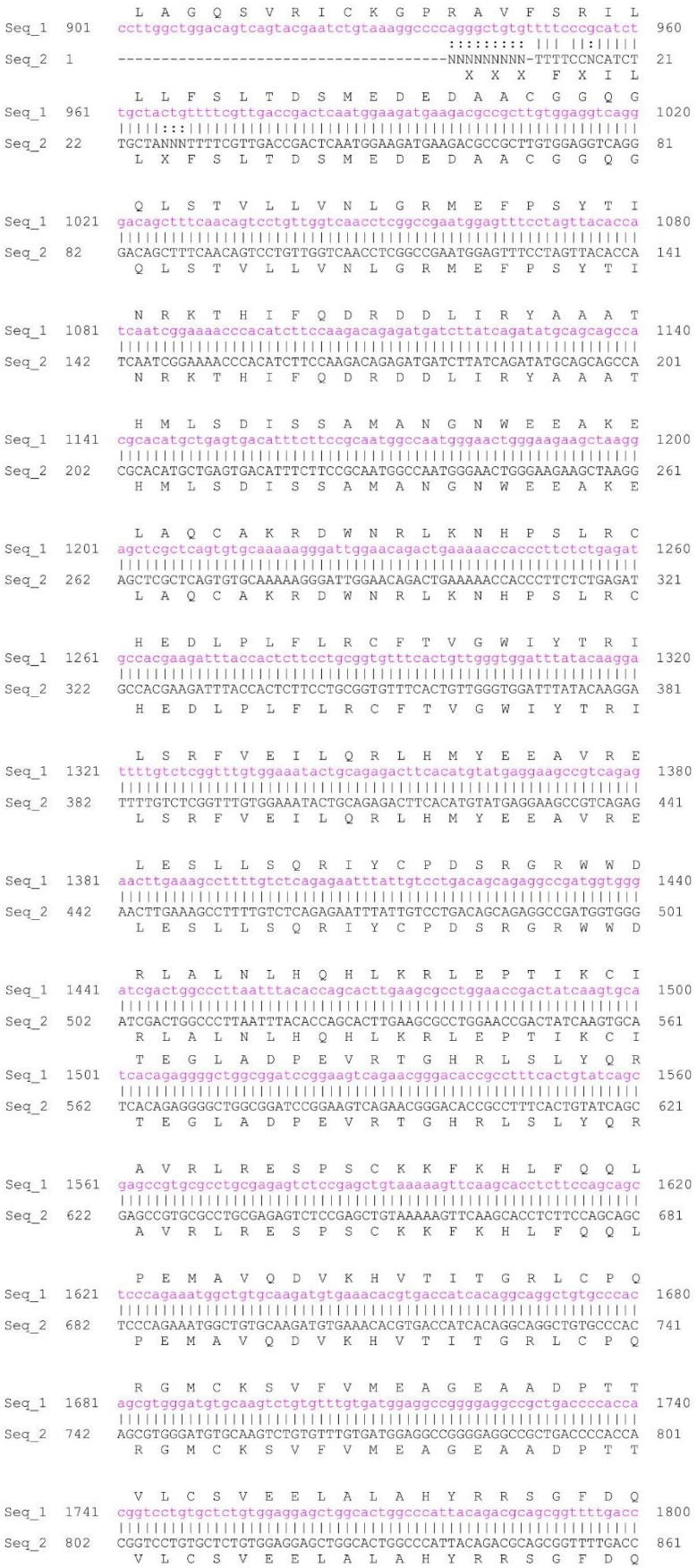


Figure 8.10 FAN1-SUMO forward primer sequencing

Forward primer sequencing of the FAN1-SUMO construct. FAN1 (residues 373-1017) are shown in purple; the SUMO and His6 tag are shown in orange.


L A G Q S V R I C K G P R A V F S R I L
Seq_1 901 ccttggctggacagtcagtacgaatctgtaaaggccccagggctgtgttttcccgcatct 960
Seq_2 1 ----------NNNNNNNNN-TTTTCCNCATCT 21
L L F S L T D S M E D E D A A C G G Q G
Seq_1 961 tgctactgttttcgttgaccgactcaatggaagatgaagacgccgcttgtggaggtcagg 1020
Seq_2 22 TGCTANNNTTTTCGTTGACCGACTCAATGGAAGATGAAGACGCCGCTTGTGGAGGTCAGG 81
L X F S L T D S M E D E D A A C G G Q G
Q L S T V L L V N L G R M E F P S Y T I
Seq_1 1021 gacagctttcaacagtcctgttggtcaacctcggccgaatggagtttcctagttacacca 1080
Seq_2 82 GACAGCTTTCAACAGTCCTGTTGGTCAACCTCGGCCGAATGGAGTTTCCTAGTTACACCA 141
Q L S T V L L V N L G R M E F P S Y T I
N R K T H I F Q D R D D L I R Y A A A T
Seq_1 1081 tcaatcggaaaacccacatcttccaagacagagatgatcttatcagatatgcagcagcca 1140
Seq_2 142 TCAATCGGAAAACCCACATCTTCCAAGACAGAGATGATCTTATCAGATATGCAGCAGCCA 201
N R K T H I F Q D R D D L I R Y A A A T
H M L S D I S S A M A N G N W E E A K E
Seq_1 1141 cgcacatgctgagtgacatttcttccgcaatggccaatgggaactgggaagaagctaagg 1200
Seq_2 202 CGCACATGCTGAGTGACATTTCTTCCGCAATGGCCAATGGGAACTGGGAAGAAGCTAAGG 261
H M L S D I S S A M A N G N W E E A K E
L A Q C A K R D W N R L K N H P S L R C
Seq_1 1201 agctcgctcagtgtgcaaaaagggattggaacagactgaaaaaccacccttctctgagat 1260
Seq_2 262 AGCTCGCTCAGTGTGCAAAAAGGGATTGGAACAGACTGAAAAACCACCCTTCTCTGAGAT 321
L A Q C A K R D W N R L K N H P S L R C
H E D L P L F L R C F T V G W I Y T R I
Seq_1 1261 gccacgaagatttaccactcttcctgcggtgtttcactgttgggtggatttatacaagga 1320
Seq_2 322 GCCACGAAGATTTACCACTCTTCCTGCGGTGTTTCACTGTTGGGTGGATTTATACAAGGA 381
H E D L P L F L R C F T V G W I Y T R I
L S R F V E I L Q R L H M Y E E A V R E
Seq_1 1321 ttttgtctcggtttgtggaaatactgcagagacttcacatgtatgaggaagccgtcagag 1380
Seq_2 382 TTTTGTCTCGGTTTGTGGAAATACTGCAGAGACTTCACATGTATGAGGAAGCCGTCAGAG 441
L S R F V E I L Q R L H M Y E E A V R E
L E S L L S Q R I Y C P D S R G R W W D
Seq_1 1381 aacttgaaagccttttgtctcagagaatttattgtcctgacagcagaggccgatggtggg 1440
Seq_2 442 AACTTGAAAGCCTTTTGTCTCAGAGAATTTATTGTCCTGACAGCAGAGGCCGATGGTGGG 501
L E S L L S Q R I Y C P D S R G R W W D
R L A L N L H Q H L K R L E P T I K C I
Seq_1 1441 atcgactggcccttaatttacaccagcacttgaagcgcctggaaccgactatcaagtgca 1500
Seq_2 502 ATCGACTGGCCCTTAATTTACACCAGCACTTGAAGCGCCTGGAACCGACTATCAAGTGCA 561
R L A L N L H Q H L K R L E P T I K C I
T E G L A D P E V R T G H R L S L Y Q R
Seq_1 1501 tcacagaggggctggcggatccggaagtcagaacgggacaccgcctttcactgtatcagc 1560
Seq_2 562 TCACAGAGGGGCTGGCGGATCCGGAAGTCAGAACGGGACACCGCCTTTCACTGTATCAGC 621
T E G L A D P E V R T G H R L S L Y Q R
A V R L R E S P S C K K F K H L F Q Q L
Seq_1 1561 gagccgtgcgcctgcgagagtctccgagctgtaaaaagttcaagcacctcttccagcagc 1620
Seq_2 622 GAGCCGTGCGCCTGCGAGAGTCTCCGAGCTGTAAAAAGTTCAAGCACCTCTTCCAGCAGC 681
A V R L R E S P S C K K F K H L F Q Q L
P E M A V Q D V K H V T I T G R L C P Q
Seq_1 1621 tcccagaaatggctgtgcaagatgtgaaacacgtgaccatcacaggcaggctgtgcccac 1680
Seq_2 682 TCCCAGAAATGGCTGTGCAAGATGTGAAACACGTGACCATCACAGGCAGGCTGTGCCCAC 741
P E M A V Q D V K H V T I T G R L C P Q
R G M C K S V F V M E A G E A A D P T T
Seq_1 1681 agcgtgggatgtgcaagtctgtgtttgtgatggaggccggggaggccgctgaccccacca 1740
Seq_2 742 AGCGTGGGATGTGCAAGTCTGTGTTTGTGATGGAGGCCGGGGAGGCCGCTGACCCCACCA 801
R G M C K S V F V M E A G E A A D P T T
V L C S V E E L A L A H Y R R S G F D Q
Seq_1 1741 cggtcctgtgctctgtggaggagctggcactggcccattacagacgcagcggttttgacc 1800
Seq_2 802 CGGTCCTGTGCTCTGTGGAGGAGCTGGCACTGGCCCATTACAGACGCAGCGGTTTTGACC 861
V L C S V E E L A L A H Y R R S G F D Q

