Thesis submitted for the degree of Doctor of Philosophy to the Department of Chemistry, University of Sheffield

Investigating the Effects of Migrating Substrates, Proliferating Cell Nuclear Antigen and Arch Mutations on Human Flap Endonuclease-1 Activity

Supervisor: Prof. Jane Grasby

Nur Nazihah Binti Md Shahari

August 2018

Declaration

This thesis is genuinely prepared by the author which has not been submitted as a complete or in partial, for other degree except where specific reference been done to other sources.

Nur Nazihah Md Shahari

Publication

Bennet, Ian A.; Finger, L David; Baxter, Nicola J.; Ambrose, Benjamin; Hounslow, Andrea M.; Thompson, Mark J.; Exell, Jack C.; **Shahari, Nur Nazihah B Md**.; Craggs, Timothy D.; Waltho, Jonathan P.; Grasby, Jane A. Regional conformational flexibility couples substrate specificity and scissile phosphate diester selectivity in human flap endonuclease 1Nucleic Acids Research 2018, 46 (11), 5618-5633.

Acknowledgements

In the name of Allah, the infinitely Compassionate and Merciful

I would like to highlight my highest gratitude especially to my supervisor Professor. Jane Grasby for her guidance and being a supportive person throughout my PhD journey. Without her patience, caring and knowledge in this area I would not at this stage today. I also would like to express much appreciation to all my research colleagues who are always being helpful and cooperative in my PhD accomplishments; Dr. David Finger (in design of migrating flap substrates), Dr. Mark Thompson (in kinetic studies on hFEN1 mutants), Dr. Ian Bennet, Dr. Sana Algasaier, Dr. Steven Shaw and Rebecca Ley. We didn't realize we were making memories; we just knew we were having fun and encouraging each other as a team.

My feeling of grateful to my beloved family; my parents and siblings for their continuous support and pray. Every year I've forced them to visit and stay in England for at least 3 months (they really don't like the weather). My wholeheartedly thanks especially to my beloved husband; Khairil Fitri who is one of my greatest blessings from Allah. The one who wiped my tears, hugged me tight, watch me succeed, seen me fail, kept me strong and always sacrifice everything for our small family. I'm a lucky wife! And I am hugely indebted to an exceptional give from Allah, princess Miza Aqilah and prince Mirza Aqil. Thank you for their endless supporting, inspiration and always understanding during this process. If mama have to choose between loving my children and breathing, I would use my last breath to tell Aqilah and Aqil… Mama love both of you so much!

Lastly, highly appreciation to my study sponsorship; Public Service Department of Malaysia for their funding and support. I would like to thank my employer; Chemistry Department of Malaysia for giving me this great opportunity in my life.

After 4 years I have been here, eventually I have managed to complete my PhD thesis which I believe it's a great experience and gaining new knowledge for me toward my personal achievement for myself and to my country. Thank you very much to everyone who has involved from the beginning until the completion of this PhD.

Abstract

Flap endonucleases (FENs), in conjunction with proliferating cell nuclear antigen (PCNA), catalyse the removal of single stranded 5'-DNA or RNA protrusions known as flaps. These bifurcated structures, which are the result of polymerase strand displacement activity during lagging strand DNA synthesis and repair, are migrating structures *in-vivo* due to sequence complementarity. *In-vitro* studies of FEN1 activity frequently use the optimal FEN1 substrate known as the static double-flap (DF), which consists of a single nucleotide 3' and a 5' flap of any length including zero. Moreover, the static DF substrates have 3' or 5' flaps that are not complementary to the template strand to reduce conformational variability. To understand better the mechanism of human FEN1 in a true biological context, nine migrating DF substrates were designed to determine the effect of increasing substrate-conformational complexity on the rates of FEN1 phosphate diester hydrolysis under multiple- and single-turnover conditions. The multiple-turnover analyses reveal that human FEN1 produces a single 5' flap product and nicked DNA even on equilibrating double flap substrates that have many potential conformers. This shows that hFEN1 action always recognises and hydrolyses a single conformer bearing a single 3'-nucleotide flap. The k_{cat}/K_M and k_{STmax} conditions showed generally that the rate of reaction was inversely proportional to the number of potential conformers. Due to the decrease in hydrolytic efficiency with complex migrating flaps, whether PCNA could assist FEN1 in the hydrolysis of such substrates was investigated. Surprisingly, we found that PCNA did not stimulate the reactions of FEN1 with static or migrating double or single flaps.

During the FEN1 reaction 5'-flaps are threaded through a hole in the protein known as the helical arch, which covers the active site. Threading is thought to occur while the arch is disordered, but when substrate is present it is believed that the arch forms its helical shape to catalyse the reaction. We investigated the role of conserved arch residues by mutation of basic amino acids to alanine to remove the positive charge or by using glycine mutations to destabilise the helices. Evaluation of the rates of reaction and the ability to stabilise the threaded state was undertaken by adapting a streptavidin capture assay. All mutants had some impact on the rate of reaction, but relatively small deficiencies in threading were revealed in the mutants. The ability of arch mutants to transfer the substrate to the active site were assessed using an ECCD assay with tandem 2-aminopurine containing substrates. The data show that the arch must be fully ordered to promote the distortion of the DNA

required for active site transfer and subsequent hydrolysis. Basic arch residues are required to position the DNA for hydrolysis, with α5 residues producing the largest impacts.

Abbreviations

K_M	Michaelis constant
k_{ST}	single-turnover rate constant
k_{cat}	catalytic rate constant
2-AP	2-aminopurine
aa	amino acid
Ala / A	alanine
Arg / R	arginine
Asp / D	aspartic acid
BER	base excision repair
bp	base pair
BSA	bovine serum albumin
CD	circular dichroism
Cm	chloramphenicol
DDR	DNA damage repair
dHPLC	denaturing high pressure liquid chromatography
DMSO	dimethyl sulfoxide
DNAP	DNA polymerase
dNTP	deoxynucleoside triphosphate
double flap	double flap
DR	direct repair
dRP	deoxyribose phosphate
ds	double stranded
DTT	dithiothreitol
E	enzyme, macromolecule
ECCD	exciton coupling circular dichroism
EDTA	ethylenediaminetetraacetic acid
EMSA	electrophoretic mobility shift assay
EP	enzyme-product
ES	enzyme-substrate
ESI	enzyme-substrate-inhibitor complex
ETDM	electronic transition dipole moments

FA	fluorescence anisotropy
FAM	fluorescein
FRET	fluorescence resonance energy transfer
HEPES	hydroxyethylpiperazineethane sulphonic acid
His / H	histidine
HJ	holliday junction
IDCL	Inter-domain connector loop
IMAC	immobilised metal ion affinity chromatography
IPTG	isopropyl β D-I-thiogalactopyranoside
ITC	isothermal titration calorimetry
Kan	kanamycin
LB	Luria-Bertani broth
Leu / L	leucine
LFER	linear free energy relationship
lp-BER	long-patch base excision repair
Lys / K	lysine
MMR	mismatch repair
NER	nucleotide excision repair
nt(s)	nucleotide(s)
OD	optical density
P, Q	product
PARP1	poly (ADP-ribose) polymerase 1
PCNA	proliferating cell nuclear antigen
PCR	polymerase chain reaction
Phe / F	phenylalanine
Pro / P	proline
pY	pseudo Y
RRB	reduced reaction buffer
RT	room temperature
SA	streptavidin
SDS-PAGE	sodium dodecyl sulphate- polyacrylamide gel electrophoresis
SOC	super optimal broth with catabolite repression

sp-BER	short-patch base excision repair
ss	single stranded
ST	single-turnover
TBE	trisborateEDTA
TBSE	tris borate saline EDTA
Tris	tris(Hydroxymethyl)aminomethane
Tyr	tyrosine
WT	wild-type

Contents

Chapter 1: Introduction

1.1 Deoxyribonucleic acid

Most organisms including humans, animals, fungi, protists and bacteria carry genetic information in the form of Deoxyribo-Nucleic-Acid (DNA). In the 1950s, Chargaff discovered that the four components of DNA cytosine, C, adenine, A, guanine, G or thymine, T[1] vary in different types of organism but that the amount of dA was always equal to the amount of dT with a similar relationship between dG and dC[2]. Later, investigations into the structure of DNA by X-ray diffraction revealed its helical nature. Watson and Crick continued the study of DNA structure using modelling methods and proposed that DNA structure consisted of two antiparallel strands and twisted around each other. As shown in **Figure 1.1**, Watson-Crick were first proposed the pairing of polynucleotide strands DNA[1].

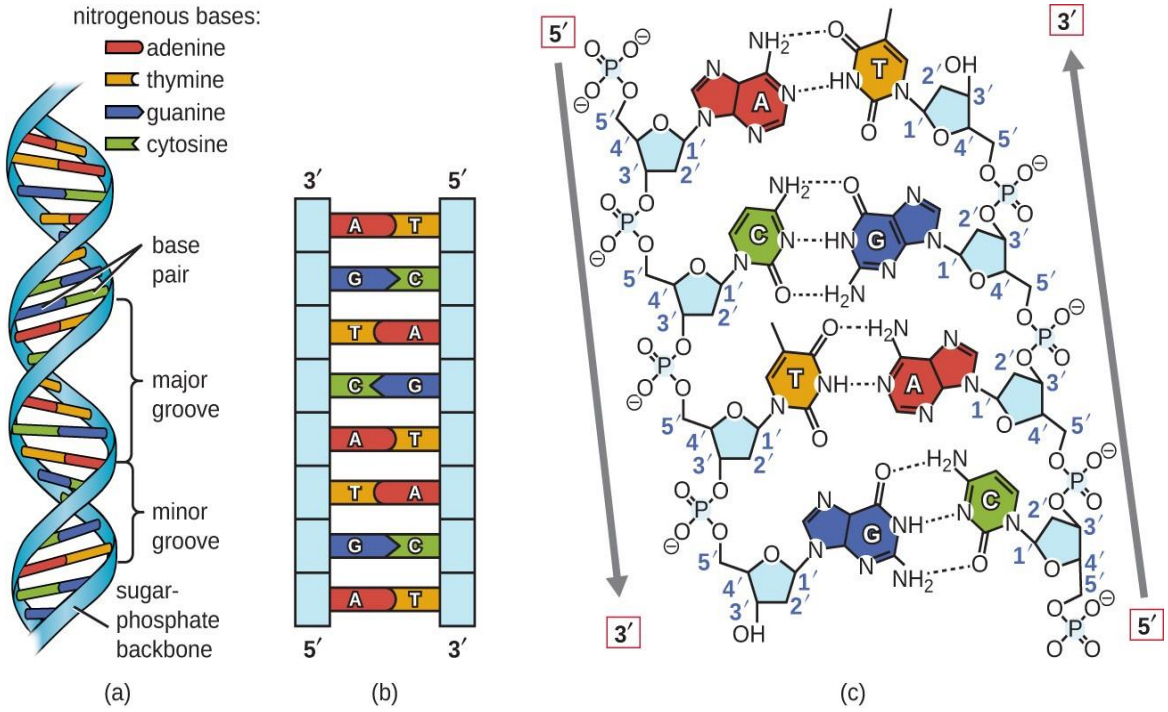

Figure 1.1: *The Watson and Crick model for DNA structure. (a) The rungs are form by pyrimidines-purine base pairs with the outside of the double helix composed of the sugar-phosphate backbones. (b) Antiparallel of two DNA strands. (c) The carbon numbering in pentose sugar indicates the direction of each strand*[3].

The double helix is formed from two strands or two polynucleotide chains intertwined to form the DNA duplex structure. Monomers consist of nucleic acids known as nucleotides that are the basis to construct polynucleotides. There are three components of nucleotides; a phosphate group, four nitrogenous heterocyclic bases (either A, C, G and T) and a pentose sugar. Each DNA strand has either 5'-3' or 3'-5' direction because of the structure linked by phosphodiester bond between 5'-OH group of one nucleotide and the 3'-OH of the neighbouring nucleotide[1,4,5].

The backbone of DNA is constructed of alternating phosphate and sugar. The sugar has the nitrogenous bases attached at 1'-carbon and is called as 2'-deoxy-D-ribose because of the hydroxyl group missing at 2'-carbon. Nitrogenous bases that carry genetic information are either monocyclic pyrimidines (T, C and uracil for RNA) or bicyclic purines (A and G). A forms two hydrogen bonds with T (in DNA) and U (in RNA), while three hydrogen bonds form between C and G. DNA is comprised of two antiparallel strands in which the deoxyribonucleotide repeat units are linked by esterification of phosphate to the 5'-hydroxyl of one sugar and the 3'-hydroxyl of the next sugar to construct the polymeric chain. The DNA sugar-phosphate backbone is incredibly stable[4,5].

In Ribo-nucleic-Acid (RNA) the sugar is D-ribose since it has the hydroxyl group at 2' position. RNA is a single-stranded molecule, which also carries genetic information by copying and transferring the data from DNA to the synthesis of protein and convey the genetic in into another chemical form. The difference between RNA and DNA is that whereas RNA contains uracil, DNA uses thymine as one of the two pyrimidines. Uracil and thymine have similar structure but thymine has a 5-methyl group on its ring. Multiple non-covalent hydrogen bonds form to complementary nitrogenous bases with specific preferences. **Figure 1.2** shows the differences between DNA and RNA[4,5].

These different features between DNA and RNA contribute to their functions. Generally both (DNA and RNA) are responsible for carrying the genetic information, however the DNA molecule is more stable because of the 2'-deoxyribose sugar. Whereas the hydroxyl group on the second carbon in RNA can attack the phosphodiester linkage and reduces the stability of the backbone[5].

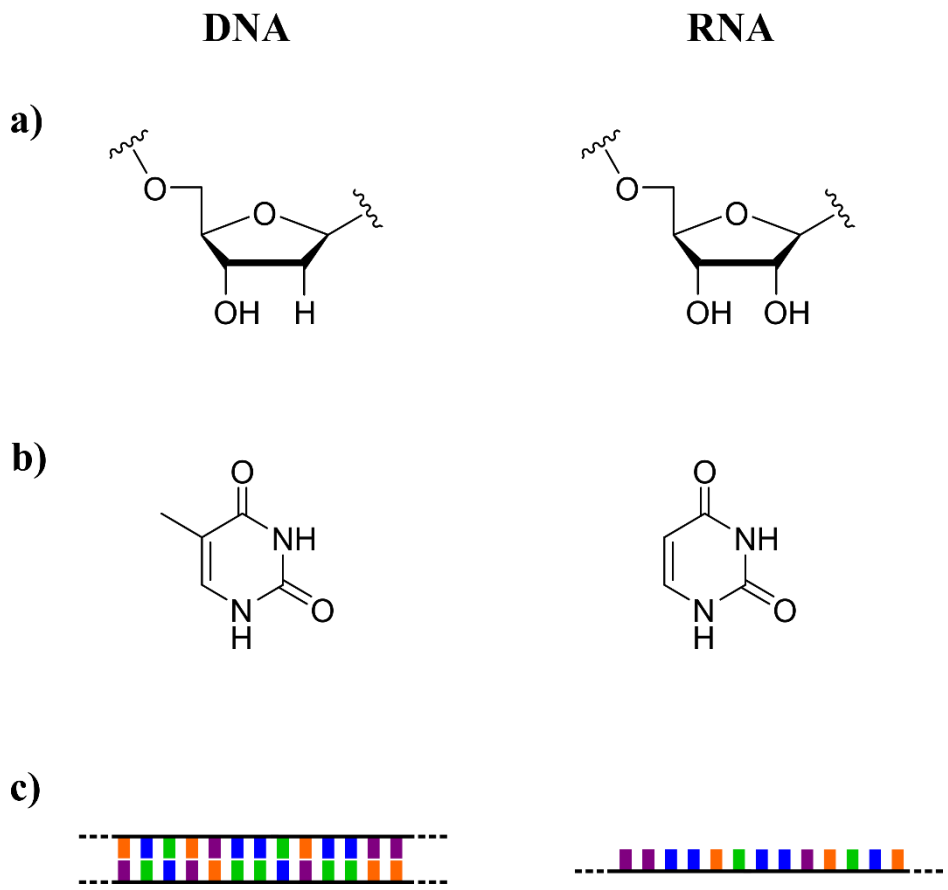

Figure 1.2: *The three differences between DNA and RNA. a) The second carbon; the DNA has hydrogen group and the RNA has hydroxyl group at the sugar. b) The DNA has thymine (methyl group), the RNA has uracil at the nitrogen bases. c) The DNA has double-stranded, the RNA has single stranded.*

1.2 Phosphodiester hydrolysis

Phosphodiester hydrolysis of nucleic acids in cells is catalysed by enzymes called nucleases. The nucleases found in eukaryotic organisms are some of the most important enzymes in the cell. They catalyse the hydrolysis of the inter-nucleotide phosphodiester bonds of DNA or RNA and are essential for DNA replication, gene expression, and DNA repair. There are two groups of nucleases; endonucleases cleaving within the DNA or RNA strand and exonucleases cleaving at end of the nucleic acid strand. Enzymes have their specific structure and pathway to cleave including the active site, the structure of protein and the position of the substrate binding site[6–9].

The phosphodiester linkage is the backbone of DNA and is extremely stable under aqueous conditions[10]. However, this bond needs to be hydrolysed due to important roles in biological processes, for example the removal of an RNA primer during DNA replication[11]. The intermolecular nucleophilic attack of water molecules or hydroxide on phosphorus can cleave either at 3' to produce a 3'-phosphate and 5'-OH products or at 5' and generate product terminating 3'-OH and 5'-phosphate. However, the preferred nucleophile attack generates a at 5'-monophosphate where the 3'-OH can be used for other DNA reactions while the 5'-phosphate can be further used as a substrate in DNA metabolism[12] as can be seen in **Figure 1.3**.

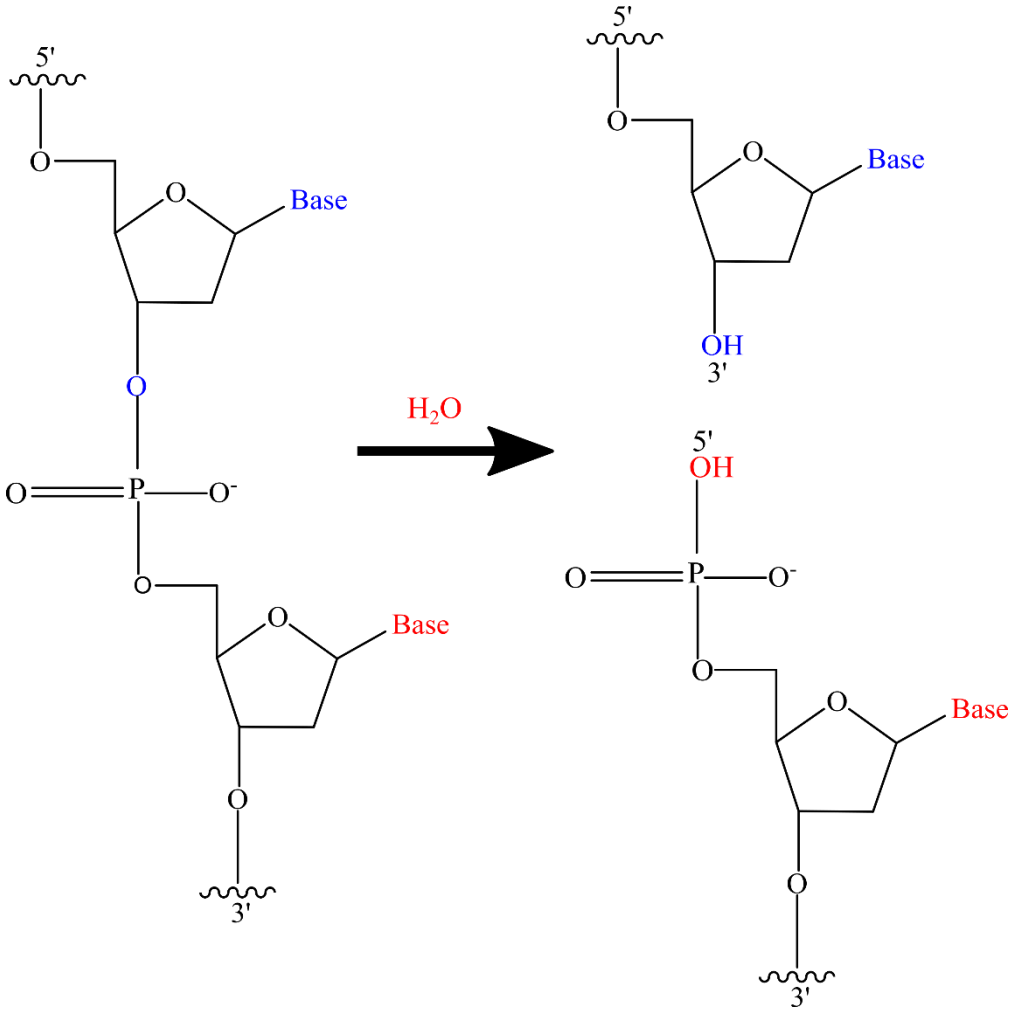

Figure 1.3: *Phosphodiester hydrolysis producing a 3'-OH and 5'-phosphate.*

1.3 Flap endonuclease-1

Flap endonuclease-1 (FEN1) enzymes are present in all living things and catalyse phosphodiester hydrolysis during DNA replication and repair[13]. FEN1s are structure-specific endonucleases that also have a 5'-3' exonuclease activity on double-stranded DNA[14]. They were first discovered and called DNAse IV as an enzyme[15] that demonstrated specific binding and activity at 5'-flap structures[16]. It is a member of the *Rad2* structure-specific nuclease family[17] and also originally known as *maturation factor* (MF1) and responsible to complete lagging strand replication[18]. Observing *in-vitro* studies of *simian virus 40 DNA*, it was proposed that the FEN1 is capable of removing the RNA moiety of the Okazaki fragment yielding nicked DNA that can be ligated by Lig1[19].