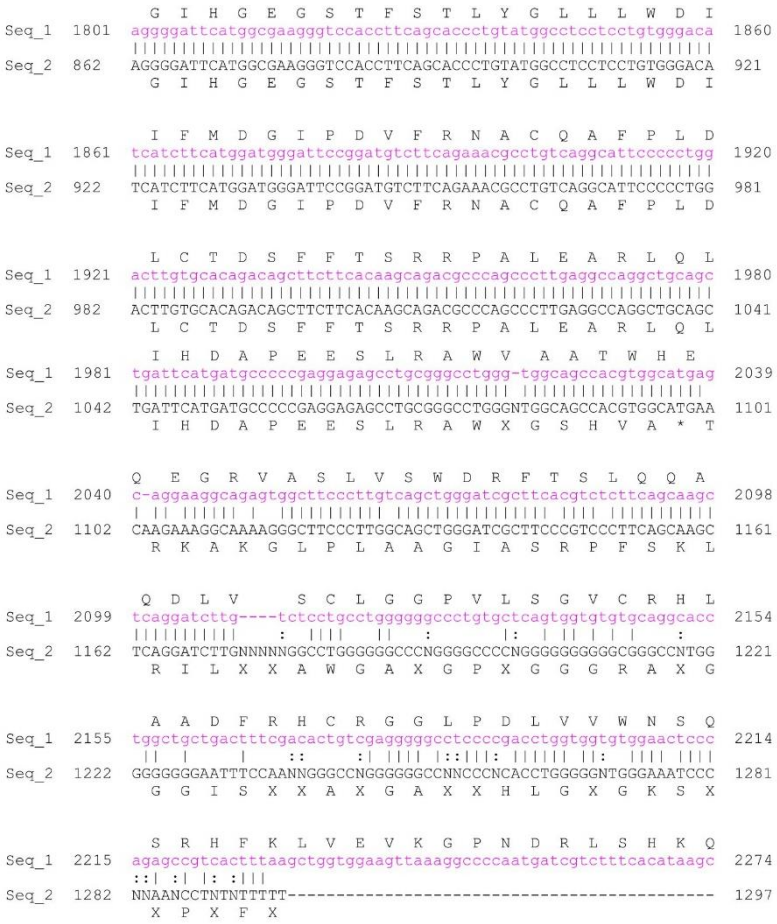
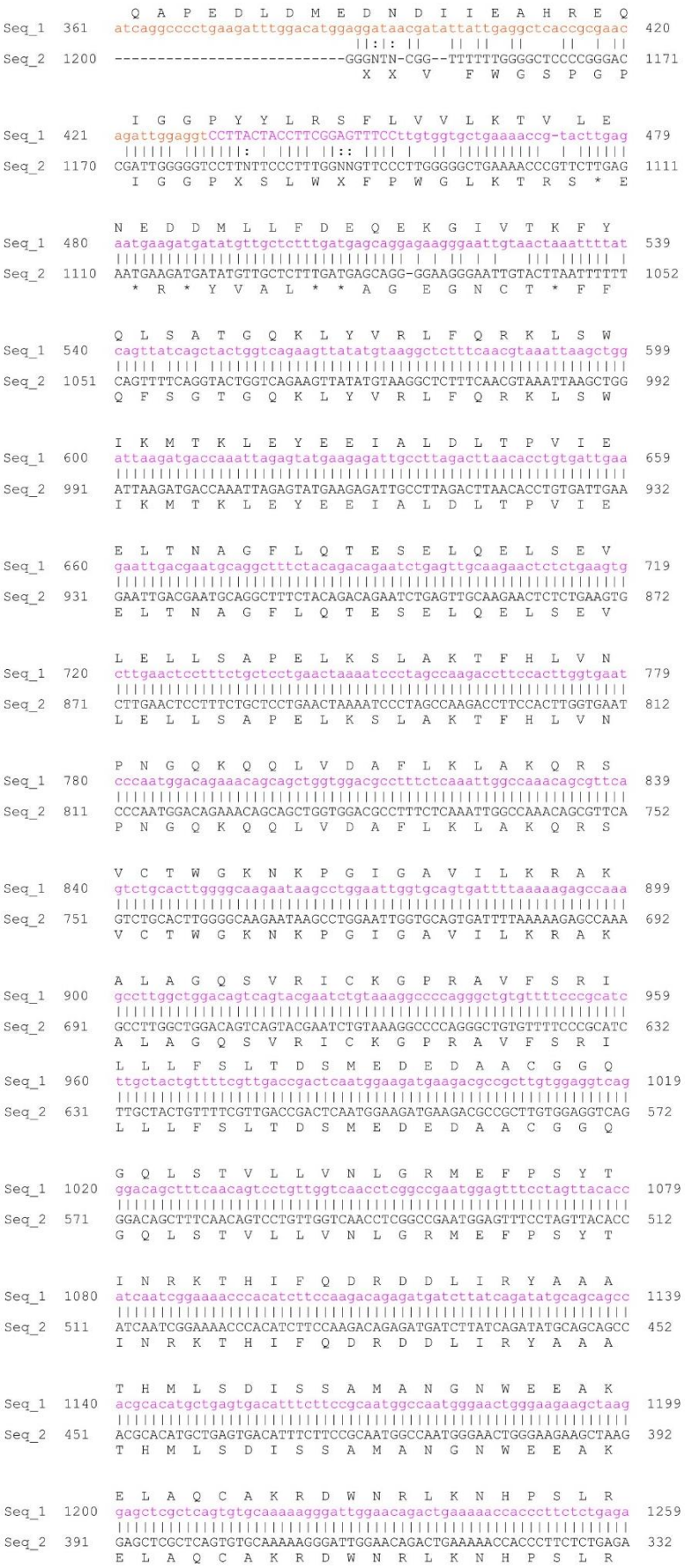


Figure 8.11 FAN1-SUMO internal forward primer sequencing

Internal forward primer sequencing of the FAN1-SUMO construct. FAN1 (residues 373-1017) are shown in purple.


Q A P E D L D M E D N D I I E A H R E Q
Seq_1 361 atcaggcccctgaagatttggacatggaggataacgatattattgaggctcaccgcgaac 420
Seq_2 1200 ----------GGGNTN-CGG--TTTTTTGGGGCTCCCCGGGAC 1171
X X V F W G S P G P
I G G P Y Y L R S F L V V L K T V L E
Seq_1 421 agattggaggtCCTTACTACCTTCGGAGTTTCCttgtggtgctgaaaaccg-tacttgag 479
Seq_2 1170 CGATTGGGGGTCCTTNTTCCCTTTGGNNGTTCCCTTGGGGGCTGAAAACCCGTTCTTGAG 1111
I G G P X S L W X F P W G L K T R S * E
N E D D M L L F D E Q E K G I V T K F Y
Seq_1 480 aatgaagatgatatgttgctctttgatgagcaggagaagggaattgtaactaaatttttat 539
Seq_2 1110 AATGAAGATGATATGTTGCTCTTTGATGAGCAGG-GGAAGGGAATTGTACTTAATTTTTT 1052
* R * Y V A L * * A G E G N C T * F F
Q L S A T G Q K L Y V R L F Q R K L S W
Seq_1 540 cagttatcagctactggtcagaagttatatgtaaggctctttcaacgtaaattaagctgg 599
Seq_2 1051 CAGTTTTCAGGTACTGGTCAGAAGTTATATGTAAGGCTCTTTCAACGTAAATTAAGCTGG 992
Q F S G T G Q K L Y V R L F Q R K L S W
I K M T K L E Y E E I A L D L T P V I E
Seq_1 600 attaagatgaccaaattagagtatgaagagattgccttagacttaacacctgtgattgaa 659
Seq_2 991 ATTAAGATGACCAAATTAGAGTATGAAGAGATTGCCTTAGACTTAACACCTGTGATTGAA 932
I K M T K L E Y E E I A L D L T P V I E
E L T N A G F L Q T E S E L Q E L S E V
Seq_1 660 gaattgacgaatgcaggctttctacagacagaatctgagttgcaagaactctctgaagtg 719
Seq_2 931 GAATTGACGAATGCAGGCTTTCTACAGACAGAATCTGAGTTGCAAGAACTCTCTGAAGTG 872
E L T N A G F L Q T E S E L Q E L S E V
L E L L S A P E L K S L A K T F H L V N
Seq_1 720 cttgaactcctttctgctcctgaactaaaatccctagccaagaccttccacttggtgaat 779
Seq_2 871 CTTGAACTCCTTTCTGCTCCTGAACTAAAATCCCTAGCCAAGACCTTCCACTTGGTGAAT 812
L E L L S A P E L K S L A K T F H L V N
P N G Q K Q Q L V D A F L K L A K Q R S
Seq_1 780 cccaatggacagaaacagcagctggtggacgcctttctcaaattggccaaacagcgttca 839
Seq_2 811 CCCAATGGACAGAAACAGCAGCTGGTGGACGCCTTTCTCAAATTGGCCAAACAGCGTTCA 752
P N G Q K Q Q L V D A F L K L A K Q R S
V C T W G K N K P G I G A V I L K R A K
Seq_1 840 gtctgcacttggggcaagaataagcctggaattggtgcagtgattttaaaaagagccaaa 899
Seq_2 751 GTCTGCACTTGGGGCAAGAATAAGCCTGGAATTGGTGCAGTGATTTTAAAAAGAGCCAAA 692
V C T W G K N K P G I G A V I L K R A K
A L A G Q S V R I C K G P R A V F S R I
Seq_1 900 gccttggctggacagtcagtacgaatctgtaaaggccccagggctgtgttttcccgcatc 959
Seq_2 691 GCCTTGGCTGGACAGTCAGTACGAATCTGTAAAGGCCCCAGGGCTGTGTTTTCCCGCATC 632
A L A G Q S V R I C K G P R A V F S R I
L L L F S L T D S M E D E D A A C G G Q
Seq_1 960 ttgctactgttttcgttgaccgactcaatggaagatgaagacgccgcttgtggaggtcag 1019
Seq_2 631 TTGCTACTGTTTTCGTTGACCGACTCAATGGAAGATGAAGACGCCGCTTGTGGAGGTCAG 572
L L L F S L T D S M E D E D A A C G G Q
G Q L S T V L L V N L G R M E F P S Y T
Seq_1 1020 ggacagctttcaacagtcctgttggtcaacctcggccgaatggagtttcctagttacacc 1079
Seq_2 571 GGACAGCTTTCAACAGTCCTGTTGGTCAACCTCGGCCGAATGGAGTTTCCTAGTTACACC 512
G Q L S T V L L V N L G R M E F P S Y T
I N R K T H I F Q D R D D L I R Y A A A
Seq_1 1080 atcaatcggaaaacccacatcttccaagacagagatgatcttatcagatatgcagcagcc 1139
Seq_2 511 ATCAATCGGAAAACCCACATCTTCCAAGACAGAGATGATCTTATCAGATATGCAGCAGCC 452
I N R K T H I F Q D R D D L I R Y A A A
T H M L S D I S S A M A N G N W E E A K
Seq_1 1140 acgcacatgctgagtgacatttcttccgcaatggccaatgggaactgggaagaagctaag 1199
Seq_2 451 ACGCACATGCTGAGTGACATTTCTTCCGCAATGGCCAATGGGAACTGGGAAGAAGCTAAG 392
T H M L S D I S S A M A N G N W E E A K
E L A Q C A K R D W N R L K N H P S L R
Seq_1 1200 gagctcgctcagtgtgcaaaaagggattggaacagactgaaaaaccacccttctctgaga 1259
Seq_2 391 GAGCTCGCTCAGTGTGCAAAAAGGGATTGGAACAGACTGAAAAACCACCCTTCTCTGAGA 332
E L A Q C A K R D W N R L K N H P S L R