In eukaryotes, FEN1 acts as both an endonuclease and exonuclease and is reported to play important roles in biological processes such as Okazaki fragment maturation[20], apoptotic DNA fragmentation, telomere maintenance, stalled replication fork rescue, and long-patch base excision repair. Studies on FEN1 expanded rapidly because of the complex structures and properties of FEN1. Genetic and biochemical studies suggested that an interaction between FEN1 and other proteins involved in metabolic pathways such as proliferating cell nuclear antigen (PCNA)[21], will authorize miscellaneous activities of FEN1.

There are many crystal structures of FEN1 from wide range of organisms including yeast[22], humans[9,17,23], thermophilic bacteria[24,25], bacteriophage[26–28] and archaea[24,29,30]. The first FEN X-ray crystal structure was published without the DNA substrate in 1995 using the FEN domain from *Thermus aquaticus (Taq)* DNA polymerase. Following this structures from bacteriophage T5FEN, T4RNaseH and the flap endonucleases from *Pyrococcus furiosus* (*Pf*), *Methanococcus jannaschii* (*Mj*), *Pyrococcus horikoschii* (*Ph*) and *Homo sapiens* (*H*) were reported. Later in 2004, the first structure to be co-crystallised with DNA bound was *Archaeoglobus fulgidus* (*Af*), although the DNA was only part but not the whole of a substrate. Later the structure of T4RNaseH, the T4 FEN, bound to DNA substrate but without active site metal ions was published.

1.4 The biological role of FEN1

FEN1 plays an important role in intracellular processes including the maintenance of cells such as telomere maintenance and suppression of trinucleotide repeats[31]. Flap endonucleases are nucleases which have the capability to catalyse the structure-specific endonucleolytic and exonucleolytic[32] hydrolysis of nucleic acids to give 5'-monophosphate and a 3'-hydroxyl products. They are essential for genomic stability and DNA processing at the key steps of metabolism including DNA replication and repair[33].

1.4.1 DNA replication

DNA replication occurs rapidly and effectively in eukaryotic and prokaryotic cells. Before the division happens, cells have to make exact copies of their DNA through replication and reproduction processes. The accuracy of this process is vital for the next generation to have the same genetic instructions as the parents. The DNA structure assists the efficiency of the replication processes. Various enzymes are involved during the replication process including pol-δ/Ɛ, helicase, pol-α, FEN1, DNA lig-1, and PCNA. In the beginning, the double strands of DNA unwind forming two parent strands to act as a template[34,35]. Complementary base pairing of nucleotides to this parent template guides formation of the new double strands. To form the new strand, nucleotides will be adding to the 3' since the direction of DNA replication occurs in 5'-3' direction[35].

E.coli have been used as a model organism (in the beginning) to understand the DNA replication processes, then followed with more complex eukaryotic organisms. Once the replication fork is created, various enzymes are required for replication to begin. The double helix is unwound by helicases by breaking the hydrogen bonds between two polynucleotides to form two new single stranded regions. The presence of topoisomerases are important to prevent the double strand (outside of the replication fork) forming supercoils. DNA polymerase will take part once replication fork is stable by adding new nucleotides on the parent template to lengthen the new strands, also known as daughter strands.

The model of DNA replication can be seen at **Figure 1.4**. After the DNA double helix is unzipped, both strands will act as a template for the synthesis of a new strand. Primase is responsible for the synthesis of the short RNA primer by polymerising complementary ribonucleotides. The DNA polymerase pol-α extends the synthesis and forms the RNA-DNA primer (~20 – 25 base pairs)[36]. Polymerisation process moves continuously by pol Ɛ, whereas pol-δ syntheses the new strand as fragments. Polymerisation of both strands occurs concurrently with a unidirectional nature of DNA synthesis since new nucleotides are added to the 3' end. The consequence of this is that, one of the daughter strand will grow continuously in the 5'-3' direction; this is called the leading strand. The other strand, called the lagging strand, undergoes discontinuous DNA synthesis growing in the 3'-5' direction with replication fork movement[34,37].

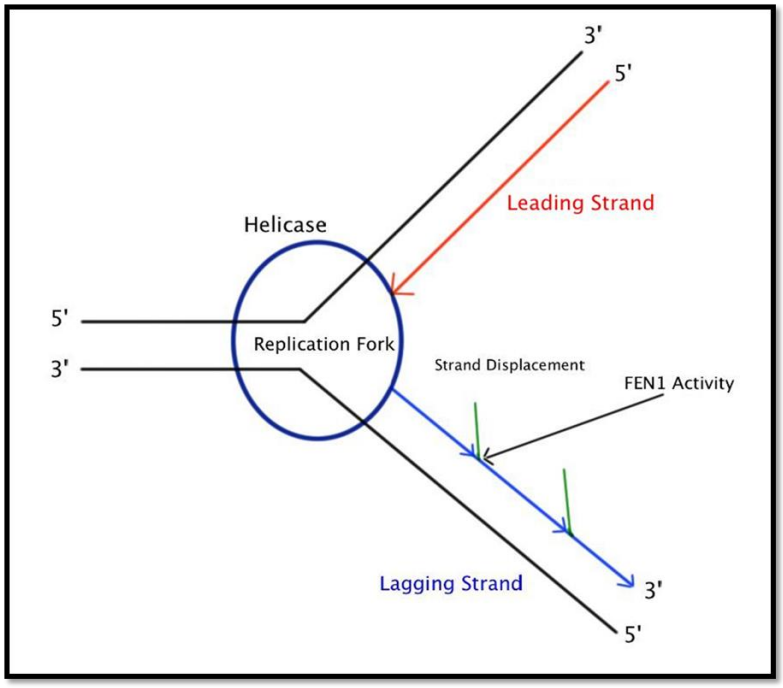

Figure 1.4: *The model of DNA replication fork. The helicase unzipped the DNA and split into a leading and lagging strand. The leading strand will polymerase continuously, whereas the lagging strand has discontinuous DNA synthesis.*

On the lagging strand, the addition of new nucleotides actually occur in the 5'-3' direction in short fragments; these are called Okazaki fragments. Short sequences of nucleotides will be added to the primer by DNA polymerase-α and then these are extended by polymerase-δ to form Okazaki fragments. These steps continue until the complete strand is replicated.

In eukaryotes polymerase-δ catalyses strand displacement synthesis; that is the extension of one Okazaki fragment displaces the 5'-end (primer) of the next Okazaki fragment form the template DNA as a single-stranded flap[38]. The flap or primers must be removed precisely to leave a nicked structure that is easily ligated by DNA ligase. Precision is required to prevent gaps or overlaps in the structure, where a DNA repair mechanism will be required. FEN1 and RNaseH are responsible for removing the flap or primers. New identical copies of the DNA appear, once the replication of leading and lagging strands have been completed[39].

This unique process demands a complicated mechanism to prevent accidental genomic lesions. Previous studies revealed many mechanisms to synthesise a new daughter strand at the lagging strand. One study proposed that when the pol-δ meets the 5'-end from previous Okazaki fragment, it may continue the synthesis by replacing the mature primer with new DNA. Then Dna2 protein and the RPA continue cleaving the flap, follow with FEN1 cleaves the flap and forms nicked DNA duplex and finally, DNA Lig-1 seals the nick[40] to produce a new DNA daughter strand[35,37].

1.4.2 Okazaki fragment

Replication of double stranded DNA is a critical process to produce efficient and precise genome duplication. Two antiparallel strands will form; lagging and leading strands which have discontinuous and continuous replication respectively because of the direction of polymerization occurs in a 5'-3' direction. Therefore, to continue the synthesis at the lagging strand, RNA primers are essential which will be extended initially by DNA polymerase alpha. Circular plasmid and *Simian virus 40* (SV40) studies, proved that FEN1 acted as an enzyme to degrade the RNA primers in Okazaki fragments[41]. In one model it was suggested the RNaseH will remove the RNA primers and leave a residual ribonucleotide upstream at the RNA-DNA junction. The PCNA, encircles the DNA duplex at the 3'-terminus and the polymerization continues with pol-δ producing un-annealed 5'-flap. FEN1 eliminates the 5'flap formed by strand displacement synthesis leaving a nick for DNA Lig-1 to form an intact DNA strand[34].

Studies have proven that, FEN1 plays an important role during Okazaki fragment maturation by removing the flapped DNA. Double flap DNA molecules are the most common substrate structure; these two-way junction DNAs have a one nucleotide 3'-flap on the upstream duplex and a 5'-flap of various length of nucleotides protruding from the downstream duplex. The upstream and downstream can be seen in **Figure 1.5**. In double-stranded DNA, with the anti-parallel nature of DNA, the upstream is one nucleotide 3' flap and the downstream has the 5' flap.

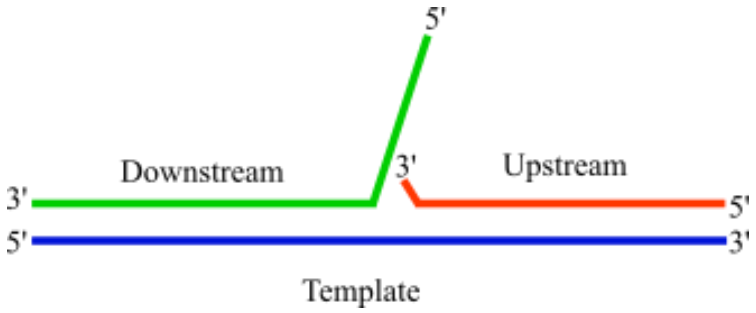

Figure 1.5: *Double flap DNA. The green is downstream, red is upstream and blue is the template of the DNA.*

1.4.3 Base excision repair

Studies have revealed that lesions in DNA occur from alkylation, oxidation, deamination and depurinatiation or depyrimidination damage. Base excision repair (BER) is a common and pivotal biological processes in DNA repair capable of removing modified bases that occur. The DNA glycosylases, an AP endonuclease, DNA polymerase and DNA ligase are four proteins that essential for BER processes. The whole host of auxiliary proteins are important for the correctness and effectiveness of the cells' repair processes[33,42].

Figure 1.6 shows the diagram for BER process. Firstly, the DNA glycosylase will acknowledge the damaged bases by hydrolysing the glycosidic bond between sugar and the base removing it and as a result creating an abasic (AP) site. A number of glycosylases are able to function at specific types of DNA damage and generate the same AP site that will be used by the handful of downstream enzymes. Then an AP lyase or AP endonuclease

will hydrolyse the DNA backbone and a new nucleotide can be inserted by polymerase-β. The AP lyase (an activity present in glycosylase) will cleave the phosphodiester bond 3' of the AP site and generate the 3'-unsaturated aldehydic sugar and a 5'-phosphate end. Then, AP endonuclease incises on the 5'-side at the 3'-unsaturated aldehydic sugar to generate a 3'-OH and 5'-deoxyribose phosphate (dRP)[43]. DNA pol-β will replace the dRP at 5'-side and finally ligation will occur by DNA ligase[43–45].

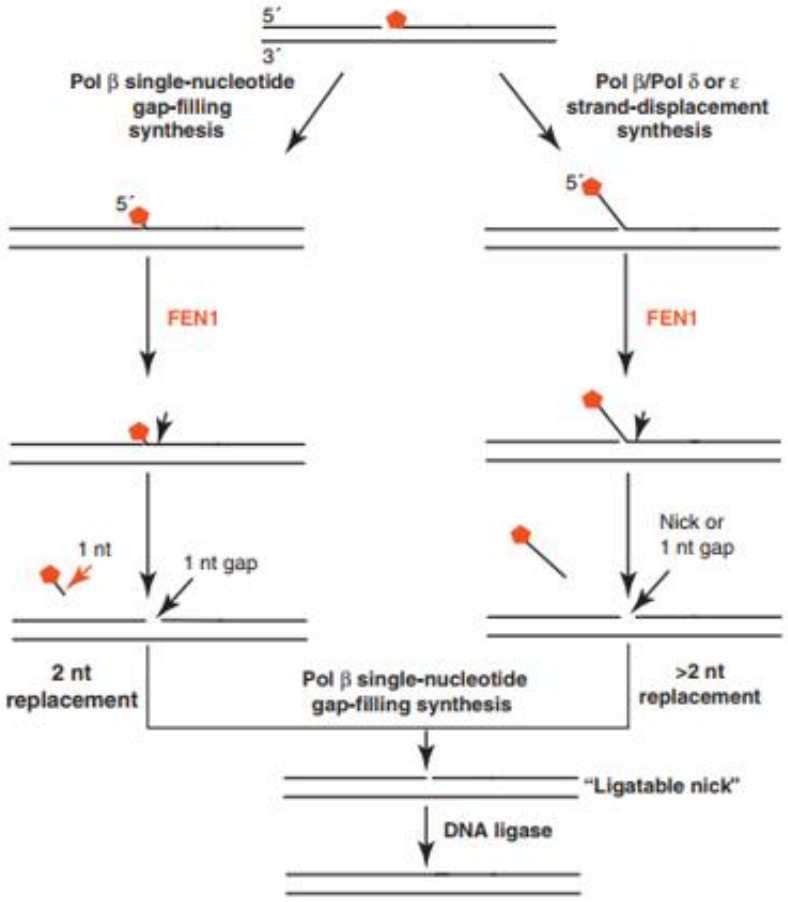

Figure 1.6: *The function of FEN1 in BER during the DNA repair processes*[43].

Diverse factors that influence whether the long or short patch BER pathways are used include the level of cell dividing, the stage of cell cycle and the type of lesion. Once the electrophilic centre of the aldehyde is off-track, the switch of sp-BER and lp-BER will occur. In sp-BER, pol-β has a lyase activity in addition to its polymerase activity. While in lp-BER many proteins are recruited at the AP site including pol-β, pol-δ, PCNA, DNA

ligase and FEN1. In lp-BER, FEN1 will cooperate with pol-β through the 'hit-and-run' mechanism or capturing the 5'-flap associated with a modified sugar created by pol-β or pol-δ/Ɛ to remove a modified sugar. This process will continue in multi-nucleotide replacement where the FEN1 cleavage results in a ligatable nick and DNA ligase 1 joining the DNA strand[44,45].

1.4.4 The miscellaneous activities of FEN1

For optimal activity during DNA replication, DNA repair, the cell cycle and apoptosis processes, studies have shown that FEN1 interacts or acts in concert with other proteins including the proliferating cell nuclear antigen (PCNA)[46], Dna2[47], Apurinic/apyrimidinic (AP) endonuclease 1[48], Replication Protein A (RPA)[49], Werner syndrome ATP-dependent helicase (WRN)[48,50,51] and p300[52]. By interacting with the C-terminal portion of FEN1 it is claimed that the WRN protein can stimulate the FEN1 cleavage of branch-migrating double-flap structures and increase the FEN1 activity during Okazaki fragment maturation. In contrast, it is suggested that p300 inhibits FEN1 endonuclease activity[52], impairing flap cleavage by catalysing the acetylation of FEN1 *in-vivo* and *in-vitro*. The role of RPA is to block the single-stranded DNA from forming secondary structures and it is claimed to be able to stimulate FEN1 cleavage on short lengths and inhibit FEN1 action at longer flaps[49]. AP endonuclease is suggested to interact with FEN1 when it acts with FEN1 in the DNA BER pathway to repair the damage and create a nick at the phosphodiester backbone.

The biological importance of FEN1 is underscored by several lines of evidence. FEN1 is mandatory to upregulated proliferation in cancer cells and could be a possible target for therapeutic drugs. In cancer cells, FEN1 is responsible for DNA repair. If there ae errors in FEN1 mediated DNA repair, cell proliferation increase. Therefore, inhibition of FEN1 will kill off the cancer cells faster than normal cells in a similar way to chemotherapy. Using FEN1 as a target for treatment has advantages and disadvantages, including dose – related toxicity to the host[53].

Knockouts of FEN1 homologues in animals demonstrated that FEN1 is required for the organism to survive. In mice studies, a heterozygous knockout exhibited some considerable tumour growth, while a homozygous knockout of FEN1 was embryonically lethal; this indicates that FEN1 has an essential role in genomic stability[54]. Moreover, FEN1 deletions resulted in defects in DNA processing pathways, increased susceptibility to mutagens, trinucleotide expansion and telomeric destabilisation. In DT40 chicken cells, viability in maintained despite homozygous knockout, but cells are more sensitive to DNA damaging agents[55].

In other research using *Saccharomyces cerevisiae*, null mutants of FEN (Δrad27) exhibit a complex mutator phenotype and temperature sensitive growth defects. Additionally, FEN1 is obligatory in DNA repair processes based on the experiments with the homologue of FEN1 (Rad27). This research showed the mutants of Rad27 were sensitive to the DNA alkylation agent, methyl methane sulphonate. Furthermore, studies in yeast demonstrated FEN1 mutation caused a rise in the rate of spontaneous mutation, showed a high frequency of chromosome recombination and also increased inter/intra-chromosomal recombination. The mutations became worse when merged with a nuclease deficiency mutation in the 3'-5' exonuclease domain of pol-δ[56–58].

Recently, there has been interest in the investigation of the specific association between two important proteins involved in DNA repair and replication[11] processes; FEN1 and PCNA, the processivity factor for DNA polymerases δ and ε[59]. Previous studies proposed that *in-vivo* PCNA recruits FEN1 to branched DNA substrates near the replication fork[30] because FEN1 interacts with a hydrophobic cleft on the front face of the PCNA[30,60].

1.5 FEN1 superfamily

The FEN1 superfamily have been found in bacteriophage viruses T4 and T5. By nature, these viruses have minimalist self-contained biological units. Consequently it is believed that the FEN1 superfamily are some of the most important proteins for DNA replication, repair and recombination processes. Studies of other enzymes from mouse and human cells revealed that other members of the 5'-nuclease superfamily have specific nucleolytic activity such as; human FEN1, exonuclease-1 (EXO1)[17], gap endonuclease-1 (GEN1)[61]

and xeroderma pigmentosum complementation group G (XPG)[32]. Although the FEN1 superfamily have similar structural features and use divalent metal ions for DNA hydrolysis, each of the superfamily members have high specificity with substrate structure and tight regulation of function. The differences in structural characteristics of the FEN1 superfamily are important to explain the substrate recognition. However, generally the FEN1 superfamily have ability to recognise the 5'-duplex end and hydrolyse the phosphodiester bond[6].

Several key features of FEN1 superfamily members have been preserved throughout evolutionary time. These include the acidic active site metal coordinating residues which are supported by the conserved positively charged sidechains to catalyse the reaction. Most FEN1 proteins (bacteria, yeast, archaea and mammals, but not bacteriophage) preserve a 3'flap recognising loop and all have a helical arch (I domain) over the active site. In addition, some of the FEN1 superfamily have extremely different I domain to FEN1, while other regions (N and C-domain) are conserved features. By sequence homology and X-ray structures, each of these enzymes are able to recognise DNAs junction and have three separate binding sites for DNA[9]. All of these FEN1 superfamily members incise DNA one base pair into the duplex region of substrate[9].

Crystal studies have shown that the nuclease core domain of human FEN1 (hFEN1) folds into seven-stranded β-sheets, surrounded by 15 α-helices that form a long positive charged groove. hFEN1 consists of three main regions, intermediate (I), N and C terminal domains. The intermediate domain is responsible for connecting the terminal N and C to produce the saddle structure. Full length hFEN1 is comprised of 380 amino acids including the helical arch or cap, the helix-two-turn-helix, the hydrophobic wedge and the 3'flap binding pocket[62] and others. A crystallographic study between hFEN1 and other organisms, shows that the archaeal FENs have sequence and structural similarities (**Figure 1.7**)[6,37].

The similarities and differences between human FEN1 and human EXO1 are useful in understanding the function of the enzyme and the product release. Physically, EXO1 is much larger than FEN1 but has smaller nuclease domain. EXO1 was found in many organisms including humans and it was first purified from *S. pombe*. EXO1 is vital for genome maintenance during the mismatch repair (MM) and double-strand break repair (DSB) and carries out processive 5'-exonucleolyic reactions on DNA nicks, gaps and blunt

ends[7,17,61]. During DNA replication, there are many processes which can give errors such as insertion, transition, deletion and also double strand DNA breaks. Orans et. *al.*, published the structure of human EXO1 demonstrating similarities with the human FEN1 structures for binding and substrate placement over the active site. Furthermore, EXO1 binds at ss/ds DNA junctions and it was proposed that the ends of the DNA are frayed apart to allow for the 5' end to be placed over the active site metals[17].