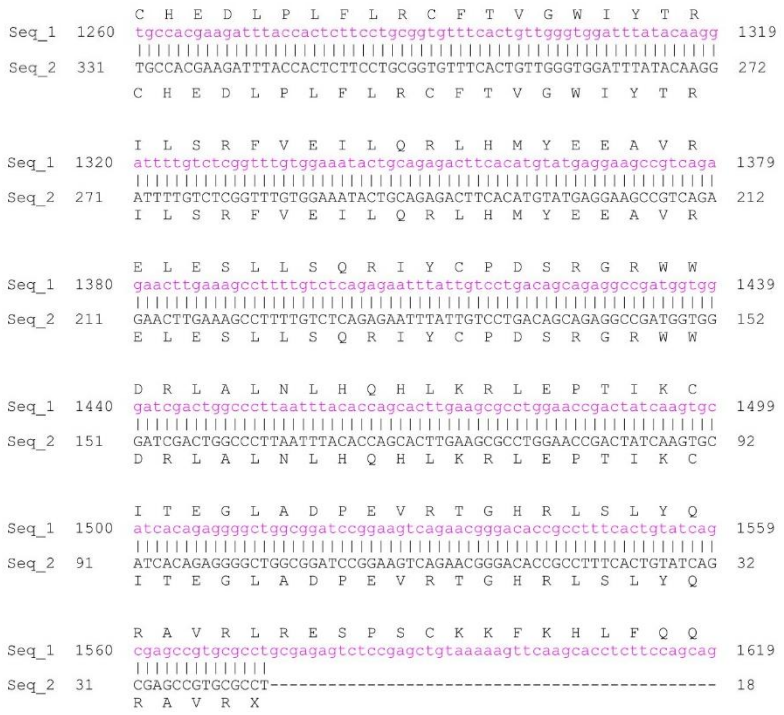


Figure 8.12 FAN1-SUMO internal reverse primer sequencing

Internal reverse primer sequencing of the FAN1-SUMO construct. FAN1 (residues 373-1017) are shown in purple.

Alignment of Sequence_1: [pESumo-hFAN1-373-1017_9-T7-Term[1].seq] with Sequence_2: [pE-SUMOpro Amp_hFAN1-373-1017.txt.xdna]

```
                                                                  D  G  G
Seq_1  1048  ------------------------------------------------CCGATGGTGGG  1038
                                                              ||||||||||
Seq_2  1381  aacttgaaagccttttgtctcagagaatttattgtcctgacagcagaggccgatggtggg  1440
               L  E  S  L  L  S  Q  R  I  Y  C  P  D  S  R  G  R  W  W  D

             F  X  W  P  L  I  X  P  T  P  *  S  A  W  N  R  L  S  S  A
Seq_1  1037  TTCGNCTGGCCCTTAATTTCNCCCACCCCTTGAAGCGCCTGGAACCGACTTTCAAGTGCA  978
              |||:|||||||||||||| : ||| | ||||||||||||||||||||| |||||||||
Seq_2  1441  atcgactggcccttaatttacaccagcacttgaagcgcctggaaccgactatcaagtgca  1500
               R  L  A  L  N  L  H  Q  H  L  K  R  L  E  P  T  I  K  C  I

             X  Q  R  G  W  R  I  R  K  S  E  R  D  P  R  L  S  L  Y  Q
Seq_1  977   TCNCAGAGGGGCTGGCGGATCCGGAAGTCAGAACGGGACCCCCGCCTTTCACTGTATCAG  918
              ||:||||||||||||||||||||||||||||||||||  ||||||||||||||||||||
Seq_2  1501  tcacagaggggctggcggatccggaagtcagaacgggaca-ccgcctttcactgtatcag  1559
               T  E  G  L  A  D  P  E  V  R  T  G  H     R  L  S  L  Y  Q

             R  A  V  R  L  R  E  S  P  S  C  K  K  F  K  H  L  F  Q  Q
Seq_1  917   CGAGCCGTGCGCCTGCGAGAGTCTCCGAGCTGTAAAAAGTTCAAGCACCTCTTCCAGCAG  858
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  1560  cgagccgtgcgcctgcgagagtctccgagctgtaaaaagttcaagcacctcttccagcag  1619
              R  A  V  R  L  R  E  S  P  S  C  K  K  F  K  H  L  F  Q  Q

             L  P  E  M  A  V  Q  D  V  K  H  V  T  I  T  G  R  L  C  P
Seq_1  857   CTCCCAGAAATGGCTGTGCAAGATGTGAAACACGTGACCATCACAGGCAGGCTGTGCCCA  798
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  1620  ctcccagaaatggctgtgcaagatgtgaaacacgtgaccatcacaggcaggctgtgccca  1679
              L  P  E  M  A  V  Q  D  V  K  H  V  T  I  T  G  R  L  C  P

             Q  R  G  M  C  K  S  V  F  V  M  E  A  G  E  A  A  D  P  T
Seq_1  797   CAGCGTGGGATGTGCAAGTCTGTGTTTGTGATGGAGGCCGGGGAGGCCGCTGACCCCACC  738
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  1680  cagcgtgggatgtgcaagtctgtgtttgtgatggaggccggggaggccgctgaccccacc  1739
              Q  R  G  M  C  K  S  V  F  V  M  E  A  G  E  A  A  D  P  T

             T  V  L  C  S  V  E  E  L  A  L  A  H  Y  R  R  S  G  F  D
Seq_1  737   ACGGTCCTGTGCTCTGTGGAGGAGCTGGCACTGGCCCATTACAGACGCAGCGGTTTTGAC  678
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  1740  acggtcctgtgctctgtggaggagctggcactggcccattacagacgcagcggttttgac  1799
              T  V  L  C  S  V  E  E  L  A  L  A  H  Y  R  R  S  G  F  D

             Q  G  I  H  G  E  G  S  T  F  S  T  L  Y  G  L  L  L  W  D
Seq_1  677   CAGGGGATTCATGGCGAAGGGTCCACCTTCAGCACCCTGTATGGCCTCCTCCTGTGGGAC  618
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  1800  cagggattcatggcgaagggtccaccttcagcaccctgtatggcctcctcctgtgggac  1859
              Q  G  I  H  G  E  G  S  T  F  S  T  L  Y  G  L  L  L  W  D

             I  I  F  M  D  G  I  P  D  V  F  R  N  A  C  Q  A  F  P  L
Seq_1  617   ATCATCTTCATGGATGGGATTCCGGATGTCTTCAGAAACGCCTGTCAGGCATTCCCCCTG  558
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  1860  atcatcttcatggatgggattccggatgtcttcagaaacgcctgtcaggcattcccccctg  1919
              I  I  F  M  D  G  I  P  D  V  F  R  N  A  C  Q  A  F  P  L

             D  L  C  T  D  S  F  F  T  S  R  R  P  A  L  E  A  R  L  Q
Seq_1  557   GACTTGTGCACAGACAGCTTCTTCACAAGCAGACGCCCAGCCCTTGAGGCCAGGCTGCAG  498
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  1920  gacttgtgcacagacagcttcttcacaagcagacgcccagcccttgaggccaggctgcag  1979
              D  L  C  T  D  S  F  F  T  S  R  R  P  A  L  E  A  R  L  Q

             L  I  H  D  A  P  E  E  S  L  R  A  W  V  A  A  T  W  H  E
Seq_1  497   CTGATTCATGATGCCCCCGAGGAGAGCCTGCGGGCCTGGGTGGCAGCCACGTGGCATGAG  438
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  1980  ctgattcatgatgcccccgaggagagcctgcgggcctgggtggcagccacgtggcatgag  2039
              L  I  H  D  A  P  E  E  S  L  R  A  W  V  A  A  T  W  H  E

             Q  E  G  R  V  A  S  L  V  S  W  D  R  F  T  S  L  Q  Q  A
Seq_1  437   CAGGAAGGCAGAGTGGCTTCCCTTGTCAGCTGGGATCGCTTCACGTCTCTTCAGCAAGCT  378
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  2040  caggaaggcagagtggcttcccttgtcagctgggatcgcttcacgtctcttcagcaagct  2099
              Q  E  G  R  V  A  S  L  V  S  W  D  R  F  T  S  L  Q  Q  A

             Q  D  L  V  S  C  L  G  G  P  V  L  S  G  V  C  R  H  L  A
Seq_1  377   CAGGATCTTGTCTCCTGCCTGGGGGGCCCTGTGCTCAGTGGTGTGTGCAGGCACCTGGCT  318
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  2100  caggatcttgtctcctgcctgggggccctgtgctcagtggtgtgtgcaggcacctggct  2159
              Q  D  L  V  S  C  L  G  G  P  V  L  S  G  V  C  R  H  L  A

             A  D  F  R  H  C  R  G  G  L  P  D  L  V  V  W  N  S  Q  S
Seq_1  317   GCTGACTTTCGACACTGTCGAGGGGGCCTCCCCGACCTGGTGGTGTGGAACTCCCAGAGC  258
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  2160  gctgactttcgacactgtcgaggggcctccccgacctggtggtgtggaactcccagagc  2219
              A  D  F  R  H  C  R  G  G  L  P  D  L  V  V  W  N  S  Q  S
```