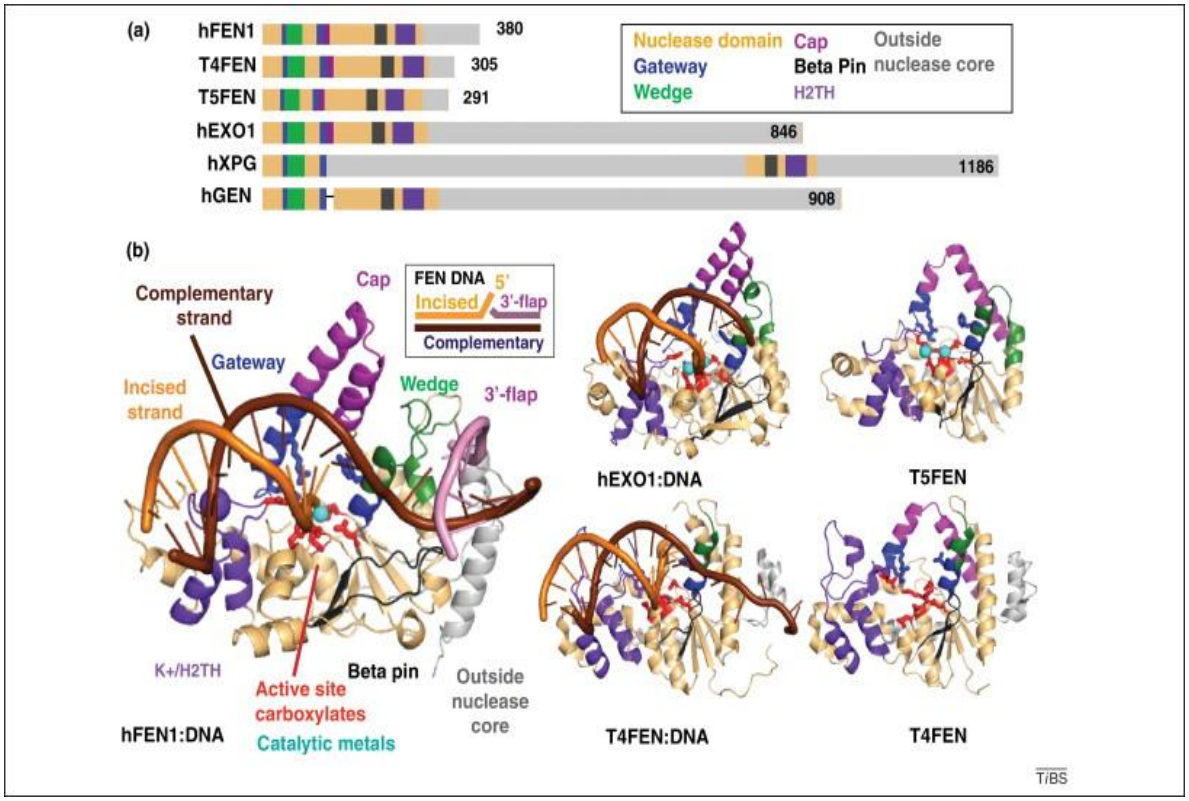

Figure 1.7: *FEN superfamily architecture. (a) Six of FEN superfamily showed each of them have a conserved architecture by sequence alignment to interact with DNA. (b) The juxtaposition of FEN superfamily structures*[6].

Another FEN1 superfamily member is GEN1 that was isolated from humans, yeast and *Drosophila melagonaster* (DmGEN)[63]. GEN1 is proposed to play a role in homologous recombination acting at four-way Holiday junctions (HJ) that form during replication by acting in a dimeric form to generate the bilateral cleavage. GEN1 also has a capability to endonucleolytically cleave 5'-flaps at lower efficiency. GEN1 is a monomer in free solution. Moreover, GEN1 has the ability to fit with the general mechanism of action of

other holiday junction resolvases, to produce symmetric nicks in the structure and to oligomerise in a substrate-induced fashion[7,61,64].

XPG is a part of a multi-protein complex require to cleave a bubble at the 3'-end subsequent to a 5' cut during nucleotide excision repair (NER)[65] and transcription-coupled DNA excision repair[66,67]. *In-vitro* studies revealed, XPG can cleave various DNA structures including single-strand 5'-overhangs. However, *in-vivo* the unique features of XPG allow it to recognize DNA bubble structures (**Figure 1.8**). Bubble structures are a cellular reaction to photo-damaged nucleotides and to bulky chemical mutagenic adducts[57]. The intermediate domain XPG plays an important role allowing reaction of the bubble structure[37]. Since the FEN1 arch is a narrow space and requires DNA with free 5'-termini to pass through this is adapted in XPG to have more than 600 amino acids while other FEN superfamily at this region only have approximately 70 amino acids[68]. This spacer region allows accommodation of the bigger substrate. The activity of a FEN1-XPG hybrid with a FEN1 arch has been tested and revealed that the catalytic efficiency is reduced for the bubble structure[69].

The conservation of the core elements of FEN structure between different types and between different members of the superfamily is an obvious indicator of the importance of the reaction for DNA maintenance (protection). Members of this superfamily act on a variety of DNA structures found in cells. Therefore, the basic structural components are conserved throughout evolution in a multitude of organisms, but the structure has been modified for specific tasks, proving the versatility of the core structure. FEN1 is studied as a paradigm of this superfamily as it contains all the relevant structural elements. A summary of the FEN1 superfamily functions and biological pathways were shown in **Table 1.1** and **Figure 1.8** shows the preferred substrate for FEN1 superfamily.

Table 1.1: *FEN1 superfamily, functions and their biological pathway.*

Structure	Functions	Biological Pathway
hFEN1	5’-endo/exo-nucleoytic cleavage of DNA double flap.	DNA replication and lp-BER.
EXO1	5’-exonucleoytic cleavage of DNA nicks, gaps, blunt ends.	Double strand break repair and mismatch.
GEN1	5’-endonucleoytic cleavage of DNA Holliday junction resolvase.	Double strand break repair.
XPG	5’-endonucleoytic cleavage of DNA bubble.	Nucleotide excision repair.

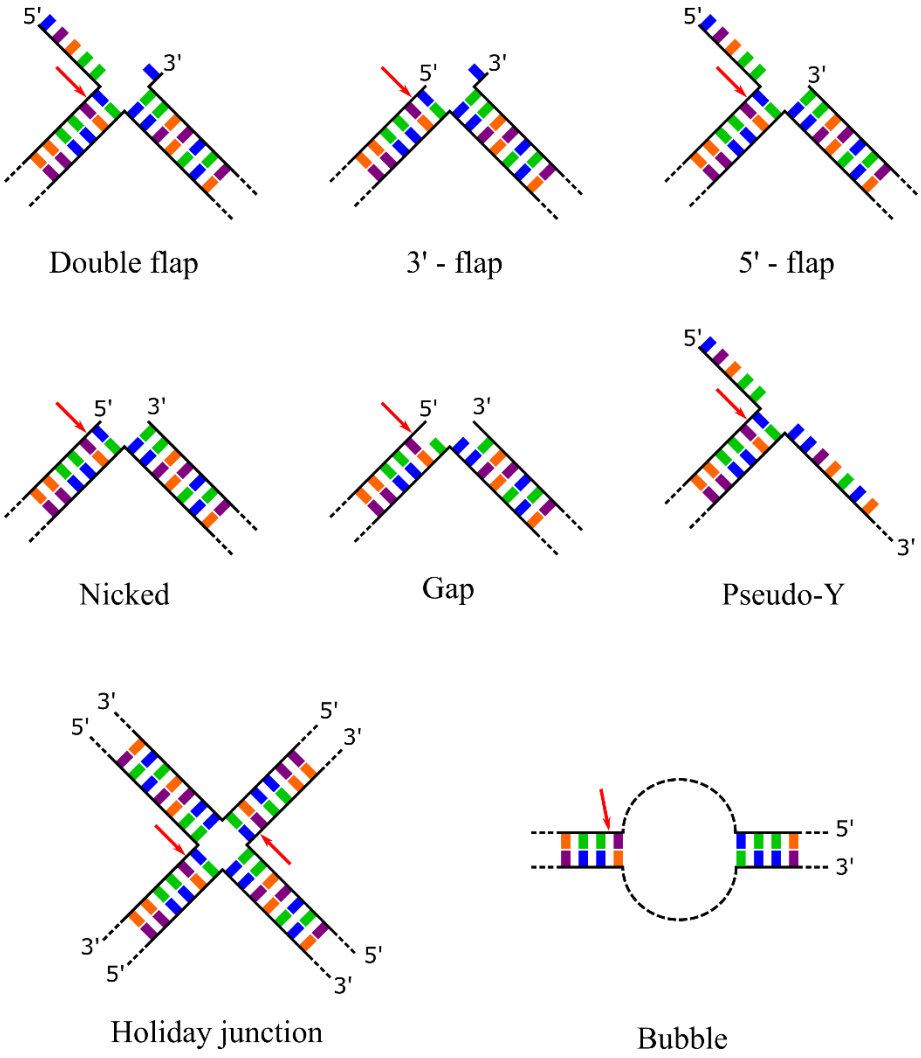

Figure 1.8: *DNA structures within normal cellular activity. The FEN1 superfamily have their specific preferred substrate which occur in different biological pathways. Red arrow indicates the position of FEN1 superfamily incise the DNA.*

1.6 Structural studies and the mechanism of FEN1superfamily

Generally, FEN1 superfamily members have a similar protein structure. Crystal structures reveal some interesting features in FEN1 and the superfamily including the flexible loop containing a helical motif[9,29,70] sometimes called the arch or cap (**Figure 1.9**). The arch interacts with the 5'-flap prior to reaction. At the base of helical arch, also referred to as the gateway, the FEN1 active site binds divalent metal ions that can execute endo-nucleolytic and exo-nucleolytic cleavage. A long adaptable arch and a seven or eight carboxylate active site with specific metal ion organization are the two unique features of the FEN1 superfamily. These characteristics contribute to the different roles of each enzyme during DNA metabolism on top of their substrate specificity. The N-terminal domain, intermediate (I) linker domain and C-terminal domain are the three domains that construct the structure of eukaryotic FEN1 proteins[9,71]. Other homologues of FEN1 can be defined as proteins that share high homology of N and C domains with FEN1[41] but sometimes differ in the I domain and the nature of C-terminal extensions that allow family members to play different roles in DNA metabolism. The full length hFEN1 structure comprises 380 (42 kDa) amino acids with 1-332 amino acid residues for nuclease core domain, and 333-380 residues for the C-terminal tail[6,37] (**Figure 1.9**).

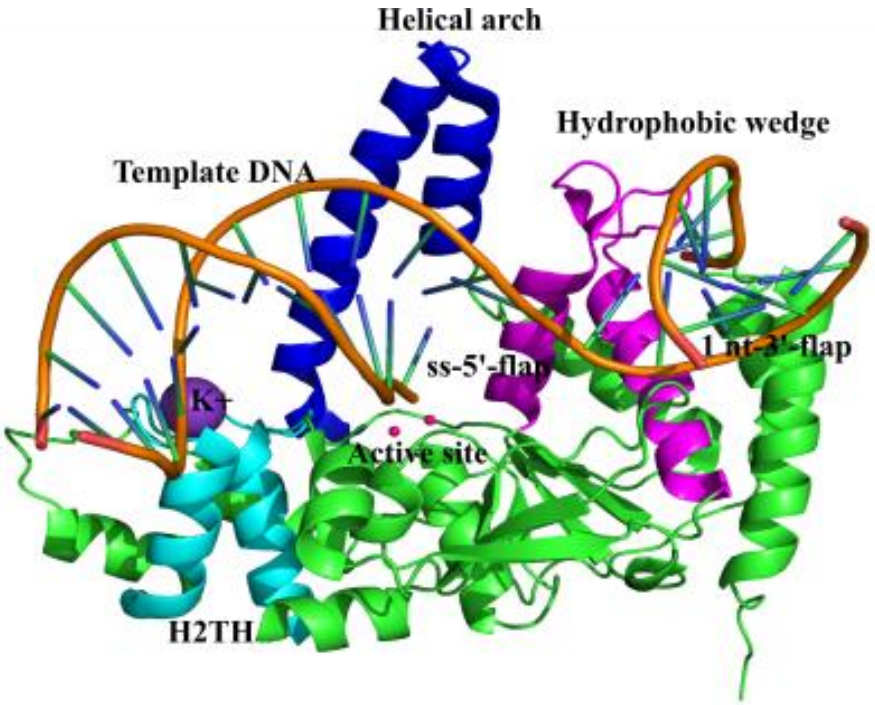

Figure 1.9: *The structures of human FEN1 with DNA substrate (PDB:3Q8K). The cartoon shows the ss-5'-flap, the 1nt-3'-flap and template from DNA. hFEN1 consists of 380 amino acid residues which comprise of 15-alpha and 7-beta sheet; helical arch (α4 and α5), hydrophobic wedge (α2 and α3), active site, H2TH.*

1.6.1 Helical arch

The helical arch is the intermediate domain of FEN1 structure and T5FEN was the first organism in which the helical arch structure was visualised[26]. The arch of T5FEN is made up by two alpha-helices; one has positively charged residues and another consists of hydrophobic residues[26]. When the arch is in ordered form, only single stranded DNA, but not double-stranded DNA can fit through. FEN1 crystal structures have revealed that diverse factors influence the degree of helicity of the arch including crystal contacts and substrate or product binding to the protein. Three ranges of structural rigidity have been found; either the helical arch is completely disordered, partially structured or completely structured[72]. Thus it is proposed that the helical arch has an inherent flexibility and ability to stretch. As an example, the conformation of the helical arch in substrate-free structures of the bacteriophage T5FEN and *Af*FEN are in helical ordered conformations, while in

T4RNaseH the arch is partially disordered and in *Mj*FEN and hFEN1 without substrate they were completely disordered[13,73,74].

The delineation of these regions is based on alignments of 5'-nucleases, with the gateway residues being well conserved in all superfamily members. In the crystal structure of FEN1, the helical cap consists of all α5 (Q121 – L130) and the C-terminal half of α4 (R105 – A116), whereas the gateway comprises by a part of α4 (P90-R104) and α2 (A35 – V46), and it is thought that the helical cap in hFEN1 enforces its specificity to discontinuous substrates. Alignments of FEN1s and EXO1s show that the helical cap is also present in hEXO1, which is consistent with hEXO1 also preferring discontinuous substrates. In GEN1, the helical cap is missing, whereas XPG has a large ~600 amino acid insert. The absence of helical cap in GEN1 is thought to allow it to cleave continuous DNA Holliday junctions by acting as a dimeric interface[7,61,63,64]. The large insert in XPG, which is predicted to be disordered, is important for protein-protein interactions with other nucleotide excision repair proteins and is a site for many post-translational modifications. The role of this large disordered region in XPG remains to be elucidated, but its interactions with other NER pathway members may be important for triggering scissile phosphate diester hydrolysis by XPG[57,66–69].

In substrate-free structures of human FEN1, the arch is disordered. However, upon the binding of substrate or product the arch adopts a helical conformation. hFEN1 has two helices with a short linker, whereas T5FEN only has α5 at the same region to run behind the main N-terminal domain of the protein. K93 and R100 in human FEN1, K83 and R86 in T5FEN are the two basic residues that are positionally conserved in all FEN1 proteins. Mutating the helical arch residues K93, R100 (hFEN1) and K83 (T5FEN) residues to alanine demonstrated these were critical to FEN1 activity[9,75].

1.6.2 The divalent metal ion active site

The active site is the cleft where catalysis happens and where the substrate is productively bound. The FEN1 superfamily are metalloenzymes. In FEN1 reaction, metal ions play a catalytic role (key role) in hydrolysis[76]. The studies on different mechanisms including human, archaea, and bacteriophage, revealed that two divalent metal ions bound to the active sites with varying spaces[6,26,28,37,77].

Nucleases with two metal ions (Mg^{2+}, Mn^{2+} or Zn^{2+}) are more common and Mg^{2+} is the most frequently associated with nucleic acid enzymes. For example in FEN1 divalent metal ions, are coordinated by 7-8 conserved carboxylates, located within the active site[9]. Three or four acidic amino acids are usually used to create a negatively charged pocket to coordinate the metal ions either through direct interaction (inner-sphere binding) or outer sphere binding through indirect interaction (via hydrogen bonding links) with metal bound waters. The inner and outer sphere binding can be as a bridge for activation of nucleophile and stabilisation of the DNA substrate. For example , in human FEN1 studies, the two metal ions are less than 4 Å apart allowing them both to contact the same scissile phosphate in the transition state[19].

In theory for the 'two metal ion mechanism' the inter-metal ion gap should be approximately 4 Å. However, diverse metal ion-metal ion distances have been published for FEN1 homologues of between 5-8 Å with most exceeding the 4 Å separation[78]. It was suggested that for a 'two metal ion mechanism' a conformational change in the protein would have to occur upon substrate binding to bring the two metal ions near each other to interact with the same scissile phosphate diester[76]. Small-Angle X-ray Scattering (SAXS)[79] and Fourier Transform Infra-Red (FTIR) spectroscopy[80] were used to detect the conformational changes. Furthermore, the presence of third ion binding during the formation of T5FEN-substrate complex was detected kinetically and ultimately crystallographically allowing a 'the two metal ion mechanism'[26].

To start the hydrolysis of the phosphodiester bond, the presence of Y40 (from α2), K93 and R100 (from α4), help to stabilise the substrate during and after the reactions and are important for DNA positioning. These two metal ions are held in place with four residues (Asp34, Asp86, Glu158, Glu160) to make the inner sphere and three residues (Asp179,

Asp181, Asp233) interact with H_2O to make outer sphere carboxylic groups that are vital to FEN1 catalysis[6,37] (**Figure 1.10**).

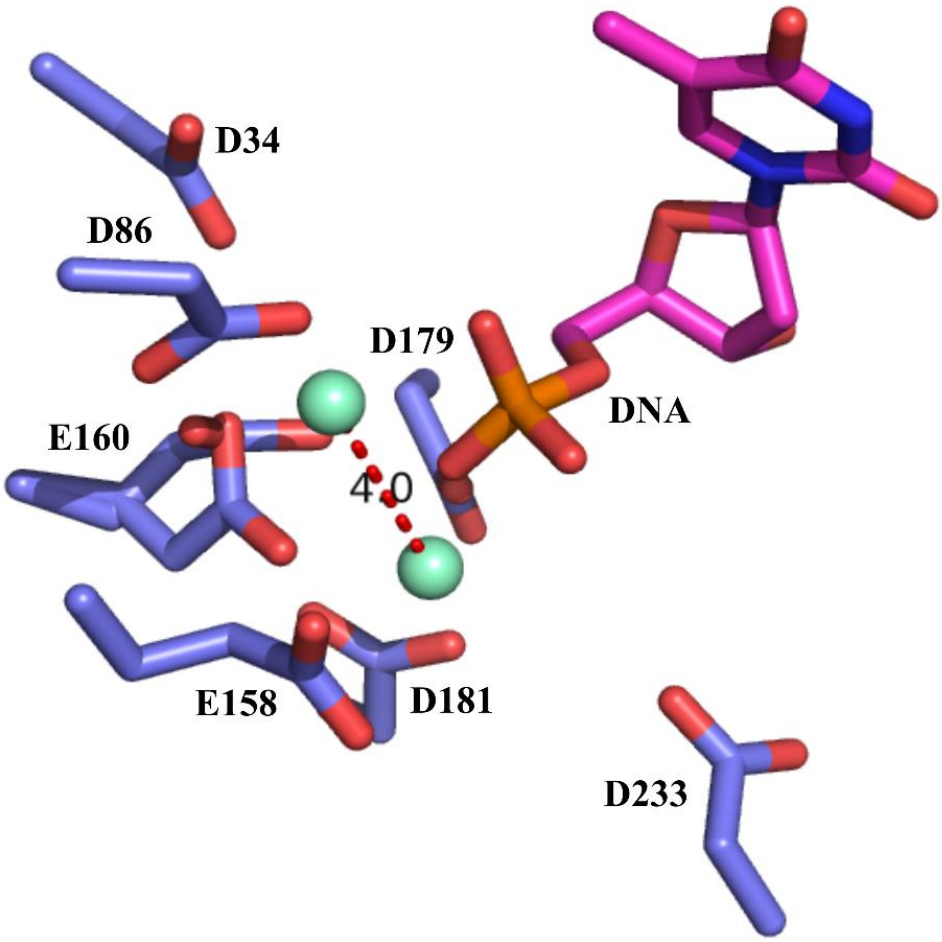

Figure 1.10: *The seven main residues (D179, D181, D233, D34, D86, E160 and E158) of hFEN1 to coordinate two metal ions.*

1.6.3 Helix-two-turn-helix

Another of the conserved structural elements in FEN1 and family proteins is the classical DNA binding site, helix-two turn helix (H2TH) motif, also known as helix-three turn helix (H3TH) motif[14] in archaeal enzymes. The H2TH or H3TH is the downstream duplex binding site, for non-specific binding of DNA[27,28,81,82]. Crystallographic images of a truncated hFEN1 and DNA revealed the presence of potassium ions (K^+) in crystal structures. The K^+ ion coordinates the DNA phosphate backbone, protein carboxyl

oxygens (from Leu238 and Leu241) and hydroxyl (from Ser237) group within the H2TH motif of α10 and α11. Four residues; Arg239, Lys244, Arg245 and Lys267 and the H2TH:K^+ ion provide the surface interaction with the duplex mostly with the DNA template strand[6,9,83]. Mutation studies on T5FEN (K125A and R216A) of the extended basic residues showed a deficiency in binding cleavage and studies on double mutations hFEN1 (K244AK245A and K252AK254A) impaired the cleavage significantly[81].

1.6.4 Hydrophobic wedge

The hydrophobic wedge was first discovered in *Af*FEN and consists of α2, the connecting loop and α3[22]. The position of this protruding binding feature is adjacent to the upstream binding site and below the helical arch. The hydrophobic wedge is responsible for forcing open the dsDNA-ssDNA or dsDNA-dsDNA bifurcated junction. In wild type human FEN1, when the upstream duplex bent at 100^0 angle, the top of helices α2-α3 and the connecting loop interact with the terminal base pair[22]. There is an extra interaction between some of the hydrophobic wedge residues and the last base pair face help to stabilise the bent configuration[9]. Hydrophobic wedges undergo a disorder to order conformational changes upon substrate binding[6,9].

1.6.5 3’-flap binding pocket

Studies reported that FEN1 cleaves double flap substrates which have unpaired single nucleotide 3’-flap endonucleolyticaly or exonucleolyticaly with greater efficiency than a single 5’-flap substrate[84]. A specific binding pocket for this single nucleotide flap is observed. Structural studies of FEN1 proteins from different organism showed, only the higher organisms possess a 3’-extrahelical nucleotide binding pocket[13] to attach with their preferred double flap substrates (a 5’ displaced single-stranded DNA flap and 1 nt at 3’-flap)[22]. This binding pocket stimulates catalysis of substrate hydrolysis including at lower concentration by stabilising the enzyme-substrate complex – it orientates the positions of the scissile phosphate near to the active site to facilitate the cleavage specificity. In bacteriophage enzymes, no stimulation occurs on introduction of a 3’-flap because of the absence of the binding pocket. Theoretically, in human FEN1 an acidic block of residues

will select a one nucleotide 3'-flap by creating a 'road-block'. The binding pocket is proposed to sterically block the 3'-flap from being longer than one nucleotide[9,73]. The single nucleotide of 3'-flap is directed by a specific contact to the 3'-hydroxyl group on the nucleic acid, and DNA backbone contacts in the minor groove together with a Leu residues that stacks between the unpaired nucleobase and the next nucleobase - help to anchor this construct in place. In human FEN1, 10 amino acid residues (M65, L53, T61, Q54, N55, Q315, K314, F316, R320 and S317) surround the 3' flap binding pocket to thread the one nucleotide into the pocket.[6,13,37].