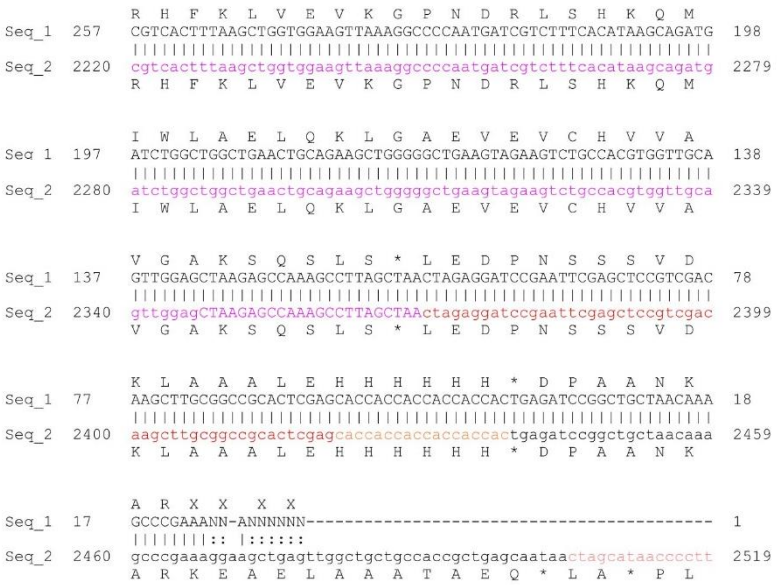


Figure 8.13 FAN1-SUMO reverse primer sequencing

Reverse primer sequencing of the FAN1-SUMO construct. FAN1 (residues 373-1017) are shown in purple; the SUMO and His6 tag are shown in orange.

8.4.3 FAN1-MBP construct sequencing

Alignment of Sequence_1: [pET_His6_MBP_TEV_hFAN1-373-1017.txt[1].xdna] with Sequence_2: [pET-HMT-hFAN1-373-1017 6-T7[1].seq]

```
Seq_1  3961  ggccacgatgcgtccggcgtagaggatcgagatctcgatcccgcgaaattaatacgactc  4020

Seq_2  12    ------------------------------------------------------------  11


                                                T  I  G  R  P  Q  R  F  P  S  S  A  G  S  G  E  L  F  N  *
Seq_1  4021  actataggg------------------------agaccacaacggtttccctctagtgccggctccggagagctctttaattaa  4080
                                                ||||||||||||||||||||||
Seq_2  12    ----------------------------------------TCCGGAGAGCTCTTTAATTAA  32
                                                S  G  E  L  F  N  *

                                                A  A  A  L  Q  D  S  S  S  R  N  N  F  V  *  L  *  E  G  D
Seq_1  4081  gcggccgccctgcaggactcgagttctagaaataattttgtttaactttaagaaggagat  4140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  33    GCGGCCGCCCTGCAGGACTCGAGTTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGAT  92
             A  A  A  L  Q  D  S  S  S  R  N  N  F  V  *  L  *  E  G  D

                                                I  H  M  K  S  S  H  H  H  H  H  H  G  S  S  M  K  I  E  E
Seq_1  4141  atacatatgaaatcttctcaccatcaccatcaccatggttcttctatgaaaatcgaagaa  4200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  93    ATACATATGAAATCTTCTCACCATCACCATCACCATGGTTCTTCTATGAAAATCGAAGAA  152
             I  H  M  K  S  S  H  H  H  H  H  H  G  S  S  M  K  I  E  E

                                                G  K  L  V  I  W  I  N  G  D  K  G  Y  N  G  L  A  E  V  G
Seq_1  4201  ggtaaactggtaatctggattaacggcgataaaggctataacggtctcgctgaagtcggt  4260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  153   GGTAAACTGGTAATCTGGATTAACGGCGATAAAGGCTATAACGGTCTCGCTGAAGTCGGT  212
             G  K  L  V  I  W  I  N  G  D  K  G  Y  N  G  L  A  E  V  G

                                                K  K  F  E  K  D  T  G  I  K  V  T  V  E  H  P  D  K  L  E
Seq_1  4261  aagaaattcgagaaagatacggaattaaagtcaccgttgagcatccggataaactggaa  4320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  213   AAGAAATTCGAGAAAGATACCGGAATTAAAGTCACCGTTGAGCATCCGGATAAACTGGAA  272
             K  K  F  E  K  D  T  G  I  K  V  T  V  E  H  P  D  K  L  E

                                                E  K  F  P  Q  V  A  A  T  G  D  G  P  D  I  I  F  W  A  H
Seq_1  4321  gagaaattcccacaggttgcggcaactggcgatggccctgacattatcttctgggcacac  4380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  273   GAGAAATTCCCACAGGTTGCGGCAACTGGCGATGGCCCTGACATTATCTTCTGGGCACAC  332
             E  K  F  P  Q  V  A  A  T  G  D  G  P  D  I  I  F  W  A  H

                                                D  R  F  G  G  Y  A  Q  S  G  L  L  A  E  I  T  P  D  K  A
Seq_1  4381  gaccgctttggtggctacgctcaatctggcctgttggctgaaatcaccccggacaaagcg  4440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  333   GACCGCTTTGGTGGCTACGCTCAATCTGGCCTGTTGGCTGAAATCACCCCGGACAAAGCG  392
             D  R  F  G  G  Y  A  Q  S  G  L  L  A  E  I  T  P  D  K  A

                                                F  Q  D  K  L  Y  P  F  T  W  D  A  V  R  Y  N  G  K  L  I
Seq_1  4441  ttccaggacaagctgtatccgtttacctgggatgccgtacgttacaacggcaagctgatt  4500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  393   TTCCAGGACAAGCTGTATCCGTTTACCTGGGATGCCGTACGTTACAACGGCAAGCTGATT  452
             F  Q  D  K  L  Y  P  F  T  W  D  A  V  R  Y  N  G  K  L  I

                                                A  Y  P  I  A  V  E  A  L  S  L  I  Y  N  K  D  L  L  P  N
Seq_1  4501  gcttacccgatcgctgttgaagcgttatcgctgatttataacaaagatctgctgccgaac  4560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  453   GCTTACCCGATCGCTGTTGAAGCGTTATCGCTGATTTATAACAAAGATCTGCTGCCGAAC  512
             A  Y  P  I  A  V  E  A  L  S  L  I  Y  N  K  D  L  L  P  N

                                                P  P  K  T  W  E  E  I  P  A  L  D  K  E  L  K  A  K  G  K
Seq_1  4561  ccgccaaaaacctgggaagagatcccggcgctggataaagaactgaaagcgaaaggtaag  4620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  513   CCGCCAAAAACCTGGGAAGAGATCCCGGCGCTGGATAAAGAACTGAAAGCGAAAGGTAAG  572
             P  P  K  T  W  E  E  I  P  A  L  D  K  E  L  K  A  K  G  K

                                                S  A  L  M  F  N  L  Q  E  P  Y  F  T  W  P  L  I  A  A  D
Seq_1  4621  agcgcgctgatgttcaacctgcaagaaccgtacttcacctggccgctgattgctgctgac  4680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  573   AGCGCGCTGATGTTCAACCTGCAAGAACCGTACTTCACCTGGCCGCTGATTGCTGCTGAC  632
             S  A  L  M  F  N  L  Q  E  P  Y  F  T  W  P  L  I  A  A  D
```

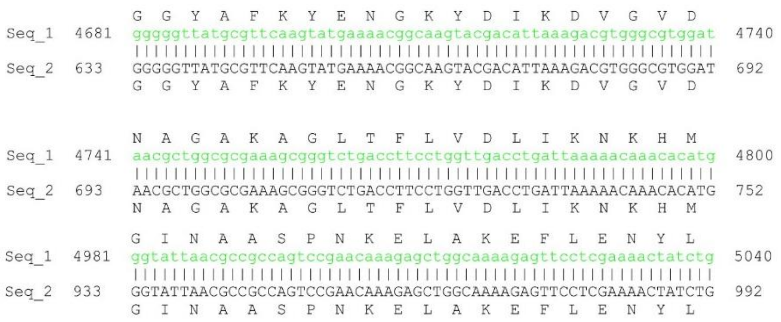


Figure 8.14 FAN1-MBP forward primer sequencing

Forward primer sequencing of the FAN1-MBP construct. FAN1 (residues 373-1017) are shown in purple; the MBP fusion tag is shown in green