1.7 FEN1 and substrate binding specificity

FEN1 is a so-called multifunctional and structure specific enzyme because of the capability to recognise different DNA structure with various "activities"[7]. The substrate specificities including with or without 5'-flap or 3'-flap on FENs superfamily have been discussed in detail. *In vitro* studies indicate that, FEN1 substrates possess multiple potential reaction sites either endonuclease or exonuclease[7]. Few aspects can decrease the effectiveness of FEN1 superfamily to cleave 5'-flap; there are; 1) more than 20 nucleotides length of 5'-flap - the presence of secondary structure, 2) the presence of non-nucleotide molecules attached to the flap and 3) the modification of the 5'-phosphate group of the reacting 5'-duplex. The mechanism for FEN1-substrate binding involves firstly the protein interacting with the ds-region downstream of the cleavage site at the ss-ds DNA junction. Another site interacts with the upstream ds-region (3'-flap binding pocket) and the last part is involved in the binding of the ss-flap region of the DNA substrates[6,9,13,37].

Studies of FEN1 from yeast *S.cerevisiae,* proposed that when the upstream primer consists of 1 nt 3'-flap, the reaction is more efficient compared with a fully annealed upstream primer[85]. A study of the 5'-nuclease of *Taq* DNA polymerase I[9], has shown that a substrate with 1 nt overlap between the 3'-end of the primer strand and the 5'-end of the downstream duplex was cleaved more efficiently compare with a substrate without[86].

Generally, all FENs cleave double flap substrates endo-nucleolytically. The reactions occur at the bifurcated site, but the main hydrolysis is a single nucleotide into the 5' downstream region. The one nucleotide 3'-flap substrate produces a specific reaction one nucleotide into the double-strand to give rise to a nicked DNA product. *In-vivo* 5'-flaps and 3'-flaps are mobile because both displaced DNAs are complementary to the template and can therefore potentially equilibrate into different lengths of double 5'-3' flaps[86–88].

Besides that, FEN1 superfamily can also cleave other flap substrate structures including a pseudo Y DNA structure (lower organisms) and 5' overhangs which do not have the upstream duplex. FENs also can cleave exo-nucleolytic substrates which does not have a single-stranded 5'-flap including, 5' blunt ended duplexes with a 3'-flap, nicked DNA and duplex DNA. For the FEN1s of higher organisms reactions of substrates that lack a single nucleotide 3'-flap or the ability to generate one are generally very slow[81].

A substrate that has a portion of secondary structure in the 5'-flap is has been termed a gap substrate. The gap endonuclease activity of FEN1 is also important. *In-vivo*, secondary structure such as hairpins in the 5' single stranded flap can appear when flapped DNA is self-complementary. Genetic disorders such as Huntington's diseases are related to the presence of tri-nucleotide repeat (TNR) sequences within the genome. These can form higher order secondary structures within the strands. FEN1 gap endonuclease activity suppresses TNR expansion. On double flap substrates that have a region of duplex within the 5'-flap, the reaction efficiency depends on the gap length[89].

To prevent any error during the processes, the right substrate must participate. Therefore, hFEN1 as a structure-specific nuclease, displays strand specificity and is totally independent of sequence for recognising the substrate. Studies revealed that hFEN1 is specific for double-stranded and does not work with single stranded DNA substrates.

1.8 FEN1 catalysed mechanism

The mechanism of FEN1 catalysis has been debated. Most of the studies proposed that the mechanism of the substrate positioning, the helical arch and divalent metal ions play crucial roles. Older studies suggested that FEN1 recognizes single-stranded flaps and could bind to 5' flaps regardless the length and nucleotide sequence[72]. An important recognition feature is the ability to bend the DNA junction (100^{o} bend), which allows the 3' flap bind at the binding pocket and the threading of 5'-flap. Additionally, FEN1 could hydrolyse DNA with and without a 5'-flap as long as the 5'-end is phosphorylated[13,37,72].

There has been much discussion about the role of the helical arch in the FEN1 reaction and how it interacts with the substrate. Three mechanisms have been suggested: tracking[90–92], threading[86] and clamping[17] (**Figure 1.11**). *Bambara et al.* postulated, that FEN1 captured the end of 5'-flap and tracked down in a threading fashion or with a ratcheted movement from a clamped enzyme until it reached the junction with dsDNA when reaction would occur[36]. However, this mechanism is not efficient for 3'-single flap (3'-SF) and gapped substrates that lack 5' flaps. Another hypothesis proposed that FEN1 binds with the double strand regions of its substrate first and then continues by accommodating the 5' portion of the substrate by threading it through the arch to form cleavage competent complex[9,71]. However, the problem of how single-strand DNA passed through the small hole of the arch remained a question.

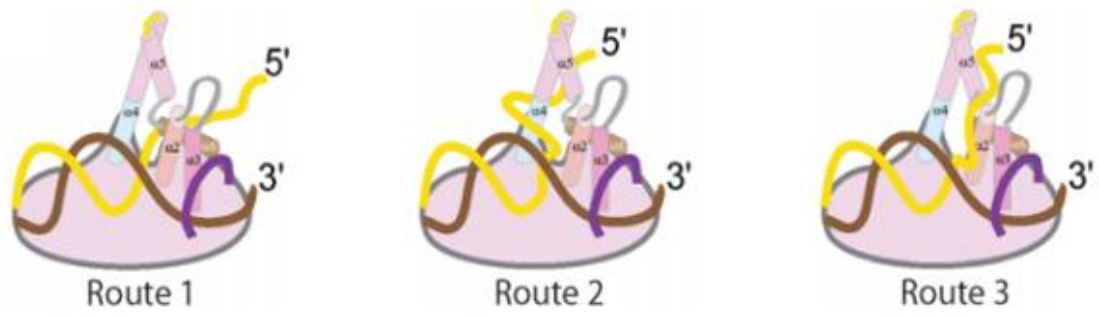

Figure 1.11*: The threading and clamping mechanism. Route 1 shows the 5' flap threaded through the helical arch (between α4 and α5), allowing cleavage when dsDNA is encountered. Route 2 and 3 show the clamping scheme; 5' flap is passed to one side of the helical arch before cleavage*[72].

Alternatively studies suggested the arch of FEN1 worked as a 'clamp' and not by threading the flap through. The clamping model proposed FEN1 recognises the dsDNA junction first, then the DNA passes to either side of the helical arch rather than threading through it. Next biochemical evidence in 2012 by *Patel et al.*, proposed a disorder to order mechanism[72] of the helical arch and this model solved some of problems with the previous threading theory. A disorder-thread-order model has been developed in which FEN1 initially recognises the dsDNA junction while the helical arch is disordered. Once the DNA substrate has bound and the 5' flap been threaded through the helical arch[74], the arch becomes ordered[72].

1.9 Human FEN1 in threading mechanism

To date, the most relevant mechanism for FEN1 accommodation of DNA substrates with a 5' ssDNA flap is a threading mechanism. The theory is the arch is in disordered position, but once ss-5'-flap thread through, it changes to an ordered form and known as 'disordered-thread-ordered' mechanism[72] (**Figure 1.12**). As we know, the ordered helical arch width allows only single-stranded DNA through, therefore this mechanism is the best technique to organise specific DNA structure. Various methods have been used to investigate the threading processes including the use of 5'-biotinylated flap substrates. Streptavidin is a 53 kDa tetrameric moiety that binds with very high affinity to biotin[72,90]. The size of conjugated streptavidin prevents the passage of a 5'-flap through even a completely disordered helical arch. The binding of streptavidin to a 5'-biotinylated substrate has a dramatic effect – on FEN1 activity[72,75,93].

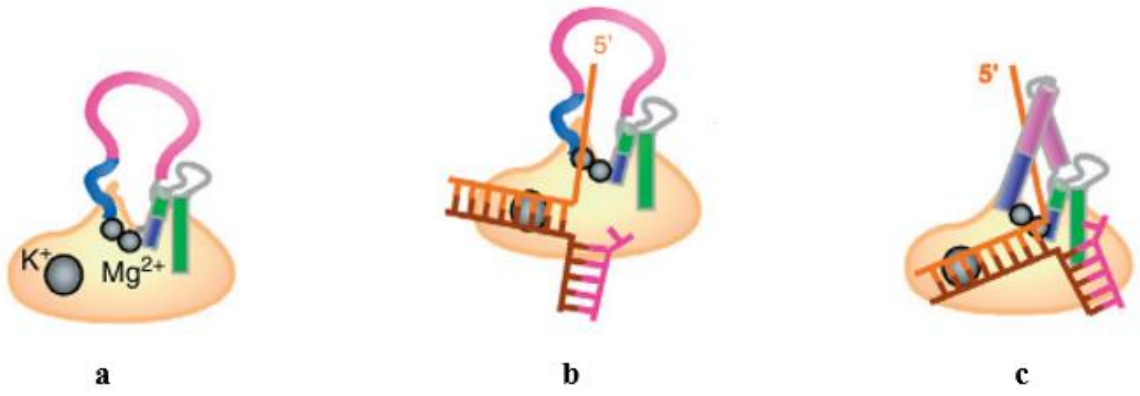

Figure 1.12: *A model for threading mechanism (model based on PDB: 1UL1). a) with free DNA, the arch are in disordered position, b) the single-strand 5'-flap threaded through the disordered helical arch and c) ordered helical arch with presence of DNA*[6].

The threading of the ss-5'-flap into the disordered helical arch, forces the scissile phosphate to move near the active site and undergo hydrolysis providing that the 3'-flap has been recognised[6]. The interaction with the active site can occur through direct contact with active site metals or contact with amino acid residues of the active site. Studies suggested the ordered position is stabilised by two residues of α2 (M37 and Y40), four residues of α4 (K93, R100, R103 and R104) and also three residues of α5 (K125, K128 and R129)[6,9,37].

The streptavidin-binding assay had two possible outcomes depending on the order of addition of components; if streptavidin is added first followed by hFEN1 later, the irreversible binding of streptavidin to the DNA takes place and '*blocked*' (**Figure 1.13**) the DNA from threading through the helical arch. When Mg^{2+} is added to initiate reaction the substrate only reacts very slowly. On the other hand, if the streptavidin was added after incubation of the 5'-biotinylated substrate with hFEN1, the substrates becomes *'trapped'* (**Figure 1.13**) within the helical arch and unable to unthread. In this case when Mg^{2+} is added to initiate reaction the substrate is hydrolysed rapidly. It was found that double flap substrates contain a region of duplex in the flap could be trapped on the protein. These substrates could not thread through a structured arch. It was therefore proposed, that the dsDNA binds to the protein when the helical arch is partially or completely disordered. The 5' flap gets threaded through and the helical arch orders around it[72,75,93].

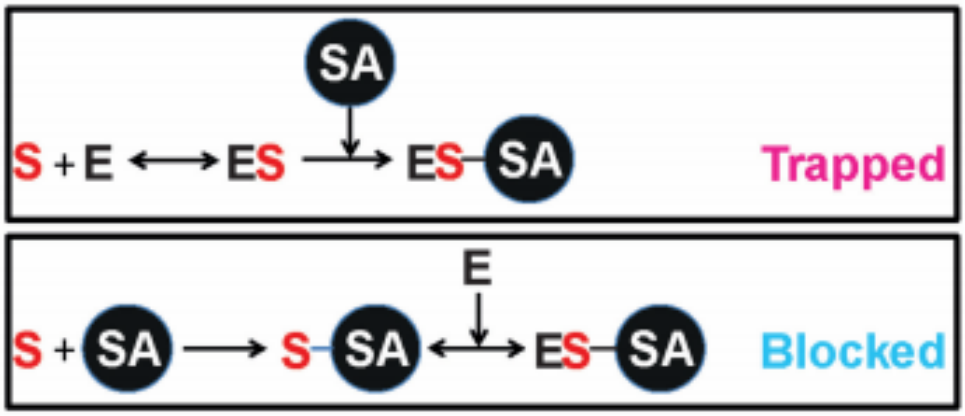

Figure 1.13: *The two schemes for the threading analysis performed by Patel et al., either trapped or blocked.* ***S*** *is a 5'-biotinyated substrate,* ***E*** *is an enzyme, SA is a streptavidin*[72].

1.10 FEN1 protein partners

During the replication and repair processes, various proteins work together. One of these interactions is between PCNA a sliding clamp protein and human FEN1. *In vitro* studies suggest that the activity of hFEN1 can be stimulated by PCNA[46]. DNA polymerases and DNA ligase also interact with PCNA[11]. In DNA replication, PCNA is proposed to coordinate DNA polymerase, FEN1 and DNA ligase to maximise the throughput of DNA and maintain fidelity[11]. The C-terminus of FEN1 has been shown to be disordered until it encounters other proteins such as PCNA. **Figure 1.14** shows the cartoon of hFEN1-PCNA interaction. hFEN1-PCNA binding orders the C-terminus into a β-strand that packs against a PCNA homotrimer and allows for regulatory interactions[37,94].

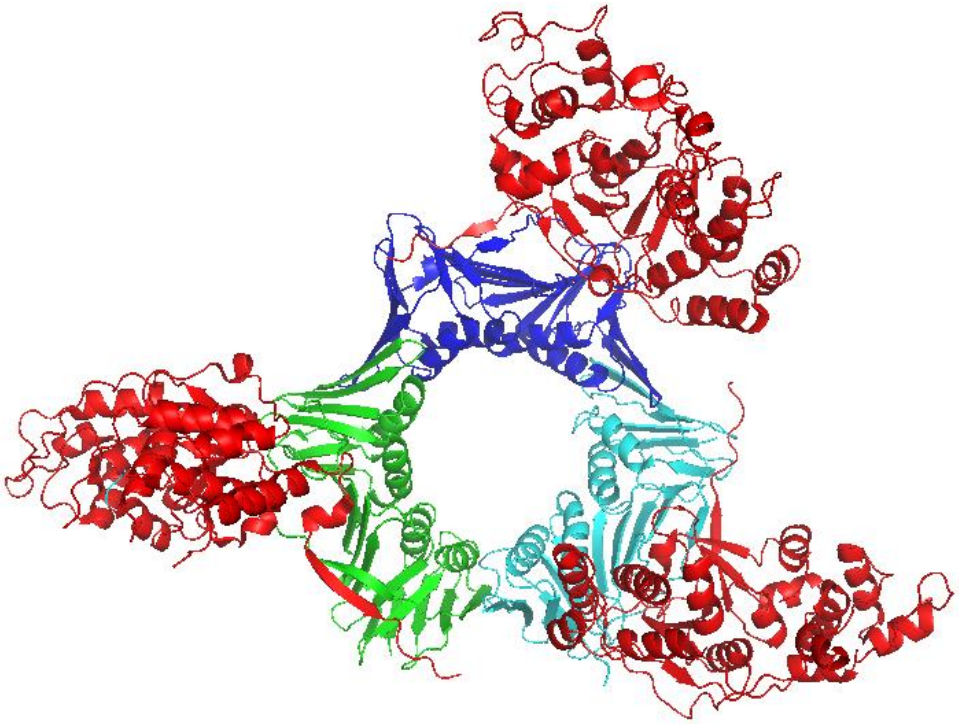

Figure 1.14: *The cartoon representation of the hFEN1-PCNA interaction visualised by PyMOL Viewer (PDB ID: IUL1), where the PCNA homotimer consists of three domains (blue, green and turquoise) each bound to hFEN1 (red).*

1.11 Proliferating cell nuclear antigen

A similar structure of DNA sliding clamp proteins have been found in archaea, bacteria and eukaryotes. In eukaryotes, the sliding clamp protein is called the proliferating cell nuclear antigen (PCNA), also known as cyclin or the DNA polymerase δ auxiliary protein[95]. Based on discovering PCNA in the nucleus of animal, plant, and yeast cells during cell separation it was proposed that PCNA plays a key role during the synthesis phase (S) of the cell cycle and it is highly conserved protein[96]. PCNA is a dynamic protein that was originally distinguished as the processivity factor for DNA polymerase delta and as necessary protein for eukaryotic chromosomal DNA replication[94].

In-vitro studies with *Simian virus 40* (SV40) DNA replication aided understanding of the function of the proteins at the replication fork including the PCNA and pol-δ[97]. A later investigation reported numerous X-ray crystal structures of PCNA bound to peptides derived from PCNA-binding proteins[98]. More recent studies have discovered that the

sliding clamp of PCNA allows it to work and cooperate with many different proteins involved in several different metabolic pathways. PCNA is a multifunctional protein with responsibility in various processes including DNA replication and DNA damage repair[94].

1.12 Structures of PCNA

PCNA is a sliding clamp in eukaryotes, which is responsible for providing replicative polymerases with the high processivity needed to duplicate the cells. PCNA has a ring structure that binds the DNA topologically and can slide freely along the DNA. Crystallographic studies demonstrated that PCNA has a ring shaped structure in solution and the central hole is adequate to accommodate the double helix of DNA. The structure of the PCNA trimer consists of three monomers, each of them containing a domain A (N-terminal) and a domain B (C-terminal). These two domains are held together by an extended β-sheet across the inter-domain boundary on each subunit[99]. The inter-domain connector loop (IDCL), residues 119 to 134, is a flexible linker and a region of PCNA where a large number of proteins bind. The six-fold symmetry and six repeating domains generate a hexagonal shape and are bound in a head-to-tail[96] order to form ring-shape homo-trimer[37,94,100] (**Figure 1.15**).

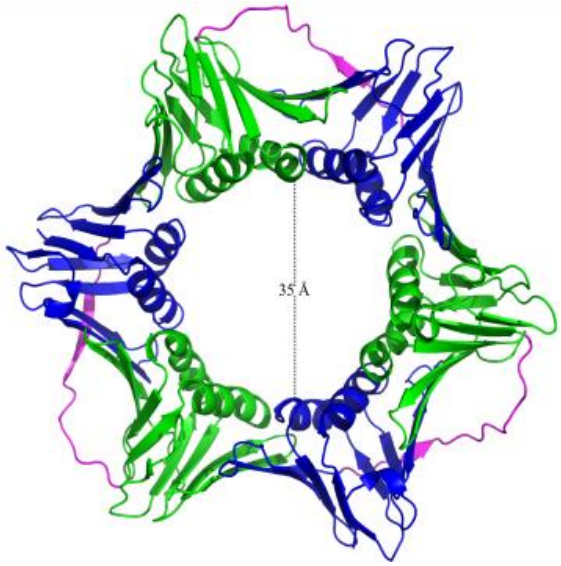

Figure1.15: *The cartoon representation of the structure of PCNA visualised by PyMOL Viewer (PDB ID: IUL1); where the purple colour represent IDCL (119 - 134 residues), the blue colour show N- terminal (1 - 118 residues) and green colour show C-terminal (135 – 258).*

The approximate diameter of PCNA ring was found to be 80 Å and the central hole was approximately 35 Å large enough for dsDNA to pass through. PCNA has non-equal surfaces between the inner side and outer side. Generally, the electrostatic potential for PCNA is negative due to the β-sheets at the outside surface of PCNA. The inside surface of PCNA ring is strongly positively charged, due to the presence of lysine and arginine[99]. Besides that, the layer of α-helices at the inner side allow PCNA to encircle the double helix DNA by interaction with the phosphate backbone of DNA. The unique topological nature of PCNA surfaces, allow it to slide along a duplex DNA stably and freely in either direction. Most of the protein-protein interactions occur at the front face of PCNA ring because it presents the IDCL. The role of the back face of PCNA is not fully investigated to date. Previous studies have revealed that PCNA can interact with a diverse set of proteins to play many cellular processes including binding with PIPs (PCNA interacting peptides) and full-length proteins[23,94].

1.13 The FEN1 and PCNA interaction

The human FEN1-PCNA complex visualised by X-ray crystallography is composed of three FEN1 molecules, bound to one PCNA trimer in the absence of DNA[101]. The three PCNA subunits tightly associate to form a closed ring, with each subunit exhibiting symmetry[23]. The main interface of FEN1-PCNA complex involves the C-terminal tail of FEN1, which forms two beta-strands connected by a short helix, the βA-αA-βB motif, participating in beta-beta and hydrophobic interactions with PCNA[23] and keeping the enzyme in an inactive locked-down orientation. These interactions are furnished by the residues forming the consensus PIP-box[102] or binding motif "Qxxhxxaa", where 'h' is an aliphatic hydrophobe (leucine, isoleucine or methionine) and 'a' an aromatic hydrophobe (phenylalanine or tyrosine), which serves to anchor the client protein to the clamp[23,94,102].

Three-dimensional crystal structures of the hFEN1-PCNA complex have revealed that PCNA also interacts with other regions of FEN1[103], including the short linker of FEN1 hinge region; ^{333}QGST^{336} connecting with FEN1 via the strand βA of the C-terminal tail and α13 of the core domain FEN1. It has been suggested that this may be responsible for the stimulation of FEN1 nuclease activity. Further backed up by mutations in the hinge region affecting the nuclease activity. Biochemical studies suggested that the FEN1-PCNA

complex has three possible functions. These are (1) recruitment of FEN1 to the site of DNA replication, (2) stimulation of the cleavage of RNA primer flaps and (3) coordination of highly ordered RNA primer processing and DNA ligation[103].

1.14 The role of PCNA

PCNA is required for replicative polymerases to duplicate an entire genome. Various studies have revealed that PCNA plays a key role in cell systems, especially eukaryotic cells. Proteins binding to PCNA have specific function in cells. The proteins interact with PCNA via PCNA-interacting peptides (PIPs). PIPs that have function in enzymatic activity including DNA replication, DNA repair and translesion synthesis[94].

1.14.1 PCNA in DNA replication

DNA replication involves many processes such as initiation, elongation, replication fork progression, DNA replication proteins, replication machinery and termination. In eukaryotic cells, DNA replication occurs at the Synthesis phase (S) of the cell cycle. PCNA has no enzymatic activity, but PCNA is a central and essential factor for DNA replication acting as a moving platform to strengthen the process. According to these particular features, the presence of the Replication Factor C (RFC) is important to localise the PCNA trimer to the 3'-OH of a DNA primer by binding to the C-terminus of PCNA[94,100].

At the leading strand in DNA replication, the PCNA trimer will encircle DNA, secure polymerases firmly, then displaces pol-α/primase and allows the recruitment of pol-δ or pol-Ɛ to replicate the leading strand continuously in an error free manner. The four subunits of DNA pol-δ are p125, (contains the catalytic activity), p66 (contacts with PCNA), p16 and p50[94,104]. However, on the lagging strand, FEN1 and DNA ligase-1 interact with PCNA at the IDCL to seal the nicks between Okazaki fragments. Studies proposed, during the Okazaki fragment maturation, pol-δ and FEN1 readily can associate with PCNA and dissociate from PCNA then replace one another on a single PCNA monomer. The switching processes in the correct sequence between pol-δ and FEN1 on a single PCNA monomer will need to be adequate for effective replication[94].

The PCNA trimer ring like structure allows itself to encircle around DNA and slide along the DNA double-helical structure freely in both directions[99]. PCNA plays important roles in coordinating the process of switching the pol-α and pol-δ (or pol-Ɛ) at the discontinuous lagging strand and the synthesis initiation of leading strand and in the completion of the Okazaki fragment maturation process[94].

1.14.2 PCNA in DNA repair

DNA damage can occur to the DNA molecules in cells at a rate of approximately one million individual molecular lesions per cell per day during metabolic activities and from environmental factors. The lesions will produce a big impact to the code of the genome such as structural damage and harmful mutations. Two types of DNA damages have been identified either endogenous damage or exogenous damage, which occur from an external agent (e.g. ultraviolet radiation from the sun). DNA repair is a process where many proteins work concurrently often with PCNA involvement. Some of the DNAs can be repaired with metabolic pathways such as mismatch repair (MMR), base excision repair (BER), and nucleotide excision repair (NER) which all involve a DNA synthesis step requiring polymerase, which further indicates a function for PCNA.

Previous studies have shown that in DNA, the PCNA can recruit and coordinate repair proteins in MMR process by interaction with mispair-binding proteins forming a ternary complex and the complex will be transferred from PCNA to ATP[105–108]. While, under lp-BER *in-vitro*, PCNA is involved in the re-synthesis and incision steps for repairing specific sites during the DNA synthesis[109–115]. In addition, the NER pathway is the process where mammalian cells eliminate carcinogenic lesions. Normally the DNA damage occurs from sunlight or other common mutagens. Two stages are involved in NER pathway; a) incision at sites of DNA damage and b) synthesis of new DNA to restore the damaged nucleotides[116]. PCNA plays a role at the early step of the NER pathway. PCNA will recruit the endonuclease XP-A and XP-G proteins to the specific site of lesion and binding with pol-δ to re-synthesise the complementary strand and finally Lig1 will ligate the fragments[109,116–118].

1.14.3 PCNA in translesion synthesis

Another pathway for repair of damaged DNA is known as translesion synthesis (TLS). TLS authorize the DNA substrate replication machinery to duplicate over DNA lesions by changing out the normal DNA polymerase[119]. TLS polymerases have the unique potential to duplicate using damaged DNA as a template. In most cases, the TLS can incorporate the correct nucleotide. PCNA is in charge of coordinating the interchange between replicative pol-δ (or pol-Ɛ) and special polymerases in TLS and recruits them at the replication fork.[120].

1.15 Aims of project

Flap endonucleases (FENs) catalyse the essential removal of single-stranded 5'-DNA or RNA protrusions known as flaps that occur during DNA replication and repair. When flaps occur in vivo they are migrating structures that can adopt a number of conformations, but it is believed that FEN1 will only act on one conformer. In addition, when FEN1 acts *in-vivo*, it is typically in complex with the protein PCNA (Proliferating Cell Nuclear Antigen). This project will investigate the effects on FEN1 reaction specificity and rate of reaction of flapped structures of increasing complexity. The project will also investigate how PCNA alters the reactions of static or migrating double or single flaps and how the interaction between FEN1 and PCNA is mediated. The project will use the techniques of molecular biology, enzyme kinetics and biophysical chemistry. Various methods have been used to investigate the mechanism of how the hFEN1 select and accommodate the DNA substrate with a single-stranded 5'-flap. However, the accurate and precise mechanism is still unclear. This project aims to further investigate the role of hFEN1 during Okazaki fragments using threading analysis on the helical arch. Site directed mutagenesis of the human FEN1 are presented in order to understand the importance of conserved residues on helical arch. Finally, investigation on how the conformational changes of substrate on mutated human FEN1 compared with wilt type human FEN1. Exciton-Coupled Circular Dichroism (ECCD) will be used to monitor any distortion of DNA at reaction site by incorporation of 2-aminopurine (2A) at specific potions.

Chapter 2: Materials and Methods

2.1 Expression and purification of human FEN1 and mutant K125A

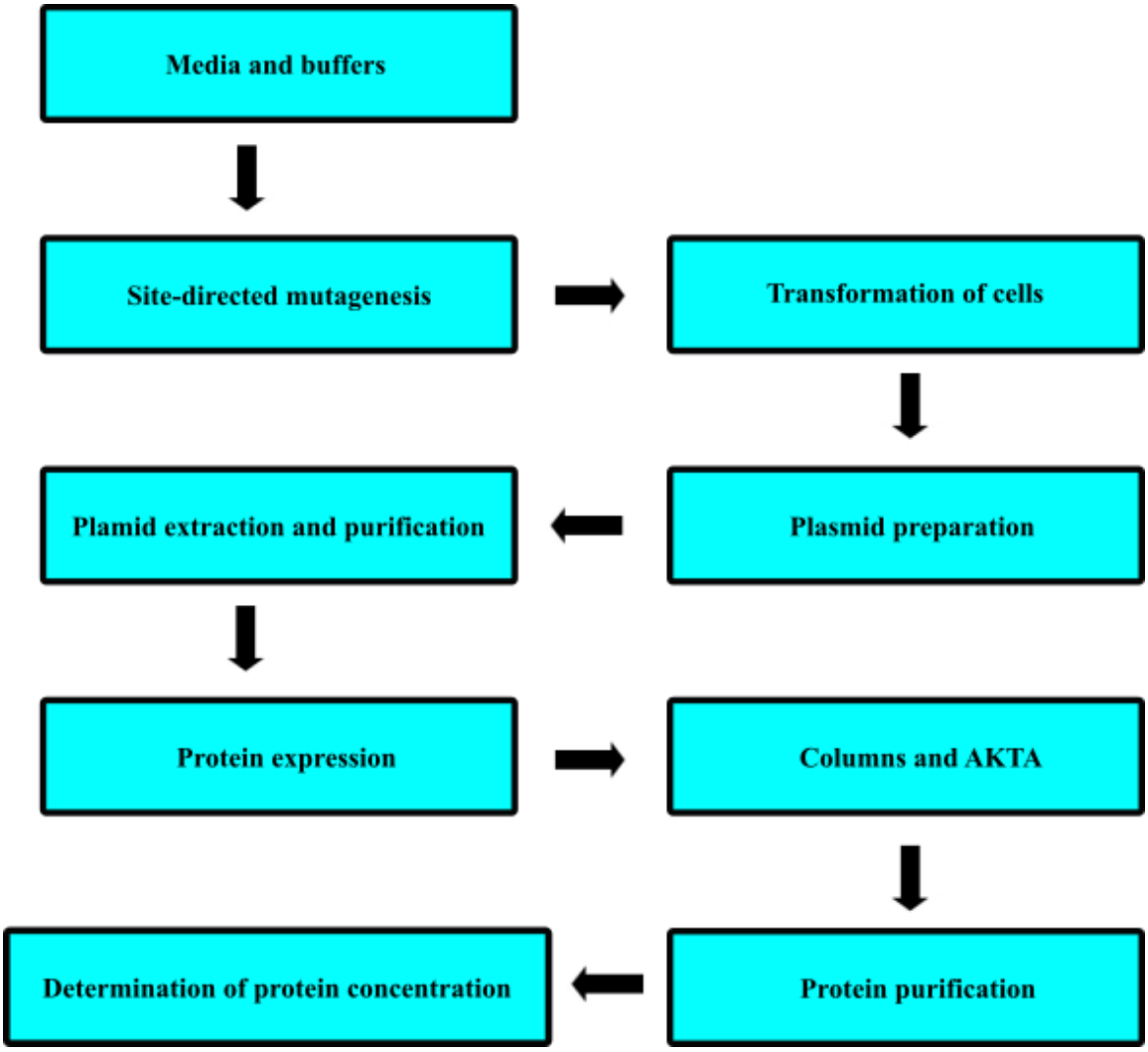

Figure 2.1: *The diagram of standard operating procedures for human FEN1 and K125 mutant expression and purification.*

Equipment

Instrument	AktaPURE (GE Lifesciences)
Column	OLIGOSep® Prep HC Cartridge
Temperature	4 °C
Flow rate	5 mL/min

Media and buffers

LB agar (1 L)	10 g tryptone, 10 g NaCl, 5 g yeast extract, 10 g agarose, H_2O, (antibiotics; 25 mg/mL kanamycin + 34 mg/mL chloramphenicol)
SOC media (1 L)	20 g tryptone, 0.6 g NaCl, 5 g yeast extract, 0.19 g KCl, 1.2 g $MgSO_4$, 3.6 g D-Glucose, H_2O
2XYT (1 L)	16 g tryptone, 5 g NaCl, 10 g yeast extract, H_2O (pH to 7.5)
TY media (1 L)	12 g Tryptone, 24 g Yeast Extract, H_2O
20XP (1 L)	1 M Na_2HPO_4 (142 g), 1 M KH_2PO_4 (136 g), 0.5 M $(NH_4)_2SO_4$ (66 g), 900 mL H_2O
50X 5052 (100 mL)	2.5 g glucose, 25 g glycerol, 10 g α-lactose, 73 g H_2O
All media	25 mL 20XP, 10 mL 50X 5052, 1 mL 1M MgSO4, 100 µL trace metals, 500 µL (34 µg/mL) chloramphenicol, 4 mL 50 µg/mL kanamycin
PBS	10 mM Na_2HPO_4, 2 mM KH_2PO_4 pH 7.4, 137 mM NaCl, 2.7 mM KCl
Cell lysis buffer	SIGMAFAST Protease Inhibitor Cocktail was dissolved in 100 mL IMAC A buffer
50X TAE buffer (1 L)	100 mL 500 mM EDTA pH 8.0, 242 g tris base, 57.1 mL glacial acetic acid
Resolving gel buffer	1.5 M tris pH 8.8, 0.4% SDS (w/v)
Stacking gel buffer	0.5 M tris pH 6.8, 0.4% SDS (w/v)
Agarose gel (100 mL)	100 mL 1X TAE, 10 µL SYBR safe DNA gel stain (10,000X) 1 g agarose
Imac A1	20 mM tris, pH 7.0, 1 M NaCl, 5 mM imidazole, 0.02% NaN_3
Imac A2	20 mM tris, pH 7.0, 0.5 M NaCl, 40 mM imidazole, 0.02% NaN_3, 0.1% Tween-20
Imac B1	0.5 M NaCl, 250 mM imidazole, pH 7.2, 0.02% NaN_3
Anion exchange A1	20 mM tris, pH 8.0, 1 mM EDTA, 0.02% NaN_3, 20 mM βME
Anion exchange B1	20 mM tris, pH 8.0, 1 mM EDTA, 1 M NaCl, 0.02% NaN_3, 20 mM βME
Heparin A1	25 mM tris pH 7.5, 1 mM $CaCl_2$, 0.02% NaN_3, 20 mM βME
Heparin B1	25 mM tris pH 7.5, 1 mM CaCl2, 1 M NaCl, 0.02% NaN_3, 20 mM βME

Storage buffer	100 mM HEPES (pH 7.5), 200 mM KCL, 2 mM $CaCl_2$, 0.04% NaN_3, 20% glycerol, 10 mM DTT

2.1.1 Site-directed mutagenesis

Primers for site directed mutagenesis were designed using http://www.genomics.agilent.com/primerDesignProgram.jsp and standard site directed mutagenesis protocols were used to create single point mutations. Briefly, the wild type plasmid (pET-28b) was sequenced by GATC Biotech. Once the sequence had been verified, a PCR reaction was set up. The mismatched primers (primer pair) were re-suspended with water to make the concentration to 100 µM. PCR samples were prepared on ice by PFU ultra reaction buffer with 10 mM dNTPs, template DNA (pET28b hFEN1-WT vector), 0.3 µM of each primer and PFU ultra HF Hot start polymerase. Samples were placed in PCR machine and run with following programme as shown in **Figure 2.2**, approximately 6 hours in total.

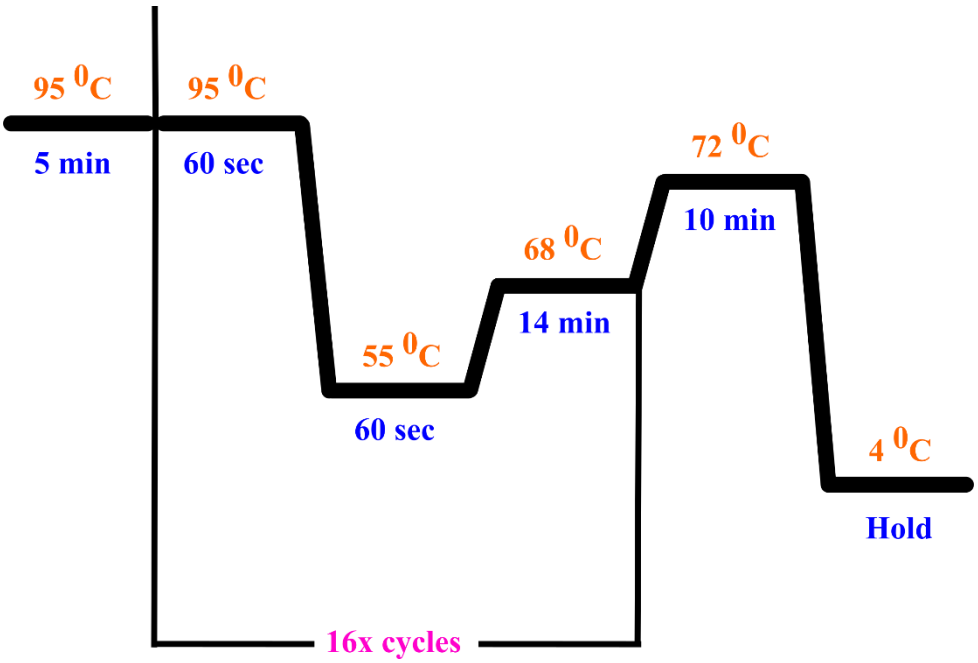

Figure 2.2: *Thermocycling protocol for PCR mutagenesis reaction. First step was initial denaturation for 5 minute at 95 ^{0}C. Then 16x cycles for DNA polymerization at 95 ^{0}C (60 sec), 55 ^{0}C (60 sec) and 68 ^{0}C (14 minutes). Continue the polymerization with extra 10 minutes at 72 ^{0}C for final extension and lastly, hold the DNA at 4 ^{0}C.*

On completion of the PCR programme, 1 µL of DPN1 was added, and the mixture was incubated for 1-2 hours at 37 ^{0}C. Agarose gel electrophoresis was used to analyse the products of PCR. Agarose (0.4 g) was dissolved in 50 mL 1x TA in the microwave for 50-60 seconds and then cooled to 50 ^{0}C in a water bath for 10 minutes with addition of 5 µL SyBr safe DNA gel stain. After loading the PCR mixture (10 µL), 100 bp marker and 1 kb marker the agarose gel was run in 1X TAE at 120 volt for one hour and imaged using the ChemiDoc MP system (BioRad).

2.1.2 Transformation of cells

The plasmid was transformed into *E. Coli* XL10 and incubated for 45 minutes in ice and then heat shocked at 42 ^{0}C for 45-60 and subsequently recovered by cooling the cells on ice for 5 minutes. 1000 µL of SOC media was added and incubated at 37 ^{0}C for one hour at 180 rpm. The cells were spun for 60 seconds and the pellet was re-suspended the with 250 µL SOC media. Finally, the cells were plated on the LB agar plates with 50 mg/mL kanamycin and incubated overnight at 37 ^{0}C.

2.1.3 Plasmid preparation

When the colonies had grown on the LB agar, one colony was added to 5 mL LB media containing 50 µg/mL kanamycin and incubated at 37 ^{0}C at 180 rpm overnight. The next day OD at 600 nm was used to confirm the starter cultures had grown.

2.1.4 Plasmid extraction and purification

A Monarch plasmid mini prep kit was used for the plasmid extraction and purification. The starter culture was divided into 1.5 mL Eppendorf tubes and pelleted at 13.3 rpm for 10 minutes at 4 ^{0}C. The pellet was re-suspended in 200 µL of buffer B1 (pink) and mixed by vortexing. 200 µL of buffer B2 (green) was added and gently inverted and incubated at room temperature for 1 minute. 400 µL of buffer B3 (yellow) was added and mixed slowly by inverting until the colour was uniformly yellow and a precipitate formed. After

incubation at room temperature for 2 minutes the tube was and spun down at 13000 rpm for 5 minutes. The supernatant was transferred to a spin column and centrifuged for 1 minute and the flow-through was discarded. The spin column was washed by adding 200 µL of wash buffer and spun discarding the flow through. This step was repeated by washing the spin column with a further 2X 400 µL of wash buffer and centrifuged for one minute to remove all remaining wash buffers. The spin column was placed in a 1.5 mL clean tube and 40 µL of elution buffer was added. After 1 minute the plasmid was collected. The concentration of the plasmid was measured by absorbance at 260 nm using a Nanodrop UV spectrophotometer. The plasmid were prepared for sequencing by adding 5 µM of T7F and 5 µM of T7R to 5 µL of purified plasmid. Sequencing confirmed the presence of the mutated genes.

2.1.5 Protein expression

The chosen expression vector was transformed into *Escherichia coli* BL21(DE3)-RILP cells using the same procedure described above. When the colonies had grown on the LB agar, a single colony was added to 25 mL 2XYT media supplemented with 15 µL of 50 mg/mL kanamycin and 25 µL of 34 mg/mL chloramphenicol. Shaking overnight at 37 ^{0}C, 180 rpm the culture was grown until an OD reading at 600 nm around 0.6 was achieved. 50 mL of the starter culture was used to inoculate a flask with 500 mL 2XYT media containing 500 µL 34 mg/mL chloramphenicol, 4 mL 50 mg/mL kanamycin, 10 mL 50X 5052 media, 100 µL trace metals, 1 mL 1 M $MgSO_4$ and 25 mL 20XP buffer. The cultures were grown for 4 hours at 37 ^{0}C, 225 rpm, then shaken (225 rpm) at 18 ^{0}C overnight. The cells were harvested by centrifugation at 6000 x g for 15 minutes at 4 ^{0}C. The supernatant was discarded and re-suspended in ice-cold phosphate buffered saline. The cells were transferred to 50 mL tubes and pelleted again by centrifugation at 4700 x g for 1 hour at 4 ^{0}C. The supernatant was removed and the cells were suspended with 30 mL of lysis buffer and 5 mL of lysis buffer containing 2% of lysozyme. The cells were incubated and mixed gently in the cold room for 1 hour. A small portion of the cell lysate was analysed by SDS PAGE analysis to confirm the presence of the protein. The lysate was stored at -20 ^{0}C until purification.

2.1.6 AKTA and columns

AKTA

To ensure that are no carry over from previous analysis, the AKTA system was washed with 0.1 M NaOH, H_2O, 20% ethanol and H_2O. The same procedure (wash or rinse the AKTA) was performed after each chromatographic step in the PCNA purification procedure. All super loops were also cleaned with sodium hydroxide and ethanol.

Columns

To prevent any contamination from previous protein purifications, all columns were washed manually via syringe according to manufacturer's protocols before beginning the purification. **Table 2.1** showed the type of columns and the buffers that have been used to wash the column.

Table 2.1: *Columns and buffers to clean up.*

Column	**Buffers**
HiTrap™ 5 mL TALON crude	$Co(NO_3)_2$, stripping buffer, IMAC B1, H_2O
HiTrap™ QFF 5 mL	0.2 M NaOH, H_2O, 20% EtOH,
HiTrap™ Heparin FF 16/10 (20mL)	0.2 M NaOH, H_2O, 20% EtOH,
HiPrep™ 26/10 Desalting	0.2 M NaOH, H_2O, 20% EtOH,
HiTrap™ SP HP	0.2 M NaOH, H_2O, 20% EtOH,
Hydroxyapatite	0.2 M NaOH, H_2O, 20% EtOH
Phenyl sepharose	0.2 M NaOH, H_2O, 20% EtOH,

2.1.7 Protein purification

The cell lysate was defrosted in cold water around for ~1 hour until completely thawed. The suspension was sonicated on ice at 75 % intensity for 10 minutes of pulsation for 5 seconds with 25 second cooling intervals. 5 mL of buffer IMAC A1 containing 1% Tween-20 was added to each lysate. The insoluble cell debris was removed by ultracentrifugation (30 000 × g at 4 ^{0}C for 30 min) leaving the clear lysate for purification. Protein purification was conducted on a AKTA pure chromatography system (GE Healthcare Life Sciences) in the cold room. All buffers were filtered and degassed prior to use for purification.