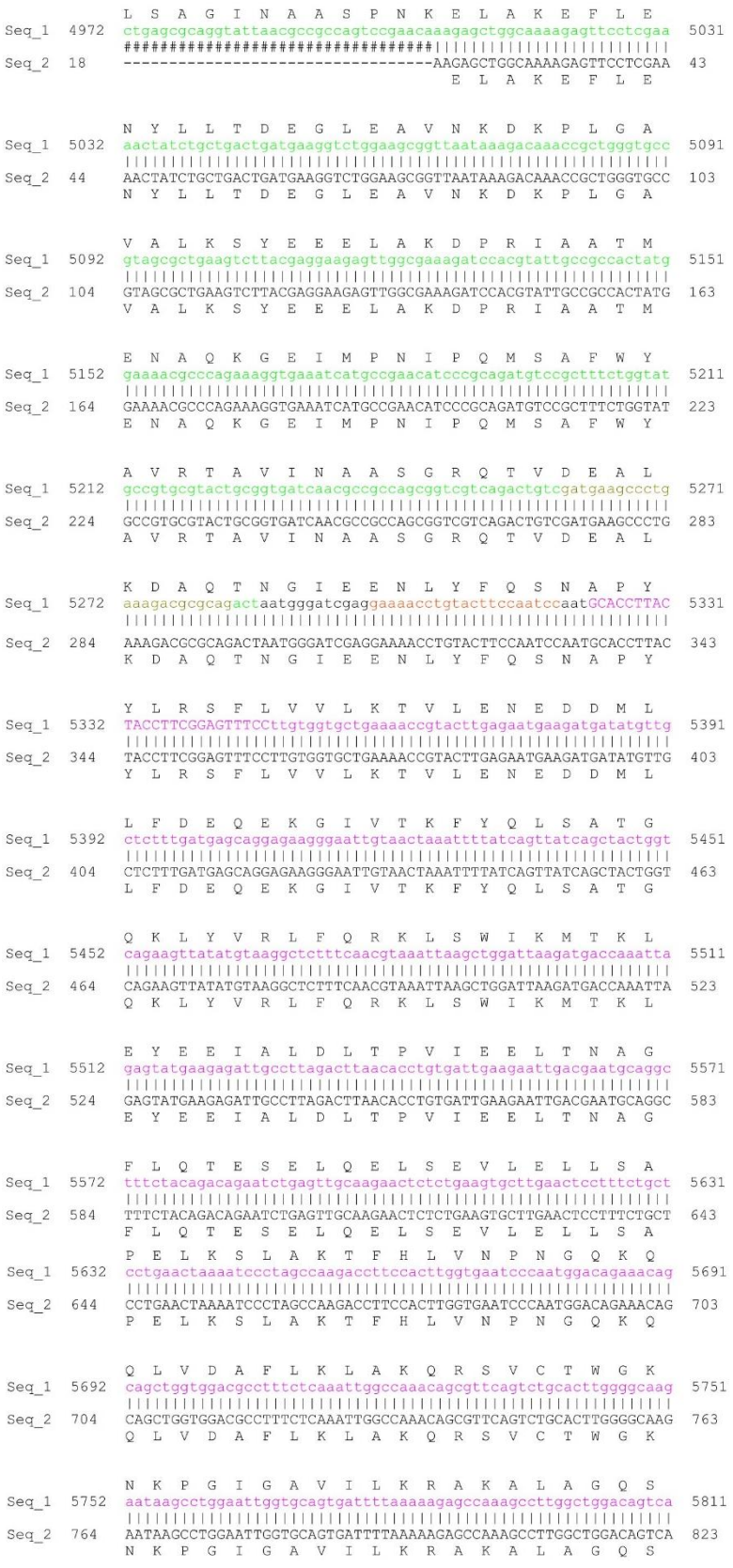


Figure 8.15 FAN1-MBP internal forward primer sequencing

Internal forward primer sequencing of the FAN1-MBP construct. FAN1 (residues 373-1017) are shown in purple; the MBP fusion tag is shown in green. The TEV protease cleavage site shown in orange.

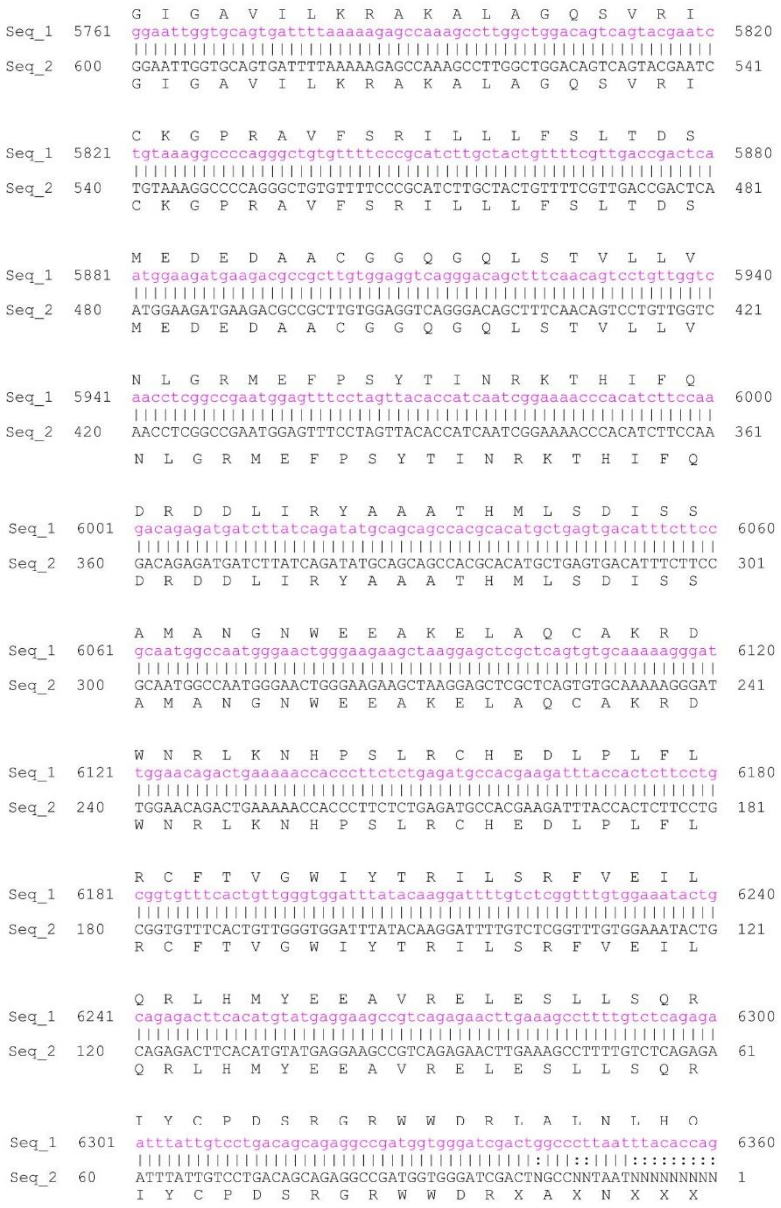
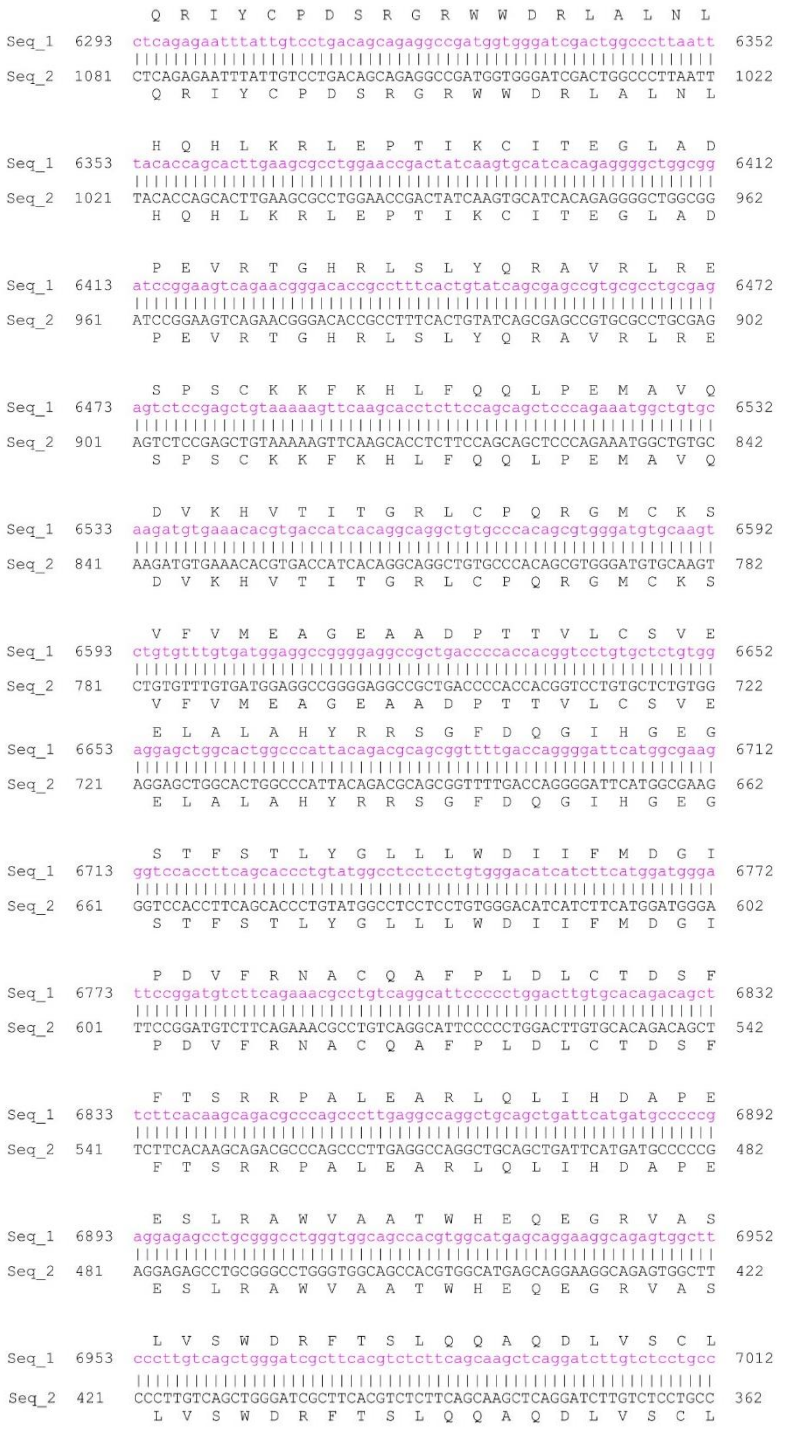


Figure 8.16 FAN1-MBP internal reverse primer sequencing

Internal reverse primer sequencing of the FAN1-MBP construct. FAN1 (residues 373-1017) are shown in purple.