First, the supernatant was applied using a 50 mL superloop to the three tandem *HiTrap™ 5 mL TALON crude* column. The column was washed with 5 column volumes of IMAC buffer A1 and 5 column volumes of IMAC buffer A2. The bound protein was eluted as single fraction with 8 column volumes of IMAC buffer B1. Fractions containing the protein of interest were identified by SDS-PAGE. The eluted fraction was diluted 1:1 ratio with anion exchange A1 buffer.

Second, using the *hFEN1_HiTrap_5mL* method, the diluted eluate was applied to *HiTrap™ QFF 5 mL* column pre-equilibrated with the anion exchange A1 buffer. The column washed with anion exchange buffer A1, and then a linear NaCl gradient of 0 to 1 M NaCl was applied over 10 column volumes. The fractions containing target protein (the protein is found in the flow through) were pooled and 10 µL of the fraction were retained for SDS-PAGE analysis. The eluted fraction was diluted further with an equal volume of cold heparin A1 buffer.

Using the *hFEN1_336_Heparin* method, the protein solution was then applied to a 20 mL *HiTrap™ Heparin FF 16/10* column, pre-equilibrated with 2 column volumes of heparin A1 buffer. The protein was applied to the column, washed with heparin buffer A1 and eluted using 15 column volumes of heparin B1 buffer containing a linear gradient of 0 to 1 M NaCl. Fractions containing target protein were pooled and concentrated to less than 10 mL using a 250 mL Amicon Ultra-filtration cell. The membrane used has a molecular weight cut off with a 10 kDa (Millipore) pressurised with 40 psi nitrogen gas.

The protein was exchanged into the appropriate storage buffer using *hFEN1_Desalt_HiPrep* method. The protein was applied to 53 mL *HiPrep™ 26/10 Desalting* column and the fractions were collected in 1.5 mL tubes. Samples were concentrated using a Vivaspin (10,000 MWCO) spin column for 15 minutes at 4000 rpm, 4 ^{0}C.

2.1.8 Determination of protein concentration

Protein concentration was determined by the absorbance at 280 nm using a Nanodrop spectrophotometer and the calculated extinction coefficient for human FEN1, 22,920 M^{-1}

cm^{-1}. Once the protein was at a concentration of 200 µM in storage buffer (100 mM HEPES (pH 7.5), 200 mM KCL, 2 mM $CaCl_2$, 0.04% NaN_3, 20% glycerol, 10 mM DTT) an equal volume of glycerol was added to create a glycerol stock. The final protein concentration was 100 µM. The proteins were stored at −20 °C and purity of protein was assessed by SDS-PAGE with various concentrations (0.1 µg/µL until 1.6 µg/µL).

2.2 PCNA purification

Human PCNA was expressed previously by Dr L. David Finger using auto induction media and Bl21(DE3)-RILP cells. These were transformed with a pET41a vector harbouring the human PCNA gene cloned into the NdeI/XhoI sites in frame with the C-terminal $(His)_6$-tag. Cells were pelleted by centrifugation at 6000 × *g*, and the supernatant was discarded. The cells were re-suspended in 1X PBS and re-pelleted at 4700 × *g*. The supernatant was discarded, and the cell pellet was re-suspended in IMAC Buffer A1 (20 mM Tris pH 7.0, 1 M NaCl, 5 mM imidazole, 0.02% NaN_3) containing 1 mM βME. Once re-suspended, a solution of IMAC Buffer A1 containing 1 mM βME and 10 mg/mL lysozyme was added. The suspension was allowed to stand for one hour at 4 °C and was then stored at -20 °C.

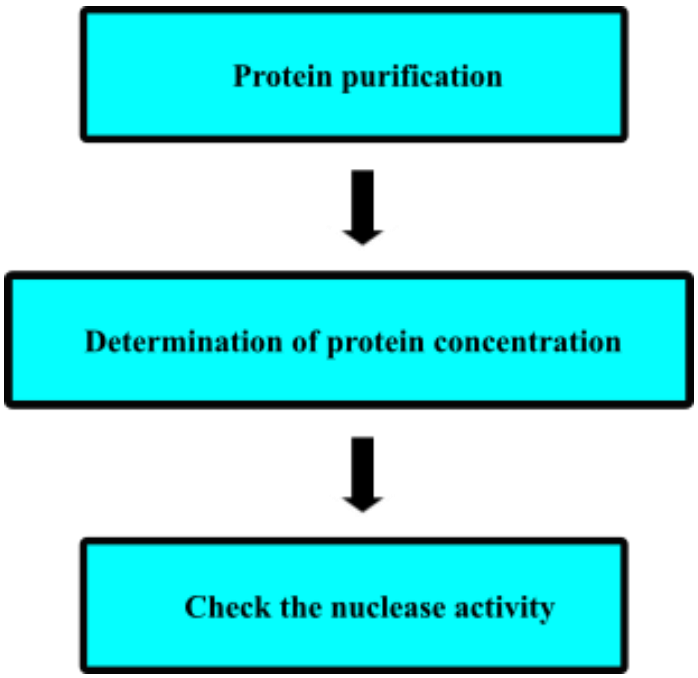

Figure 2.3: *The diagram of standard operating procedures for PCNA purification.*

Buffers

IMAC A1	20 mM Tris pH 7.0, 1 M NaCl, 5 mM imidazole, 0.02% NaN_3
PCNA IMAC A2 (1 L)	25 mM Tris pH 7.5, 100 mM NaCl, 0.17% Brij-35, 0.02% NaN_3, 10% glycerol, 5 mM imidazole
PCNA IMAC B1	500 mM NaCl, 0.02% NaN_3, 250 mM imidazole pH 7.4
PCNA IMAC B2 (1 L)	100 mM NaCl, 0.05% Brij-35, 0.02% NaN_3, 10% glycerol, 250 mM imidazole pH 7.4
PCNA Anion Exchange A1 (1 L)	25 mM Tris pH 7.5, 100 mM NaCl, 0.17% Brij-35, 0.02% NaN_3, 10% glycerol, 1 mM EDTA
PCNA Anion Exchange B1 (1 L)	25 mM Tris pH 7.5, 1000 mM NaCl, 0.17% Brij-35, 0.02% NaN_3, 10% glycerol, 1 mM EDTA
PCNA Heparin A1 (5 L)	25 mM NaH_2PO_4 pH 7.0, 0.17% Brij-35, 0.02% NaN_3, 10% glycerol
PCNA Heparin B1 (1 L)	500 mM NaH_2PO_4 pH 7.0, 0.01% Brij-35, 0.02% NaN_3, 10% glycerol
PCNA PS/A1 (1 L)	25 mM Tris pH 7.5, 0.17% Brij-35, 0.02% NaN_3, 10% glycerol, 1 mM imidazole
PCNA PS/B1 (5 L)	25 mM Tris pH 7.5, 0.02% NaN_3, 1 mM EDTA, 1.5 M $(NH_4)_2SO_4$
Stripping buffer	20 mM Na phosphate pH 7.4, 50 mM EDTA, 0.5 M NaCl, 0.02% NaN_3
PCNA SEC buffer (3 L)	100 mM HEPES pH 7.5, 200 mM KCl, 2 mM EDTA, 10 mM $Na_2S_2O_5$, 10 mM DTT, 0.01% Brij-35, 0.04% NaN_3

2.2.1 PCNA purification

The frozen cell suspension was thawed in a beaker of cold water until little lysate-ice was left whereby it was placed on ice. The viscous lysate was transferred into a beaker and sonicated on ice with 5 sec burst at 75% amplitude with 25 sec rest periods for a total of 1 min and 30 secs of burst phase. The sonication was repeated until the lysate was no longer viscous. Then, 5 mL of 1% Tween-20 solution in IMAC Buffer A1 was added and mixed. The lysate was clarified by centrifugation at 30,000 × *g* for one hour. The supernatant was poured into a beaker and kept on ice. The pellet was washed once with 1X PBS, and then,

re-suspended in 10 mM tris pH 7.5, 2% SDS. The re-suspended pellet was then set aside for further analysis later.

All protein purification steps were conducted in the cold room (4 °C) using an AktaPURE (GE Lifesciences). Two tandem 5 mL cobalt columns were charged with 0.1 M $CoCl_2$ according to the manufacturer's protocol. The column was equilibrated with 5 column volumes (CV) of IMAC Buffer A1. The PCNA clarified lysate (~100 mL) was then applied to column using a 150 mL superloop (GE Lifesciences). The column was subsequently washed with 5 CV of IMAC Buffer A1 and then 10 CV of PCNA IMAC Buffer A2. The target protein was eluted as a single fraction with 10 CV of PCNA IMAC Buffer B1. A 10 µL aliquot of the washes and the eluted fraction was retained for SDS PAGE.

The eluate from the Co^{2+}-IMAC column was diluted two-fold with water and then applied to two tandem 5 mL HiTrap™ Q FF column (2 x 5 mL in tandem) previously equilibrated with 5 CV of PCNA Anion Exchange A1. The protein was eluted using a 20 CV linear NaCl gradient (100 mM to 730 mM NaCl). Fractions containing hPCNA were confirmed by SDS-PAGE and then pooled. The pooled PCNA fraction was then dialysed against three × 1 L of Heparin A1 buffer at 4 °C.

The dialysate was then applied to two tandem 5 mL HiTrap™ SP HP (2 x 5 mL) columns equilibrated with 5 CV of PCNA Heparin A1. Note, this is a subtractive column whereby the protein of interest flows through the column, whereas the impurity that needs to be removed is retained on the column. The column was eluted using a 10 CV gradient with buffer PCNA Heparin A1 and B1. Confirmation of PCNA in the flow through was achieved by SDS-PAGE.

The flow through from the heparin column was then dialysed 3 × 1 L against PCNA Heparin A1. The dialysate was applied to a 26 × 100 mm hydroxyapatite-ultragel (Sigma) column that was equilibrated in PCNA Heparin A1. The column was then washed with 2 CV of PCNA Heparin A1. The protein was eluted from the column by linear salt gradient to 500 NaH_2PO_4 pH 7.0 (PCNA Heparin B1) and collected in 50 mL tubes. The presence of the protein was confirmed by SDS-PAGE analysis and protein-containing fractions were concentrated using Vivaspin-20 (10,000 MWCO). Concentrated protein was dialyzed against 3 × 1 L of PCNA PS/B1 buffer in the cold room.

The dialysate was then applied to a HiTrap Phenyl Sepharose (16 x 100 mm) previously equilibrated with PCNA PS/B1 Buffer. The column was washed with 2 CV of PCNA PS/B1 buffer and then the protein was eluted from the column by an inverse linear salt gradient generated using buffers PCNA PS/A1 and B1. PCNA containing fractions were detected by SDS PAGE then pooled and concentrated by ultrafiltration using a 50 mL Amicon Ultrafiltration cell with 10,000 MWCO PES membrane (Millipore) pressurized with N_2 (40 psi). The protein concentration was determined by absorbance at 280 nm using calculated extinction coefficients.

2.2.2 Determination of PCNA concentration

The concentrated protein sample was dialysed 3 × 1 L in SEC buffer at 4 °C. The protein concentration was re-measured as above and was concentrated to greater than 600 µM using a vivaspin-20 (4500 xpg at 4 °C for 10 min). The concentration of PCNA was measured again, and if above 600 µM was adjusted by the addition of SEC buffer. The volume of glycerol equal to the volume of protein solution was added to make a 300 µM PCNA stock solution in 50% glycerol. The aliquots were stored at -20 °C. The final concentration will be 300 µM of hPCNA, but as hPCNA is a trimeric protein, the concentration of $(hPCNA)_3$ is 100 µM.

2.2.3 Checking the nuclease activity of PCNA

To ensure that the PCNA preparation lacked nuclease activity, the procedures for the multiple turnover analysis of FEN1 on double flap substrate was used except only proliferating cell nuclear antigen (PCNA) was added at 0 µM (blank), 0.025 µM, 0.05 µM 1 µM and 10 µM using the UP31-9ST substrate. **Table 2.2** summarized the concentration for PCNA nuclease activity test.

Table 2.2: *Concentration of MT analysis for PCNA test of nuclease activity.*

Final [S] (nM)	S (µL)	MM (µL)	PCNA (µL)	Final [PCNA] (µM)	UREA (µL)	WAVE (µL)
50	20	160	18	0	50	60
50	20	160	18	0.025	50	60
50	20	160	18	0.05	50	60
50	20	160	18	1	50	60
50	20	160	18	10	50	60

2.3 Kinetics

Equipment

Instrument	denaturing High Performance Liquid Chromatography (dHPLC)
Detector	Fluorescence
Column	OLIGOSep™ Cartridge
Temperature	75 °C
Retention time	11 minutes
Flow rate	0.9 mL/min

Buffers

10X RB (50 mL)	0.5 M HEPES (pH 7.5), 1 M KCl, 0.08 M $MgCl_2$, 1 mg/mL BSA
5X RRB (1 mL)	500 µL 10X RB, 5 µL DTT (1 M), 495 µL H_2O
1X RRB (2 mL)	400 µL 5X RRB, 1600 µL H_2O
MM (2.8 mL)	0.63 mL 5X RRB, 2.17 mL H_2O
10X FB (50 mL)	0.5 M HEPES (pH 7.5), 1 M KCl
1X FB (10 mL)	1 mL 10X FB, 9 mL H_2O
ST quench (1 L)	50 mM EDTA, 1.5 M NaOH
10X RB quench (1 L)	0.5 M HEPES (pH 7.5), 1 M KCl, 0.08 M $MgCl_2$, 1 mg/mL BSA, 0.02% NaN_3

1X RRB quench (1 L)	100 mL 10X RB quench, 900 mL H_2O
1X RB quench (1 L)	55 mM HEPES (pH 7.5), 110 mM KCl, 8 mM $MgCl_2$, 15% glycerol, 0.02% NaN_3
Wave buffer A (5 L)	0.1% MeCN, 1 mM EDTA, 2.5 mM TBAB, H_2O
Wave buffer B (5 L)	70% MeCN, 1 mM EDTA, 2.5 mM TBAB, H_2O
Wave buffer C (5 L)	8% MeCN, H_2O
Wave buffer D (5 L)	80% MeCN, H_2O

2.3.1 Multiple turnover analysis

The DNA oligonucleotides listed in **Table 2.3** were purchased from DNA Technology A/S. The DNA substrates were prepared using the indicated oligonucleotide as 5 μM stocks by heating (95 ^{0}C) and annealing the appropriate oligos in a 1:1.5:1.5 ratio; downstream: upstream: template (**Table 2.4**) in 1X folding buffer (55 mM HEPES (pH 7.5), 110 mM KCl). Reactions were prepared with varying concentrations of substrate below K_M (2.5, 5.0, 7.5 and 10.0 nM) and pre-incubated at 37 °C in 1X RB (55 mM HEPES pH 7.5, 110 mM KCl, 8 mM $MgCl_2$, 0.1 mg/mL BSA, 1 mM DTT). Reactions were initiated with wild type human FEN1 concentrations empirically determined to give ~10% hydrolysis of substrate in 10 minutes. Once initiated, 20 μL aliquots of the reaction were removed at 2, 4, 6, 8, 10, 12 and 20 minutes and quenched in 50 μL of 250 mM EDTA. The analyses were carried out in triplicate. Quenched reactions were subsequently analysed by rp-ip-dHPLC on a WAVE system equipped with fluorescence detector (ADS Biotech – formerly Transgenomics) using a 4.5 X 50 mm OLIGOSep™ column (ADS Biotech) at 75 °C. Substrate and product peak separation were achieved using a gradient as shown in **Figure 2.4** of Buffer A (2.5 mM tetrabutylammonium bromide (TBAB), 1 mM EDTA, 0.1% acetonitrile) and Buffer B (2.5 mM TBAB, 1M EDTA and 70% acetonitrile). The ratio of product formed was determined by integration of the chromatograms and calculating the ratio of product using **Equation 2.1**;

$$\{Z_{product}/(Z_{product} + Z_{substrate})\} \times [S_0],$$

Equation 2.1

Where Z is the area under the peak and S_0 is the initial substrate concentration. Linear regression of a plot of concentration of product versus time of reaction produced slopes that are initial rates of reaction. Plots of normalized initial rates of reaction ($v_0/[E]_0$) versus substrate concentration produced a slope is the second order rate constant of the reaction (k_{cat}/K_M). Curve fitting was carried out using *GraphPad Prism 7 software.*

Table 2.3: *List of oligonucleotides.*

Oligo	**Sequence**	**Extinction Coefficient (Ɛ)**	**Molecular Weight (g/mol)**
Temp-48 (template)	/5Bio/GGT CCT ACT ACG ATT CAA GAG AGA GAC GCT GAG CTG AAC TGG ATC TGG 3’	472300	15278.1
F26-9ST (downstream)	/56-FAM/TTT TTT TTT TGA ATC GTA GTA GGA CC 3’	266460	8509.8
UP31-9ST	/5Bio/CCA GAT CCA GTT CAG CTC AGC GTC TCT CTC C3’	275300	9748.5
F26-9EQ	/56-FAM/TCT CTC TCT TGA ATC GTA GTA GGA CC 3’	262860	8449.7
UP31-9EQ	/5Bio/CCA GAT CCA GTT CAG CTC AGC GTC TCT CTC T 3’	275900	9763.5
UP30-9EQ	/5Bio/CCA GAT CCA GTT CAG CTC AGC GTC TCT CTC 3’	268100	9459.3
UP29-9EQ	/5Bio/CCA GAT CCA GTT CAG CTC AGC GTC TCT CT 3’	260600	9170.1
UP28-9EQ	/5Bio/CCA GAT CCA GTT CAG CTC AGC GTC TCT C 3’	252800	8865.9
UP27-9EQ	/5Bio/CCA GAT CCA GTT CAG CTC AGC GTC TCT 3’	245300	8576.7
UP26-9EQ	/5Bio/CCA GAT CCA GTT CAG CTC AGC GTC TC 3’	237500	8272.5
UP25-9EQ	/5Bio/CCA GAT CCA GTT CAG CTC AGC GTC T 3’	230000	7983.4

UP24-9EQ	/5Bio/CCA GAT CCA GTT CAG CTC AGC GTC 3’	222200	7679.2
UP23-9EQ	/5Bio/CCA GAT CCA GTT CAG CTC AGC GT 3’	214700	7390.0

Table 2.4: *Oligonucleotide combination used to make the substrate constructs the experiments.*

Downstream; 1	Upstream; 1.5	Template; 1.5
F26-9ST	UP31-9ST	Temp-48
F26-9EQ	UP31-9EQ, UP30-9EQ, UP29-9EQ, UP28-9EQ, UP27-9EQ, UP26-9EQ, UP25-9EQ, UP24-9EQ, UP23-9EQ	Temp-48

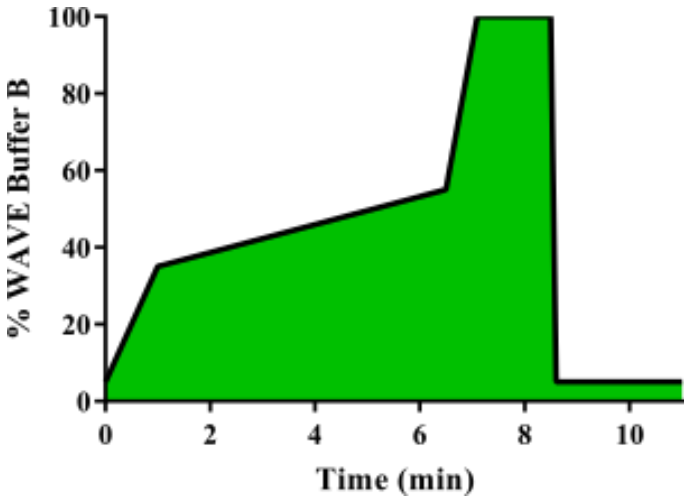

Figure 2.4: *Gradient for kinetics analysis at 75 ^{0}C by Denaturing HPLC used to separate the reaction products from substrate. The specific gradient switch points are as follows: 0 min 5% B 95% A, 1 min 35% B 65% A, 6.5 min 55% B 45% A, 7.1 min 100% B, 8.5 min 100% B, 8.6 min 5% B 95% A, 11 min 5% B 95% A.*

2.3.2 Single turnover analysis

Single turnover analysis was carried out using a RQF-63 quench flow device (Hi-Tech Sci Ltd, Sailsbury, UK) at 37 °C. The reactions between enzyme and substrate were carried out between 4.5 ms to 51241 ms (21 points) in 8 mM $MgCl_2$, 110 mM KCl, 55 mM HEPES (pH 7.5), 0.1 mg/mL BSA and 1 mM DTT. 80 µL of enzyme solution and substrate solution were injected and quenched with 80 µL of 1.5 M NaOH, 80 mM EDTA. The final concentrations of enzyme and substrate were 1 µM and 2.5 nM, respectively. The quenched reaction mixtures were removed from the sample loop and analysed using rp-ip-dHPLC equipped with a fluorescence detector as described above. The analyses were done in triplicate. The concentration of product was determined using **Equation 2.1**. The concentration of product with time was then non-linear regression fitted (*GraphPad Prism 7 software*) using **Equation 2.2**;

$$\{P\% = A1\ (1-e^{-k_{fast}*t}) + A2\ (1-e^{-k_{slow}*t})\}$$

Equation 2.2

Where A1 and A2 represent the amplitude @ span of the exponential phase for fast and slow phases respectively and k_{fast} and k_{slow} are the observed rate constant associated with the process and at t time k_{fast} was used to obtain first order rates of reaction.

2.4 Oligonucleotide purification

Equipment

Instrument	High Performance Liquid Chromatography (HPLC)
Detector	Tunable Absorbance Detector
Column	OLIGOSep® Prep HC Cartridge
Temperature	60 °C
Retention time	25 minutes
Flow rate	1.5 mL/min
Injection volume	60 µL

Buffers

Stock	1 M TEAAc pH: 7
Buffer A	0.1 M TEAAc pH: 7
Buffer B	0.1 M TEAAc pH 7, 25% acetonitrile
High salt buffer	10 mM Tris pH 8, 1 M NaCl, 1 mM EDTA, 0.02% NaN_3,
Low salt buffer	10 mM Tris pH 8, 100 mM NaCl, 1 mM EDTA, 0.02% NaN_3,

Oligonucleotide purification was achieved by HPLC using an OLIGOPrep™ (9.2 × 50 mm) eluting using a gradient as shown in **Figure 2.5** with a linear gradient Buffer A (100 mM triethylammonium acetate, 0.1% acetonitrile) and Buffer B (100 mM triethylammonium acetate, 25% acetonitrile). The peak corresponding to the oligonucleotide was collected and then applied manually to a 5 mL HiTrapQ column equilibrated with Low Salt Buffer. The column was then washed manually with 10 CV of Low Salt Buffer. The oligonucleotide was eluted manually with 10 CV of High Salt Buffer with fractionation into 1.5 mL tubes. The fractions containing oligonucleotide were determined using A260 and pooled. The oligonucleotide was then desalted using NAP-25 desalting columns according to the manufacturer's protocol. Oligonucleotides were dried in a speed vac, and then, re-suspended in water. The concentration of the oligonucleotide stocks were determined by UV spectrophotometry (A_{260}) and calculated extinction coefficients.

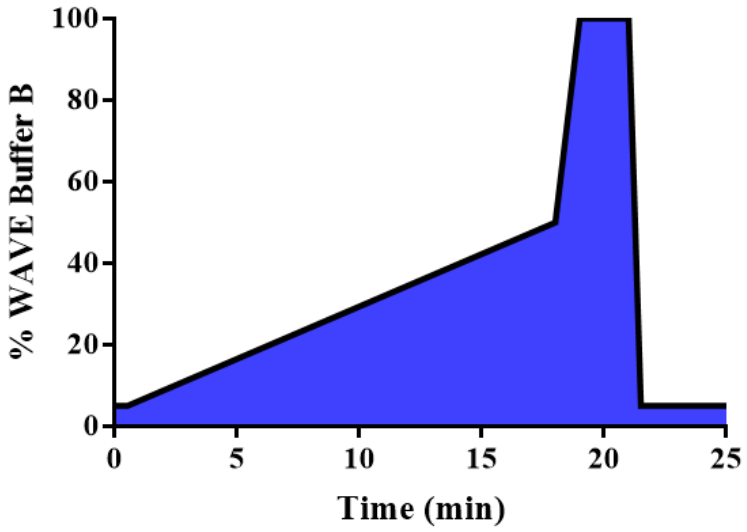

Figure 2.5: *Gradient for oligonucleotide purification analysis by HPLC used to collect the pure oligonucleotide. The specific gradient switch points are as follows: 0 min 5% B 95% A, 0.5 min 5% B 95% A, 18.0 min 50% B 50% A, 19.0 min 100% B, 21.0 min 100% B, 21.5 min 5% B 95% A, 25.0 min 5% B 95% A.*

2.5 Threading analysis

Buffers

10X CaRB (50 mL)	250 mM HEPES (pH 7.5), 0.5 M KCl, 20 mM $CaCl_2$, 1 mg/mL BSA
5X CaRB (2 mL)	1000 µL 10X CaRB, 5 µL 1 M DTT, 995 µL H_2O
1X CaRB (2 mL)	400 µL 5X CaRB, 1600 µL H_2O
10X MgRB (50 mL)	250 mM HEPES (pH 7.5), 0.5 M KCl, 160 mM $MgCl_2$, 1 mg/mL BSA
1X MgRB (2 mL)	200 µL 10X MgRB, 1800 µL H_2O
MM (2 mL)	572 µL 5X CaRB, 1430 µL H_2O
10X FB (50 mL)	0.5 M HEPES (pH 7.5), 1 M KCl
1X FB (10 mL)	1 mL 10X FB, 9 mL H_2O
Bench quench (1 L)	8 M urea, 80 mM EDTA
RQF-63 quench (1 L)	8 M urea, 300 mM EDTA

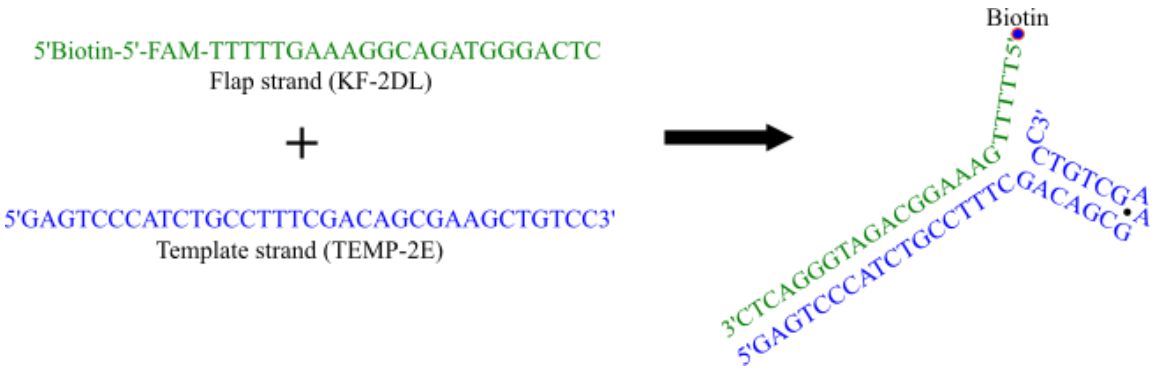

Figure 2.6: *SB5-1 structures were prepared by KF-2DL as a flap strand, which have biotin for streptavidin attach on the DNA and FAM for detection by fluorescence detector and TEM-2E as a template strand for threading analysis.*

All substrate (**Figure 2.6**) binding reactions were prepared with 100 nM of the biotinylated substrate SB5-1 in 50 mM HEPES pH 7.5, 100 mM KCl, 4 mM $CaCl_2$, 0.2 mg/ml BSA, and 2 mM DTT. The DNA substrates were prepared using the indicated oligonucleotide as 5 µM stocks by heating (95 ^{0}C) and annealing the appropriate oligonucleotides in a 1:1.1 ratio between flap and template strand in 1X folding buffer.

As can see in **Figure 2.7**, premixed complexes were prepared by incubating the 5 nM substrate with indicated 500 nM hFEN1 protein for 2 minutes at room temperature in master mixed (MM) and continued by adding 18 µL 1X Ca buffer and equilibrated for 5 minutes in room temperature. Trapped complexes refer to binding reactions that were prepared by incubating the 5 nM substrate with the indicated 500 nM hFEN1 protein for 2 minutes at room temperature in MM. Then streptavidin was added to 5 times excess of substrate (dissolved in 1X CaRRB) and equilibrated for 5 minutes at room temperature. Blocked complexes refer to binding reactions that were prepared by incubating the substrate with a 5 molar excess streptavidin for 5 minutes in room temperature in MM. Then continued by adding the indicated 500 nM hFEN1 protein and equilibrated for 2 minutes. All the protocols were done at room temperature.

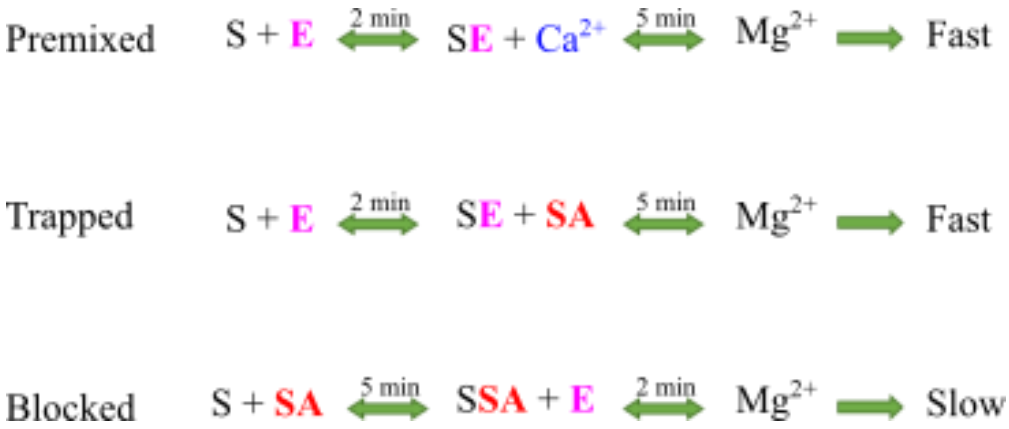

Figure 2.7: *Threading analysis procedures for premixed, trapped and blocked. (S: substrate, E: enzyme, **SA**: streptavidin).*

After the addition of the second component in both blocked and trapped binding reactions, the reactions were incubated for a further 1 minute at 37 ^{0}C. Then, 20 µL of aliquots were took out before the reaction was started as a control for the experiment. Then, magnesium reaction buffer (25 mM HEPES (pH 7.5), 0.05 M KCl, 16 mM $MgCl_2$, 0.1 mg/mL BSA) was added to the samples to initiate the reaction at 37 ^{0}C. The reaction was quenched in a quench solution of 8 M urea and 80 mM EDTA at varying time points on the bench (Premix: 10 s, 20, 30, 40, 1 minute, 2, 4, 8, 16, 30, 60, 120; Trapped: 10 s, 20, 30, 40, 1 minute, 2, 4, 8, 16, 30, 60, 120, 240, 360 and Blocked: 30 s, 1 minutes, 2, 4, 8, 16, 30, 60, 120, 180, 240, 360). Further analyses were done (especially for premixed and trapped) at lower delay times using a RQF-63 quench flow device at 37 °C. The reaction was quenched in a quench solution of 8 M urea and 300 mM EDTA at varying time (between 30.6 ms to 14041 ms). Final substrate and enzyme concentrations were 5 nM and 500 nM respectively. Quenched reactions were subsequently analysed by rp-ip-dHPLC on a WAVE system equipped with fluorescence detector at 50 ^{0}C. The specific gradient switch points are as follows: 0 min 5% B 95% A, 1 min 30% B 70% A, 5.5 min 55% B 45% A, 7.1 min 100% B, 10.0 min 100% B, 10.1 min 5% B 95% A, 12.5 min 5% B 95% A (**Figure 2.8**).

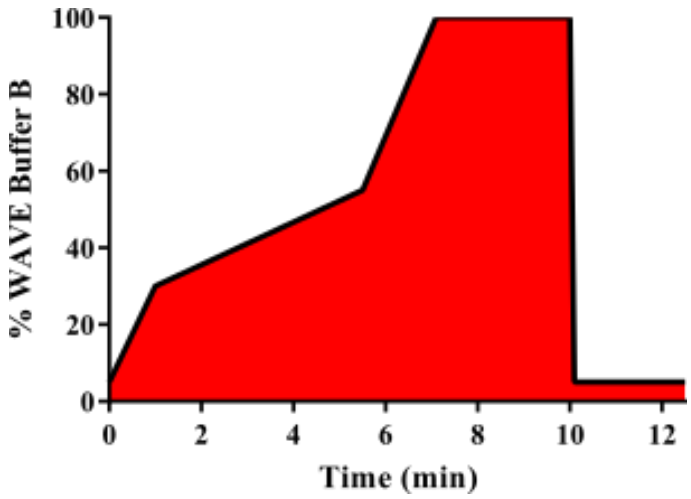

Figure 2.8: *Gradient for kinetics analysis at 50 ^{0}Cby Denaturing HPLC used to separate the reaction products from substrate. The specific gradient switch points are as follows: 0 min 5% B 95% A, 1 min 30% B 70% A, 5.5 min 55% B 45% A, 7.1 min 100% B, 10.0 min 100% B, 10.1 min 5% B 95% A, 12.5 min 5% B 95% A.*

2.6 2-aminopurine Exciton-Coupled Circular Dichroism (ECCD) Spectroscopy

Buffers

10X buffer (Ca^{2+})	0.5 M HEPES (pH 7.5), 100 mM $CaCl_2$, 1 M KCl,
10X FB	0.5 M HEPES (pH 7.5), 1 M KCl,
Storage buffer (SB)	100 mM KCl, 1 mM $CaCl_2$, 5 mM DTT, 50 mM HEPES (pH 7.5), 0.02% NaN_3, 50% glycerol

The concentration of the final stocks was measured by UV Nanodrop at 260 nm and using extinction coefficients generated by IDT oligo analyser 3.1 tool (http://www.idtdna.com/calc/analyzer). Solution of substrate oligonucleotides (100 μM) were prepared in 0.5 m HEPES (pH 7.5) and 1 M KCl from appropriate constituent oligonucleotides. The substrates with 2-aminopurine were bought from DNA Technology were annealed at 95 ^{0}C for 5 minutes and cooling to room temperature.

A mixture of 1X FB and storage buffer (SB) was used as a blank and 10 µM substrates with none template (ssFEC1) incubated in 100 mM KCl, 1 mM $CaCl_2$, 5 mM DTT, 50 mM HEPES (pH 7.5), 0.02% NaN_3, 50% glycerol was used as a control. Samples containing 10 µM of the indicated DNA construct, 55 mM HEPES pH 7.5, 110 mM KCl, 1 mM DTT, and either 10 mM $CaCl_2$ or 10 mM $CaCl_2$ + 25 mM EDTA and appropriate 12.5 µM indicated protein (hFEN1, R103A, R125A and R129A) where incubated at 20 °C for 10 minutes.

ECCD spectra were collected using a quartz cuvette with a 5 mm path length and with subsequent acquisition at 20 ^{0}C of ECCD spectra (300–480 nm) using an *Applied Photophysics Chirascan* CD spectrophotometer. ECCD spectra are an average of two repeats in set recorded in 0.8 nm step size (time per point: 0.5 s) that were baseline corrected using spectra recorded on samples containing the same components, but lacking DNA. The baseline-subtracted spectra were smoothed using *Chirascan Pro-Data* software (version : Chirascan v.4.4.2.0). The signal was normalized to account for baseline shift observed for day to day variation. To normalize for baseline shifts, the signal was averaged between 400-480 nm, a region in which 2-APs has no absorbance and was subtracted from the associated spectrum. The ECCD spectra data were analysed and plotted as Δε per mol 2-AP residue versus wavelength using *GraphPad Prisam 7 Software*. Each measurement was independently repeated and gave identical results.

Chapter 3: Migrating Flap Substrates

To date most *in-vitro* studies of human FEN1 reactions have used static (non-equilibrating) flap substrates bearing non-complementary 5' and 3' flaps, where the single stranded regions are not complementary to the template (continuous) DNA. However, when flaps occur *in-vivo*, they are migrating or equilibrating structures as the flaps can base pair with template DNA. This study will focus on how the full length wild type human flap endonuclease-1 (WT-hFEN1) acts on migrating flap substrates. The behaviour of migrating flap substrates will be compared to those of static flaps.

During Okazaki fragment maturation, there are two regions to a double-flap substrate. The downstream duplex that has a 5'-flap generated by a polymerase on the previous Okazaki fragment, whereas the upstream duplex is generated by a polymerase from the latest Okazaki fragment synthesised. *In-vivo*, the double-flap of substrates are equilibrating because the Okazaki fragments are synthesised from the same parental template during the replication of new strand[41]. As a consequence of this, the length of 3' flap and 5' flap can vary giving rise to a number of different conformations. However, FEN1 is thought to act upon the substrate in one particular conformer that is a conformer with single nucleotide 3'-flap and a 5'-flap of any length (**Figure 3.1**).

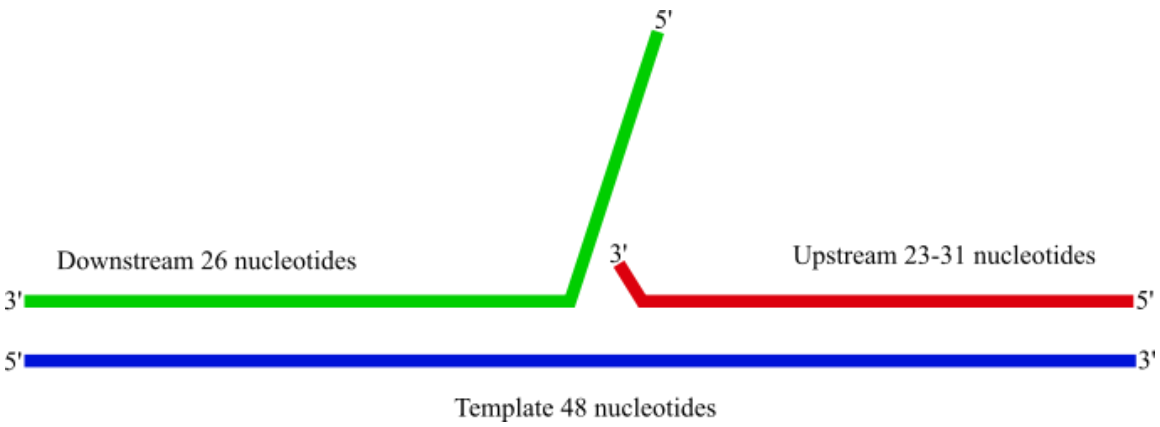

Figure 3.1: *General structures of double flap for this study.*

To investigate how the migrating flap substrates affect the ability of hFEN1 to catalyse the reaction, nine different migrating flap substrates were designed by Dr. L David Finger. For this research, the migrating flap substrate were designed to have between 1 to 9

overlapping base pairs as shown in **Table 3.1**. The continuous template strand, Temp-48 is 48 nucleotides bases long in all cases and forms a downstream double-stranded DNA (dsDNA) of variable length. As a comparison for each result in the analysis, a static substrate, that has an 8 nucleotides 5'-flap was be used as a control. As shown in **Figure 3.2**, each flap substrate was constructed by annealing three oligonucleotides; a 26 nucleotide 5'-FAM labelled strand that formed the downstream duplex, a 48 nucleotide 5'-biotinylated DNA as complementary template strand and a variable upstream 3'-flap strand with 5' biotin. Biotin was attached to allow detection of the oligonucleotides using a streptavidin antibody system. In addition, the relatively small biotin (244.3 Da) was unlikely to perturb the function of the substrate. The 5'-FAM allows flap removal to be monitored. The oligonucleotides can be used to create a static double flap control and migrating flaps. hFEN1 is predicted to cleave at a specific site location when the substrate adopts a conformer with a single nucleotide 3'-flap.

Table 3.1: *A variable 10 upstream 3'-flap strand with 5' biotin. The length of 5' flap between zero to eight nucleotides to form different length single-strand 5' flap products. Each substrate were annealed at 95 ^{0}C for 5 minutes with the ratio 1 (downstream), 1.5 (upstream) and 1.5 (template) in 100 mM KCl, 50 mM HEPES-KOH pH 7.5 and 0.02 % NaN_3.*

Construct (FAM, Temp-48)	**Flap length of cleaved conformer**		**Potential conformers**	**Number of overlapping base pairs**	**Length of nts for ssDNA product**
	5'	**3'**			
UP31-9ST	8	1	1	0	9
UP23-9EQ	0	1	1 of 2	1	1
UP24-9EQ	1	1	1 of 3	2	2
UP25-9EQ	2	1	1 of 4	3	3
UP26-9EQ	3	1	1 of 5	4	4
UP27-9EQ	4	1	1 of 6	5	5
UP28-9EQ	5	1	1 of 7	6	6
UP29-9EQ	6	1	1 of 8	7	7
UP30-9EQ	7	1	1 of 9	8	8
UP31-9EQ	8	1	1 of 10	9	9

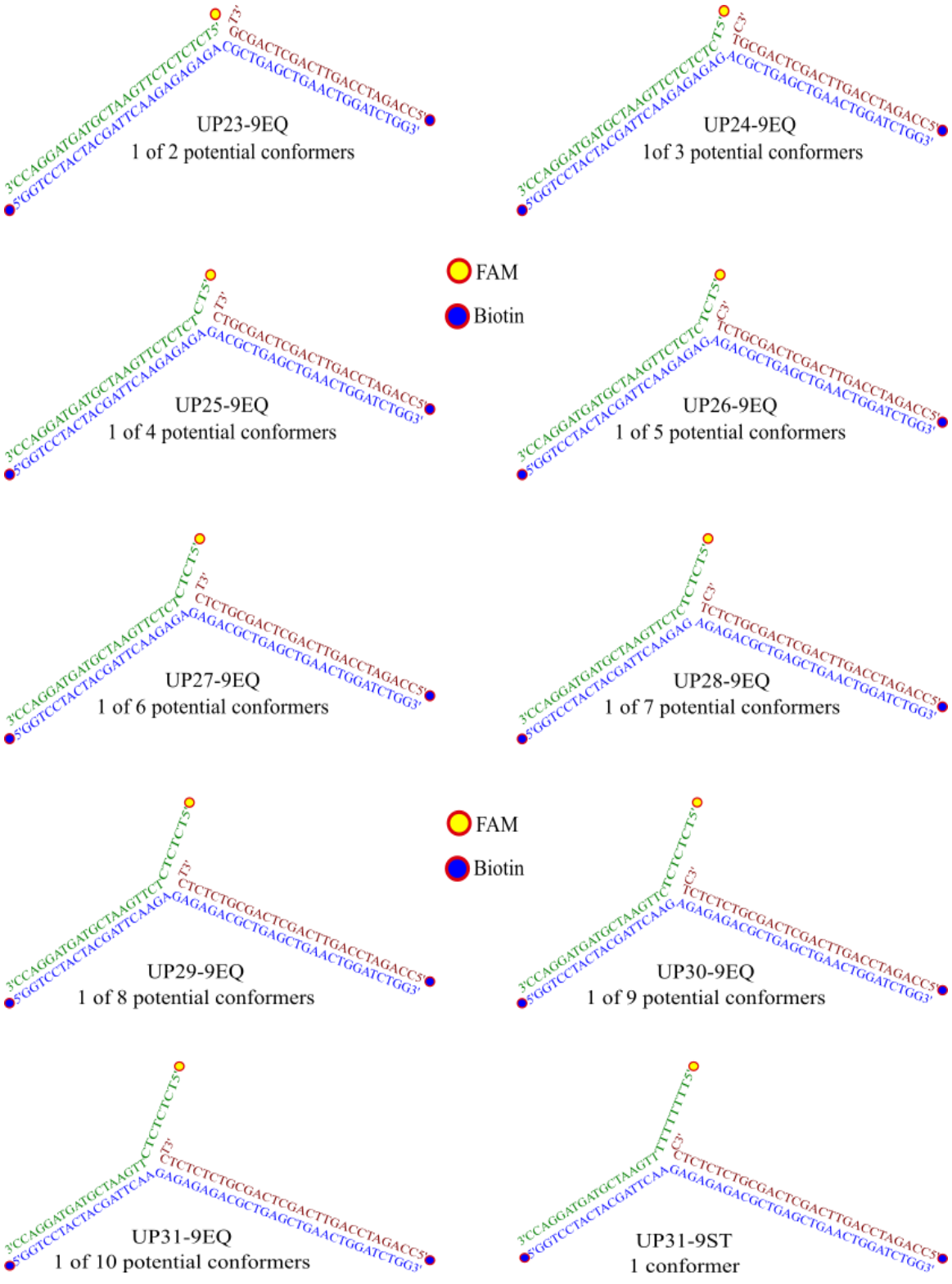

Figure 3.2: *A schematic of the designed DNAs with 5'-FAM for reaction monitoring and 5'-biotin on the template and upstream strands. This illustration of the migrating flaps shows one of the potential conformers with a single nucleotides 3' flap and corresponding length 5' flap. The static flap, 8 nts 5' flap and 1 nt 3' flap (UP31-9ST) used as a control for the analysis. The template strand is common to each structure whereas the flap and upstream strands are different.*

3.1 Full length human FEN1 purification

To carry out this study, full length wild type hFEN1 bearing an C-terminal $(His)_6$-tag was purified following expression in *Escherichia coli* BL21(DE3)-RILP using auto-induction methods as described in section 2.1. Proteins were purified using four different chromatographic media including IMAC, anion-exchange, heparin and desalting column. Initially, the target protein was captured by an IMAC column and was then further purified by anion exchange chromatography to remove nucleic acid contamination that coelutes with the protein. To remove further contaminants, the protein purification was further polished by heparin affinity chromatography. The protein was concentrated and exchanged into storage buffer. The protein purity was assessed by SDS PAGE as shown in **Figure 3.3**. The single band from each of concentrations; 0.1, 0.2, 0.4, 0.6, 0.8, 1.0 and 1.6 µg/µL proved that the human FEN1 was totally pure and not contaminated by other proteins.