Q R I Y C P D S R G R W W D R L A L N L
Seq_1 6293 ctcagagaatttattgtcctgacagcagaggccgatggtgggatcgactggcccttaatt 6352
Seq_2 1081 CTCAGAGAATTTATTGTCCTGACAGCAGAGGCCGATGGTGGGATCGACTGGCCCTTAATT 1022
Q R I Y C P D S R G R W W D R L A L N L
H Q H L K R L E P T I K C I T E G L A D
Seq_1 6353 tacaccagcacttgaagcgcctggaaccgactatcaagtgcatcacagaggggctggcgg 6412
Seq_2 1021 TACACCAGCACTTGAAGCGCCTGGAACCGACTATCAAGTGCATCACAGAGGGGCTGGCGG 962
H Q H L K R L E P T I K C I T E G L A D
P E V R T G H R L S L Y Q R A V R L R E
Seq_1 6413 atccggaagtcagaacgggacaccgcctttcactgtatcagcgagccgtgcgcctgcgag 6472
Seq_2 961 ATCCGGAAGTCAGAACGGGACACCGCCTTTCACTGTATCAGCGAGCCGTGCGCCTGCGAG 902
P E V R T G H R L S L Y Q R A V R L R E
S P S C K K F K H L F Q Q L P E M A V Q
Seq_1 6473 agtctccgagctgtaaaaagttcaagcacctcttccagcagctcccagaaatggctgtgc 6532
Seq_2 901 AGTCTCCGAGCTGTAAAAAGTTCAAGCACCTCTTCCAGCAGCTCCCAGAAATGGCTGTGC 842
S P S C K K F K H L F Q Q L P E M A V Q
D V K H V T I T G R L C P Q R G M C K S
Seq_1 6533 aagatgtgaaacacgtgaccatcacaggcaggctgtgcccacagcgtgggatgtgcaagt 6592
Seq_2 841 AAGATGTGAAACACGTGACCATCACAGGCAGGCTGTGCCCACAGCGTGGGATGTGCAAGT 782
D V K H V T I T G R L C P Q R G M C K S
V F V M E A G E A A D P T T V L C S V E
Seq_1 6593 ctgtgtttgtgatggaggccggggaggccgctgaccccaccacggtcctgtgctctgtgg 6652
Seq_2 781 CTGTGTTTGTGATGGAGGCCGGGGAGGCCGCTGACCCCACCACGGTCCTGTGCTCTGTGG 722
V F V M E A G E A A D P T T V L C S V E
E L A L A H Y R R S G F D Q G I H G E G
Seq_1 6653 aggagctggcactggcccattacagacgcagcggttttgaccaggggattcatggcgaag 6712
Seq_2 721 AGGAGCTGGCACTGGCCCATTACAGACGCAGCGGTTTTGACCAGGGGATTCATGGCGAAG 662
E L A L A H Y R R S G F D Q G I H G E G
S T F S T L Y G L L L W D I I F M D G I
Seq_1 6713 ggtccaccttcagcaccctgtatggcctcctcctgtgggacatcatcttcatggatggga 6772
Seq_2 661 GGTCCACCTTCAGCACCCTGTATGGCCTCCTCCTGTGGGACATCATCTTCATGGATGGGA 602
S T F S T L Y G L L L W D I I F M D G I
P D V F R N A C Q A F P L D L C T D S F
Seq_1 6773 ttccggatgtcttcagaaacgcctgtcaggcattccccctggacttgtgcacagacagct 6832
Seq_2 601 TTCCGGATGTCTTCAGAAACGCCTGTCAGGCATTCCCCCTGGACTTGTGCACAGACAGCT 542
P D V F R N A C Q A F P L D L C T D S F
F T S R R P A L E A R L Q L I H D A P E
Seq_1 6833 tcttcacaagcagacgcccagcccttgaggccaggctgcagctgattcatgatgccccg 6892
Seq_2 541 TCTTCACAAGCAGACGCCCAGCCCTTGAGGCCAGGCTGCAGCTGATTCATGATGCCCCCG 482
F T S R R P A L E A R L Q L I H D A P E
E S L R A W V A A T W H E Q E G R V A S
Seq_1 6893 aggagagcctgcgggcctgggtggcagccacgtggcatgagcaggaaggcagagtggctt 6952
Seq_2 481 AGGAGAGCCTGCGGGCCTGGGTGGCAGCCACGTGGCATGAGCAGGAAGGCAGAGTGGCTT 422
E S L R A W V A A T W H E Q E G R V A S
L V S W D R F T S L Q Q A Q D L V S C L
Seq_1 6953 cccttgtcagctgggatcgcttcacgtctcttcagcaagctcaggatcttgtctcctgcc 7012
Seq_2 421 CCCTTGTCAGCTGGGATCGCTTCACGTCTCTTCAGCAAGCTCAGGATCTTGTCTCCTGCC 362
L V S W D R F T S L Q Q A Q D L V S C L

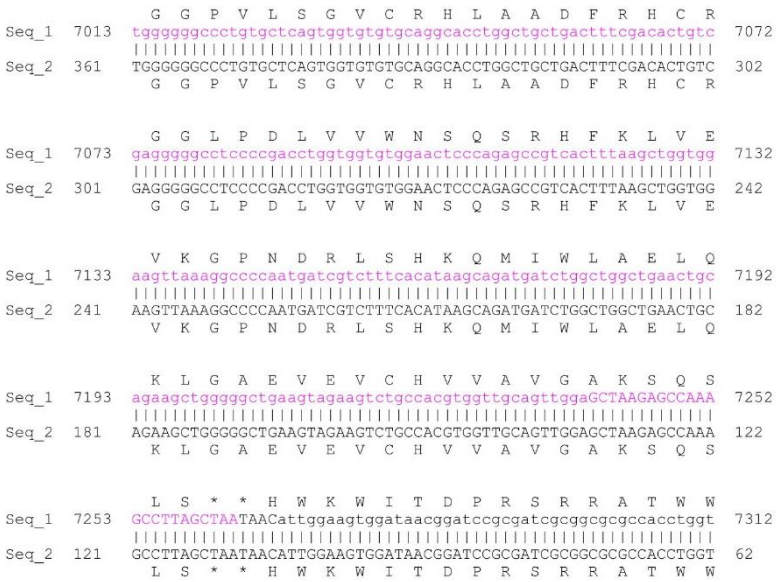


Figure 8.17 FAN-MBP reverse primer sequencing

Reverse primer sequencing of the FAN1-MBP construct. FAN1 (residues 373-1017) are shown in purple.

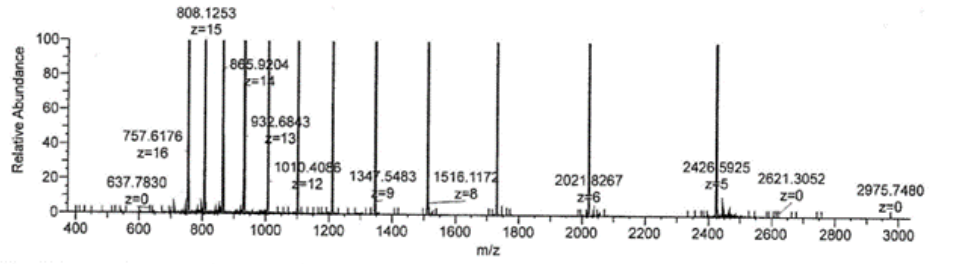
8.5 Quench flow kinetics table

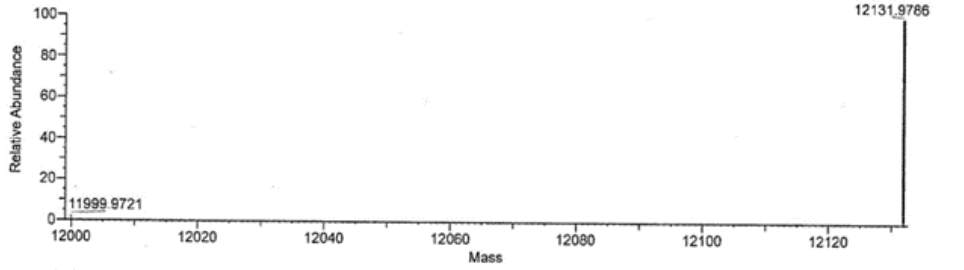
Table 16 Timepoints and RQF quench flow parameters for H2TH mutant single turnover kinetics

Loop	Mode (C/I)	Switch position	Push 1	Push 2	Delay	Time point (ms)	10 ×
1	c	2	160	-	-	12.1	-
2	c	6	180	-	-	9.1	-
2	c	3	180	-	-	19.4	-
4	c	5	255	-	-	30.6	-
4	c	3	255	-	-	57.5	-
4	c	1	255	-	-	124.2	-
4	I	4	160	120	200	240.8	-
4	I	4	160	120	800	840.8	-
4	I	4	160	120	3200	3240.8	-
4	I	4	160	120	1280	12840.8	10x
4	I	4	160	120	2560	25640.8	10x
4	I	4	160	120	5120	51240.8	10x

8.6 Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (FTMS) of FAN1 oligonucleotides

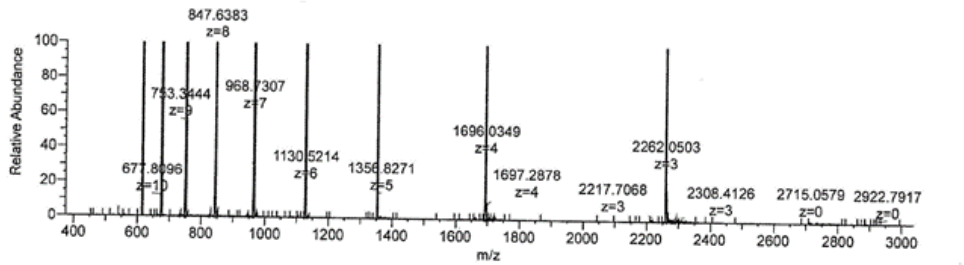
TEMPLATE-NL

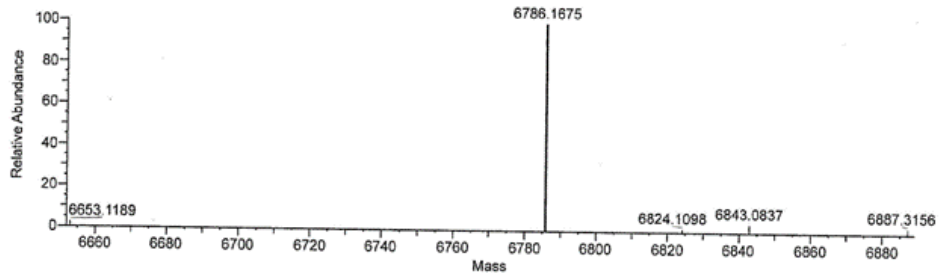




No.	Monoisotopic...	Sum Intensity	Number of Ch...	Average Charge	Delta Mass	Relative Abun...	Fractional Ab...	RT Range
2	12131.9786	47572386.28	12	12.795855986 1414	132.0065	100	97.339490630 1951	7.815
1	11999.9721	1300261.37	4	14.064771328 5785	0	2.7332271337 9767	2.6605093698 0493	7.815

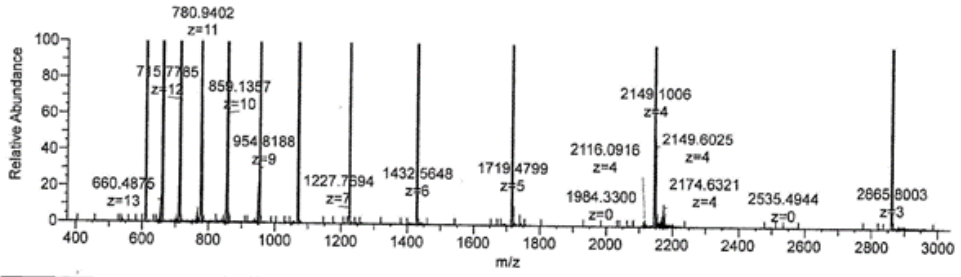
pB0-NL

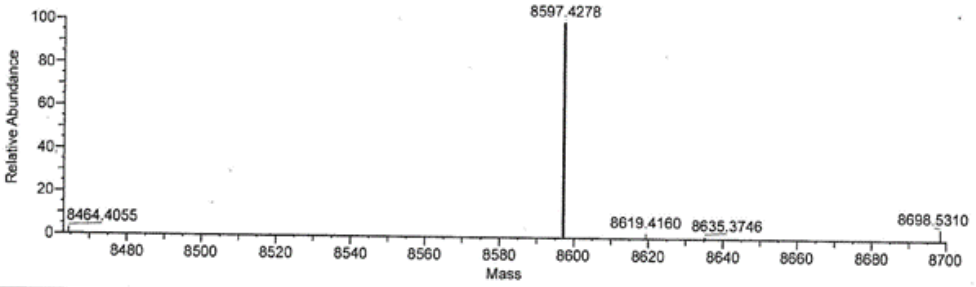




No.	Monoisotopic...	Sum Intensity	Number of Ch...	Average Charge	Delta Mass	Relative Abun...	Fractional Ab...	RT Range
2	6786.1675	92028490.95	9	6.1795128832 9833	133.04859999 9999	100	89.488078682 5975	3.917
4	6843.0837	3981118.42	8	6.0332361989 1526	189.9648	4.3259629478 9076	3.8712211265 885	3.917
5	6887.3156	2772829.78	3	3.3970736366 4041	234.1967	3.0130123306 1238	2.6962868451 3477	3.917
1	6653.1189	2201439.67	5	6.6008423858 9267	0	2.3921284020 5765	2.1406697466 2211	3.917
3	6824.1098	1854948.77	5	5.4896022742 1165	170.9909	2.0156244559 1747	1.8037435990 5711	3.917

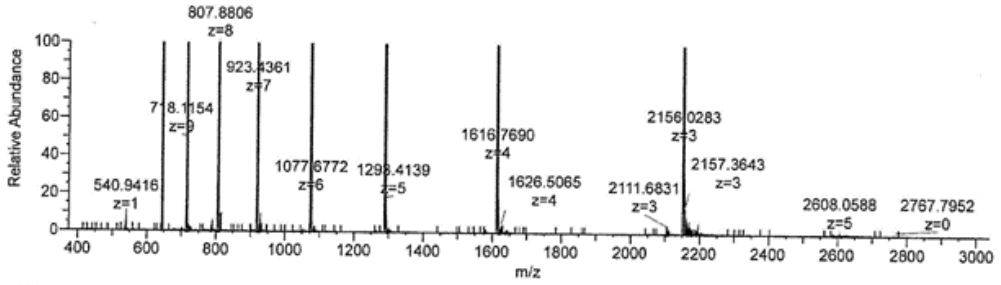
3'T_8-NL

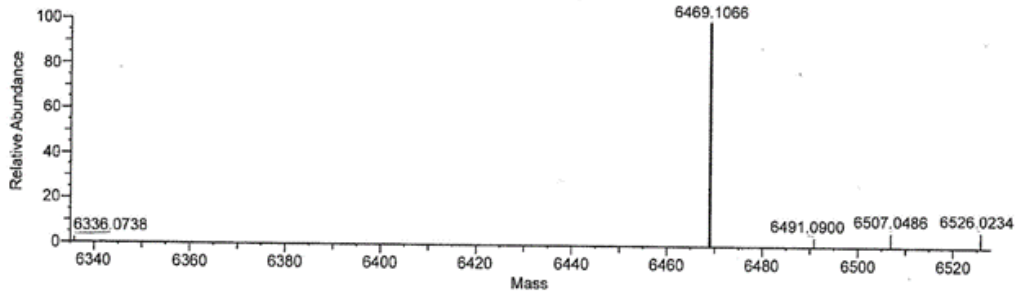




No.	Monoisotopic...	Sum Intensity	Number of Ch...	Average Charge	Delta Mass	Relative Abun...	Fractional Ab...	RT Range
2	8597.4278	145068807.85	12	7.9117210863 4458	133.0222999 9999	100	88.709907779 6636	6.781
5	8698.531	8149996.06	3	4.0154555781 8341	234.1255	5.6180209796 9057	4.9837412301 2566	6.781
1	8464.4055	4300987.77	7	7.4704399659 7598	0	2.9647915590 8359	2.6300638579 2231	6.781
3	8619.416	3555572.11	3	4.5973393966 5624	155.0104999 9999	2.4509556276 7458	2.1742404770 306	6.781
4	8635.3746	2456322.22	3	6.6393217241 6833	170.9690999 9998	1.6932118326 4966	1.5020466552 5787	6.781

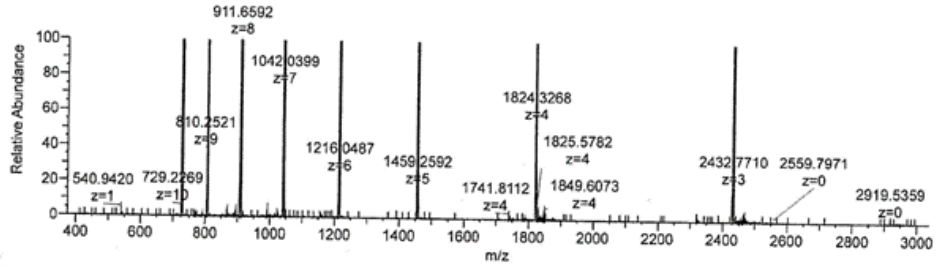
3'T_1-NL

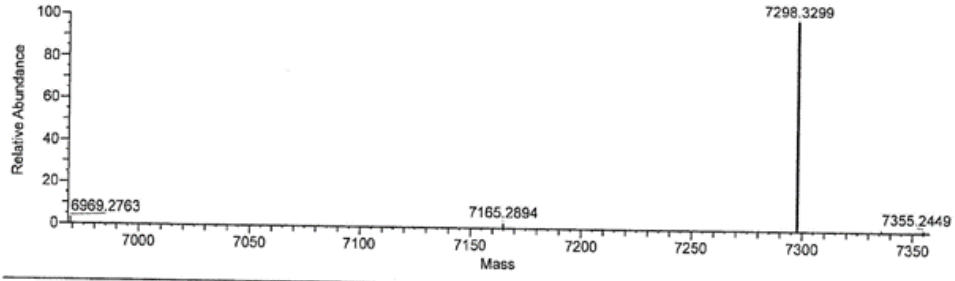




No.	Monoisotopic...	Sum Intensity	Number of Ch...	Average Charge	Delta Mass	Relative Abun...	Fractional Ab...	RT Range
2	6469.1066	105539105.08	8	5.7183194838 9102	133.0328	100	82.560542757 4164	3.943
5	6526.0234	7604045.95	7	5.5012618182 3237	189.9496	7.2049558732 1499	5.9484506743 5864	3.943
4	6507.0486	7032219.62	7	5.3520307455 721	170.9748	6.6631412258 7025	5.5011255607 7164	3.943
3	6491.09	4782679.66	6	3.7779758519 0832	155.0162	4.5316659226 6882	3.7413679817 0826	3.943
1	6336.0738	2874327.67	6	5.3940054389 2902	0	2.7234717101 5068	2.2485130257 4509	3.943

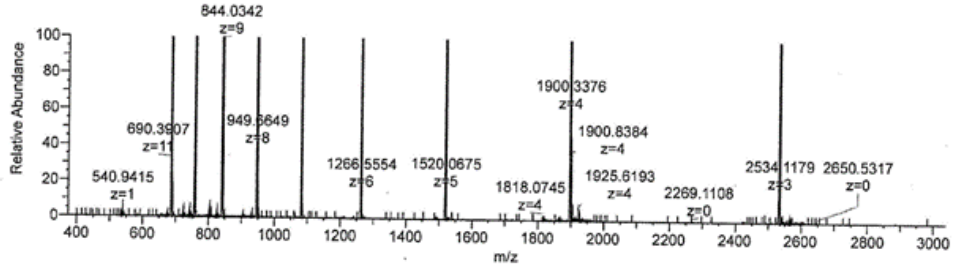
3’FAM-pB0

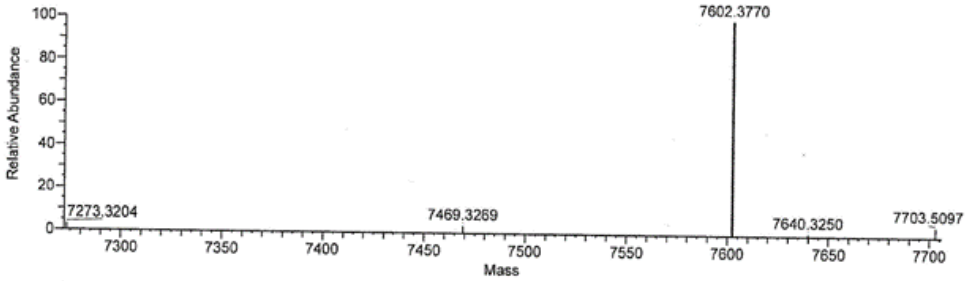




No.	Monoisotopic...	Sum Intensity	Number of Ch...	Average Charge	Delta Mass	Relative Abun...	Fractional Ab...	RT Range
3	7298.3299	60731660.7	8	6.0255852124 6495	329.0535999 9999	100	91.464843032 3295	7.496
1	6969.2763	2033083.4	6	5.5975582914 6949	0	3.3476499350 8567	3.0619227583 9799	7.496
2	7165.2894	1686380.56	5	6.0576375623 0171	196.0130999 9999	2.7767733346 373	2.5397713718 8959	7.496
5	7355.2449	1051091.64	4	4.6717643435 5583	385.9685999 9999	1.7307144706 4842	1.5829952739 1639	7.496
4	7336.2725	896695.77	5	5.3897960817 2476	366.9962	1.4764881441 8144	1.3504675634 6651	7.496

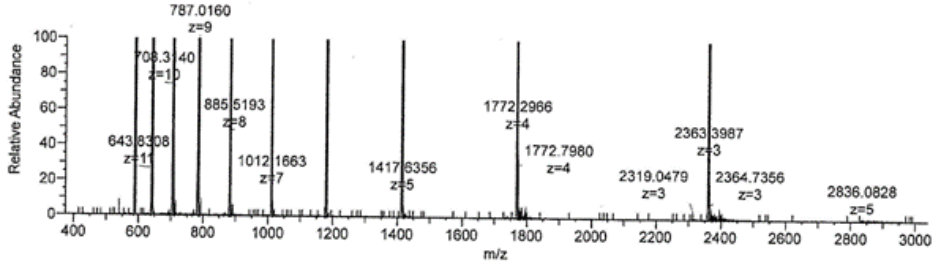
3’FAM-pB1

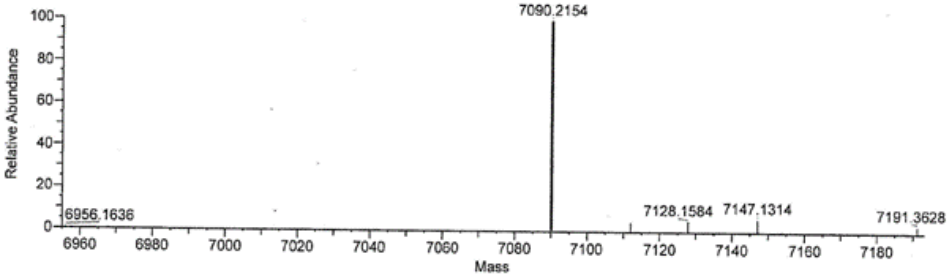




No.	Monoisotopic...	Sum Intensity	Number of Ch...	Average Charge	Delta Mass	Relative Abun...	Fractional Ab...	RT Range
3	7602.377	43415749.21	9	6.8786921051 321	329.0566000 0001	100	87.897650101 4421	7.735
6	7703.5097	2130358.35	3	3.7120980987 1597	430.1893	4.9068791596 6981	4.3130314746 6715	7.735
2	7469.3269	1616163.65	8	6.9980406607 0667	196.0065	3.7225285280 2952	3.2720151004 9375	7.735
1	7273.3204	1319824.94	6	6.6394250804 6563	0	3.0399681314 1717	2.6720605513 4841	7.735
4	7640.325	606510.2	4	5.2866712046 6921	367.0046	1.3969819962 4827	1.2279143470 4244	7.735
5	7658.2876	304920.28	4	9.3934516655 181	384.9672	0.7023264265 81088	0.6173284250 06206	7.735

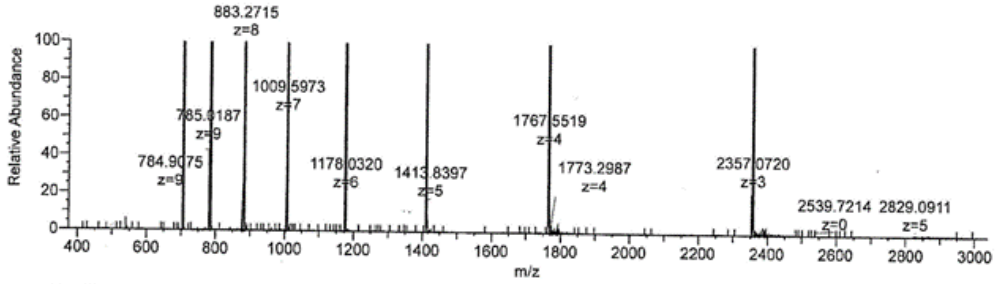
pB1-NL

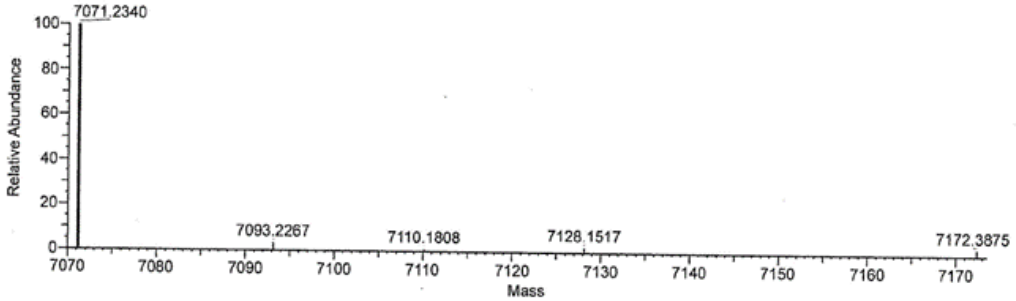




No.	Monoisotopic...	Sum Intensity	Number of Ch...	Average Charge	Delta Mass	Relative Abun...	Fractional Ab...	RT Range
2	7090.2154	115285201.93	10	6.4745663353513	134.0518	100	82.7712107333152	4.074
5	7147.1314	7507335.58	9	6.1523334200 4206	190.9678	6.5119681054 6281	5.3900348434589	4.074
4	7128.1584	6076796.6	8	5.8675701609 0445	171.9948	5.2710985436 7065	4.3629520835 4234	4.074
3	7112.1976	5417734.87	5	3.9587283166 5403	156.0340000 0001	4.6994191616 1069	3.8897661374 9858	4.074
6	7191.3628	4213956.23	3	3.4266423928 0381	235.1992	3.6552446970 2423	3.0254902909 9226	4.074
1	6956.1636	780738.23	4	9.3187919366 4968	0	0.6772232835 86784	0.5605459111 92694	4.074

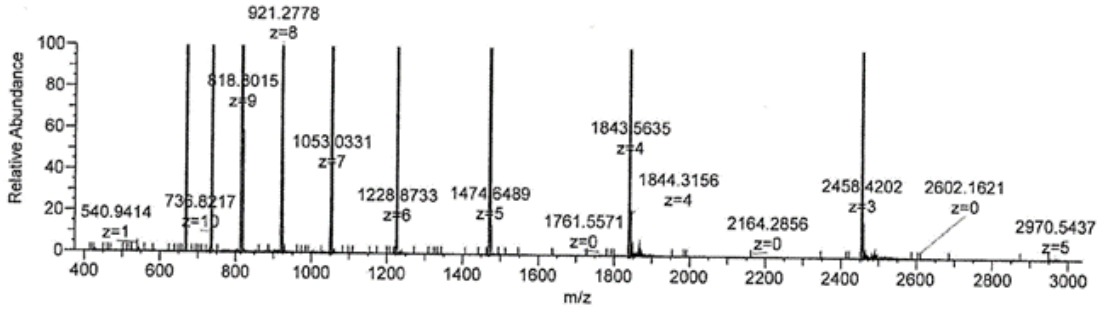
5FAM-pB0

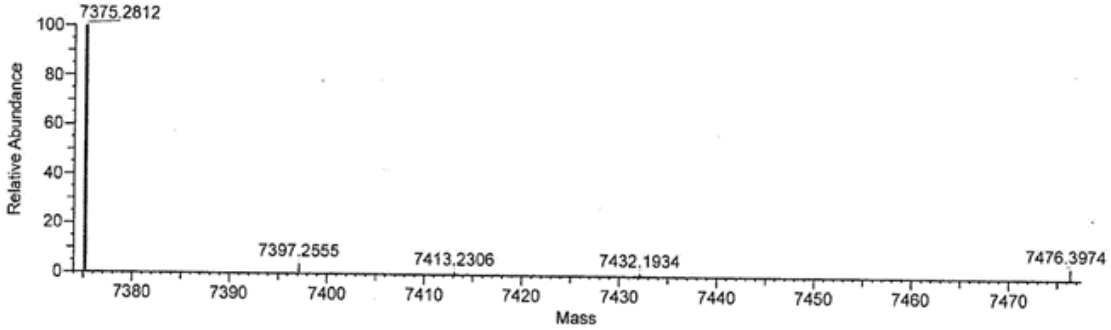




No.	Monoisotopic...	Sum Intensity	Number of Ch...	Average Charge	Delta Mass	Relative Abun...	Fractional Ab...	RT Range
1	7071.234	70303553.34	8	6.1167065310 8833	0	100	89.5682710295169	6.625
2	7093.2267	2555122.65	5	3.8750524936 4967	21.992699999 9998	3.6344146612 9456	3.2552823741 6481	6.625
3	7110.1808	1210423.52	6	5.2130236483 2894	38.946799999 9997	1.7217103012 5061	1.5421061489 6726	6.625
4	7128.1517	1993413.64	7	5.6218361114 1606	56.9177	2.8354379619 4697	2.5396527586 3047	6.625
5	7172.3875	2429069.34	3	3.5240607765 7857	101.15349999 9999	3.4551160284 2122	3.0946876887 2059	6.625

5FAM-pB1





No.	Monoisotopic...	Sum Intensity	Number of Ch...	Average Charge	Delta Mass	Relative Abun...	Fractional Ab...	RT Range
1	7375.2812	74126429.78	9	6.2300083660504	0	100	89.1728395487956	7.346
2	7397.2555	3126784.57	4	3.84556715301412	21.9742999999999	4.21817775290135	3.76146887947771	7.346
3	7413.2306	1035642.18	5	5.25076592593821	37.9493999999995	1.39712944906922	1.24586000190747	7.346
4	7432.1934	1276271.68	4	5.88770136063016	56.9121999999998	1.72174983172379	1.53533321487471	7.346
5	7476.3974	3561561.66	3	3.56827425275271	101.116199999999	4.80471226062063	4.28449835494454	7.346