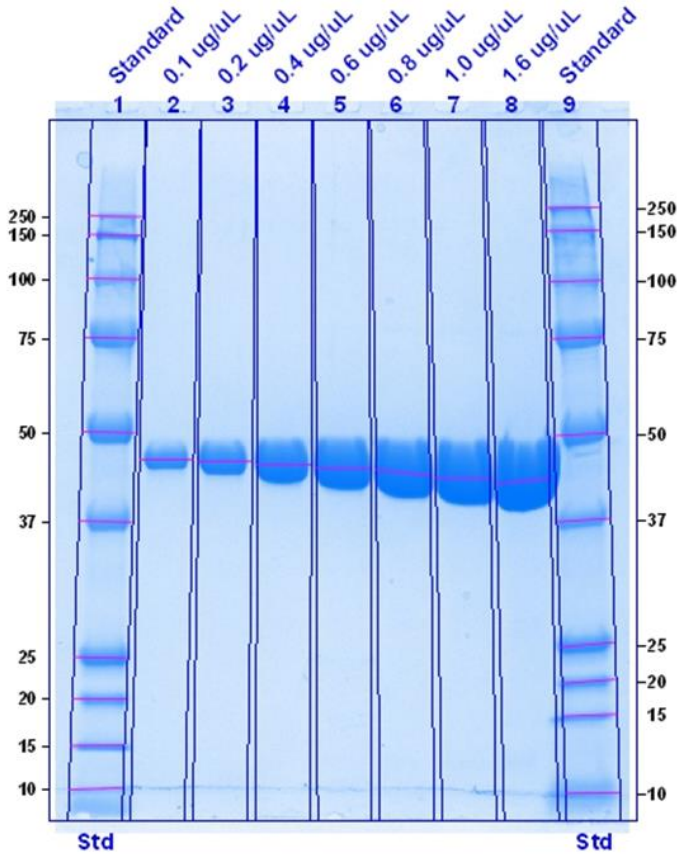

Figure 3.3: *SDS PAGE gel showing the purity of full length human wild type FEN1 after purification and concentration steps. Lane 1 and lane 9: Protein ladder (representing 10, 15, 20, 25, 37, 50, 75, 100, 150 and 250 kDa), lane 2: 0.1 µg/µL, lane 3: 0.2 µg/µL, lane 4: 0.4 µg/µL, lane 5: 0.6 µg/µL, lane 6: 0.8 µg/µL, lane 7: 1 µg/µL and lane 8: 1.6 µg/µL.*

3.2 Kinetics studies

Measuring the kinetics of hFEN1 reactions allows for a comprehensive comparison of altered conditions and reagents that can ultimately lead to a mechanistic understanding and the elucidation of structure-function relationships. Comparing measured rates of reaction with static and migrating substrates will allow us to understand the influence migrating flaps on hFEN1 activity. hFEN1 prefers to bind at a double flap DNA structure with 1 nt 3'-flap, because occupation of the 3'-flap binding site by a single 3'-flap nucleotide is crucial to reaction. However, detailed kinetic information with migrating flap substrates is lacking.

Both multiple and single turnover kinetic analyses are routinely employed to determine the rate of reaction of human FEN1-catalyzed hydrolysis. Under second order or steady-state multiple turnover (MT) conditions where substrate concentration is in a large excess to the enzyme and lower than K_M, the catalytic efficiency (k_{cat}/K_M) of hFEN1 acting on a static double flap is close to the bimolecular diffusion association rates in solution (k_{on}) of roughly 10^7-10^{10} M^{-1} s^{-1}. Typically under MT conditions, enzyme must catalyse multiple hydrolyses of substrate to create detectable amounts of product. Because substrate hydrolysis decreases the substrate concentration only the first 10% of the reaction is monitored[37].

The proposed enzyme kinetics reaction pathway with static double flap substrates is shown in **Figure 3.4**. The scheme shows that enzyme and substrate form an encounter complex that must change conformation to become catalytically competent. Once hydrolysis has occurred, the 5'-flap product (P) is thought to be released and an enzyme product complex (EN) is former with the dsDNA product. This must them dissociate for turnover to occur. If we allow the substrate to have many conformations and all can bind to the enzyme, albeit with different affinities, then there are also at least ES^1, ES^2, ES^3 until ES^n states (where n is the number of potential conformers) possible. It is also likely that the ES^n states must reach the ES state to then proceed through these conformational changes as well to become cleavage competent. An alternative model, which could also be represented by **Figure 3.4** assumes only one conformer (that with a single nt 3'-flap) can bind to the enzyme, but that substrate conformers can readily interconvert in solution.

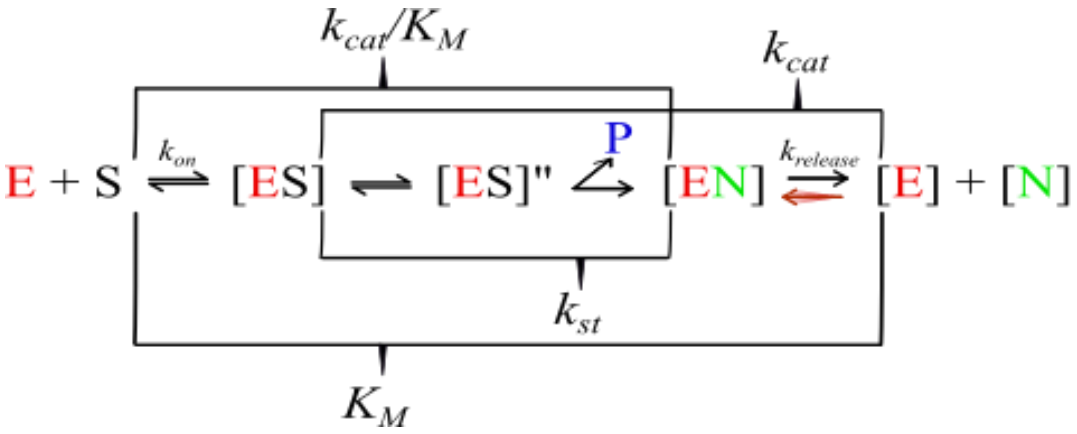

Figure 3.4: *The proposed reaction pathway of FEN1 catalysed reaction of double flaps substrates*[37]*. Where E; enzyme, S; substrate, ES; P; the flap or single-stranded product and N; nicked or double-stranded product.*

Downstream, upstream and template strands were mixed with specific ratios and heated at 95 ^{0}C for 5 minutes and then annealed in 100 mM KCl, 50 mM HEPES-KOH pH 7.5 and 0.02 % NaN_3 to pre-form a static or migrating double flap substrate. Reverse-phase ion-pairing denaturing high performance liquid chromatography (dHPLC), using tetrabutylammonium bromide (TBAB) as the interface with solid phase was utilized along with a fluorescence detector to monitor and quantify the cleavage reaction catalysed by human FEN1. In the analysis, only the fluorescein-labelled product and substrate were observed, and these were quantified by integration using the WAVE^{TM} system software to calculate the percentage product at each time point. The concentrations of product produced were calculated following **Equation 3.1**.

$$[P] = \frac{Area(product)}{Area(product) + Area(substrate)} \times [S]_{Total}$$

Equation 3.1

Under the substrate-limiting multiple turnover analysis, the specific substrates concentrations were prepared and initiated with an appropriate enzyme concentration and measured by discontinuous assay at 37 ^{0}C. A linear relationship between product concentration (nM) versus time (minutes) demonstrate that initial rates (v_0 min^{-1}) under steady state conditions were being calculated. The initial rates (v_0) were determined by linear regression and normalised by dividing by enzyme concentration $[E]_0$. The

normalised rates of reaction ($v_0/[E]_0$) versus substrate concentration [S] were measured three times and plotted. Linear regression using *GraphPad Prism 7 software* was used to acquire k_{cat}/K_M rate constants (gradient).

Under maximal single turnover (ST) conditions, where the enzyme is in excess of the substrate and adding more enzyme does not increase the rate of reaction, the rate of reaction is the decay of the ES complex. The maximal MT rate (k_{cat}) is slower than the respective ST maximal rate when enzyme is saturating, suggesting that product release ($k_{release}$) is the rate-limiting step under saturating substrate conditions. During the human FEN1 catalytic activity, there are additional steps to substrate binding and catalysis. The enzyme and substrate undergo a series of distinct conformational changes before it is ready for hydrolysis including the threading of the DNA flap, helical arch ordering, un-twisting of DNA nucleotides and finally phosphodiester hydrolysis. Thus, using ST conditions at time points faster than enzyme-product release allows us to assess the effects of these additional steps in our reaction scheme.

3.2.1 Kinetic characterisation of full length WThFEN1

The kinetic parameters of full length WThFEN1 were determined using a static double flap substrate with a 5 nts 5'-flap and a 1 nt 3'-flap (DF5,1) (**Figure 3.5**) and 5'-fluorescein (FAM) similar to the substrates that have been used previously by Dr. Jack Exel[121]. The full Michaelis-Menten curve was analysed, under conditions where the $[E] \ll [S]$ at pH 7.5 and at 8 mM $MgCl_2$. The final concentration of the substrates were 5, 10, 25, 50, 75, 100, 250, 500, 750, 1000, 2500 and 5000 nM with the final enzyme concentration less than 0.01% of that of the substrate concentration. Only the initial phase (10% or less) of product formation was used to determine the rate of reaction. The formation of single strand and nicked products were assessed by dHPLC and calculated via **Equation 3.1**. Each substrate was analysed in triplicate at different concentrations. A plot of the normalized initial rates of reaction $v_0/[E]_0$, min^{-1} versus substrate concentration was curve fitted to the Michaelis-Menten equation as shown in **Figure 3.6**. The variation of the concentration of substrates selected ensures that data are collected above and below the suspected K_M.

$$V_O = \frac{V_{max}[S]}{K_M + [S]}$$

Equation 3.2

Where,

$$K_M = \frac{k_{cat}}{k_{cat}/K_M}$$

Equation 3.3

Where V_0; initial reaction velocity, S; substrate concentration, K_M; the Michaelis constant and V_{max}; the maximum reaction velocity (all the active sites are saturated). While K_M is the substrate concentration at half of the maximal rate.

Figure 3.5: *The structure of static double flaps, 5 nucleotides 5'-flap and 1 nucleotide 3'-flap and 5' FAM labelled*[74].

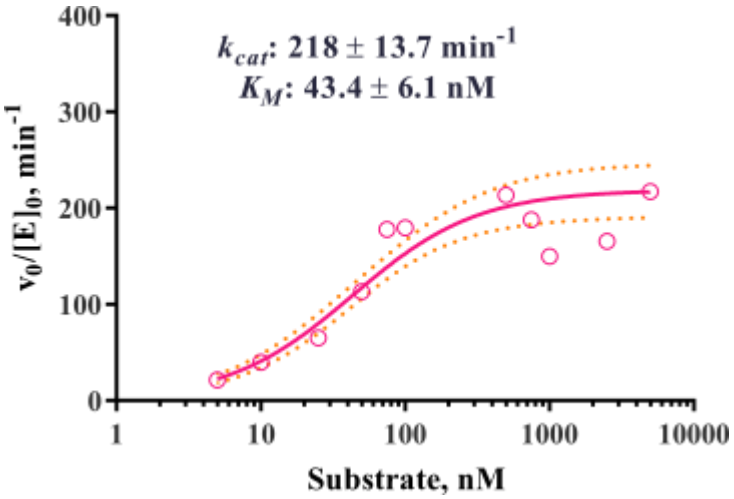

Figure 3.6: *Michaelis-Menten plots for endonucleoytic cleavage of DF5,1 by new batch of full length human WTFEN1. Result of fitting data using non-linear regression of the Michaelis-Menten equation.*

The results show that initially the enzyme activity increases linearly with substrate concentration. Then, $v_0/[E]_0$, follows saturation reactions with respect to [S]. By plotting the curve using non-linear regression and the Michaelis-Menten model in *GraphPad Prism 7 software*, the value of k_{cat} and K_M can be measured. In this analysis, the value of K_M is 43 ± 6 nM, which is double the previous result of 20 ± 3 nM[121]. The value of k_{cat} is 218 ± 14 min^{-1}, slightly higher than measured by Dr. Jack Exell at 165 ± 9 min^{-1}.

The single turnover rates of substrate hydrolysis were measured to complement the multiple turnover measurements of human FEN1. The final enzyme concentration was 10-fold higher than K_M value and the substrate concentration was 2.5 nM. Data were fitted to **Equation 2.2** in section 2.3.2, and each single turnover data point is representative of at least three replicates (**Figure 3.7**). The rate was significantly faster than maximal multiple turnover rates of reaction with a value of 723 ± 59 min^{-1} as shown in **Table 3.2.** The multiple turnover parameters are the rate limited by product release as discovered by earlier FEN studies[37]. The measurement is slightly slower when compared with previous values of 918 ± 50 min^{-1} [122].

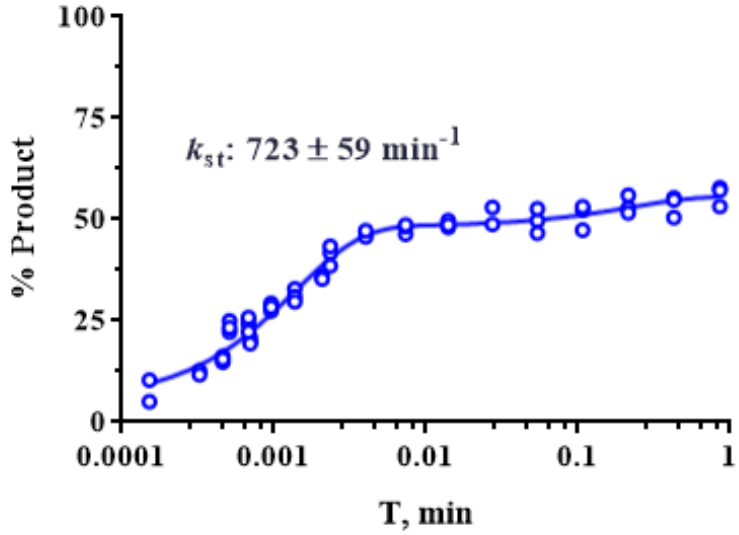

Figure 3.7: *The k_{st} rate for DF5,1 cleavage by human FEN1 protein. The [S] was 2.5 nM and the final enzyme concentrations was 450 nM. Reactions were performed in 55 mM HEPES-NaOH pH = 7.5, 8 mM $MgCl_2$, 1 mM DTT, 110 mM KCl and 0.1 µg/µL BSA and at 37 ˚C.*

Table 3.2: *Comparison of catalytic of DF5,1 with full length WThFEN1 new and old batch*[121].

Protein	k_{cat}, min^{-1}	K_M, nM	k_{st}, min^{-1}
Full length WThFEN1 new batch	218 ± 14	43 ± 6	723 ± 59
Full length WThFEN1 old batch	165 ± 9	20 ± 3	918 ± 50

3.2.2 Oligonucleotide purification

When oligonucleotides were purchased initial purification was requested. Three oligonucleotides have been further purified; 48 nucleotides template (Temp-48), 31 nucleotides static upstream (UP31-9ST) and 26 nucleotides static downstream (F26-9ST) to see if there are any differences in the substrate properties of initially-purified versus extra-purified oligonucleotides. The second order rate of reaction analyses were determined by heating and annealing downstream, upstream and template strands in 1:1.5:1.5 ratio, respectively in buffer to form a static double flap substrate; 8 nts 5'-flap

and a 1 nt 3'-flap and 5'-fluorescein (FAM) labelled. Multiple turnover analysis were conducted, final [S] were 2.5, 5.0, 7.5 and 10.0 nM and final [E] were 0.5 pM and 1 pM for 10% of products in 8 mM $MgCl_2$, pH 7.5 at 37 °C.

Results showed that there were no significant differences between initial purified and extra-purified oligonucleotides. The slope of plots of $v_0/[E]_0$ versus [S] were determined, as shown in **Figure 3.8** for initial purified and extra-purified substrates which yielded k_{cat}/K_M of 6.29 ± 0.59 $nM^{-1}min^{-1}$ and 4.58 ± 0.53 $nM^{-1}min^{-1}$, respectively. The 95% confidence intervals slope were between 3.40 to 5.77 for extra-purified and 4.98 to 7.59 for the initial. In this useful analysis, it is believed that the initial-purified and extra-purified substrates acted similarly.

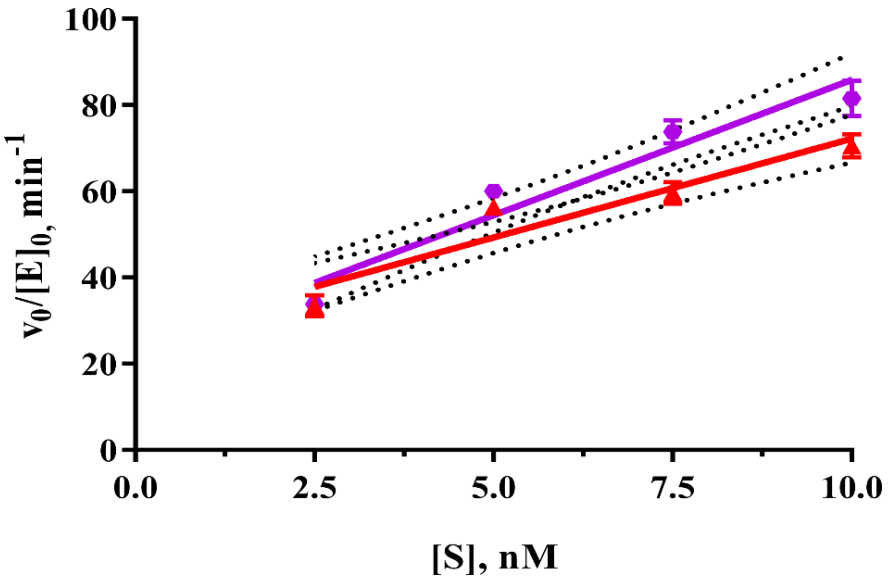

Figure 3.8: *The means of oligonucleotide purification analysis for second order rate analysis on static double flap substrate; 8 nts 5'-flap and a 1 nt 3'-flap (UP31-9ST) and 5'-fluorescein (FAM) labelled. Purple and red represent the initial purified and extra-purified substrates respectively. Errors represent standard errors of experiments performed in at least triplicate.*

3.2.3 Determination the ability of human FEN1 to process substrates that can adopt multiple conformers

For steady state analysis, reaction mixtures were prepared at lower concentrations of the substrates of 2.5, 5.0, 7.5 and 10 nM to determine k_{cat}/K_M for the ten different double flap substrates. The range of enzyme concentrations were between 0.5 to 12 pM. To examine the rate of the human FEN1 *in-vitro*, we carried out the MT analyses on the different substrates with variable numbers of potential conformers. Each measurement was the average of at least three independent experiments at 37 ^{0}C. The reaction mixture were removed and quenched at seven time points; 2, 4, 6, 8, 10, 12 and 20 minutes after initiation of reaction. The reaction was stopped by addition of EDTA to chelate the essential divalent metal ions. It was hypothesised that as the number of potential conformers increase the second order rate constants would get lower.

Sample dHPLC chromatograms of the fluorescence intensity (mV) of product and substrate versus retention time (min) are shown in **Figure 3.9**. From the analysis with 4 different concentrations of each double flap substrate, there is only one product peak appearing in the reactions of the migrating substrates, despite their ability to adopt multiple conformers to different degrees. This demonstrates the ability of hFEN1 to hydrolyse in only one place despite the various length of nucleotides and the ability of substrates to adopt varying conformations.

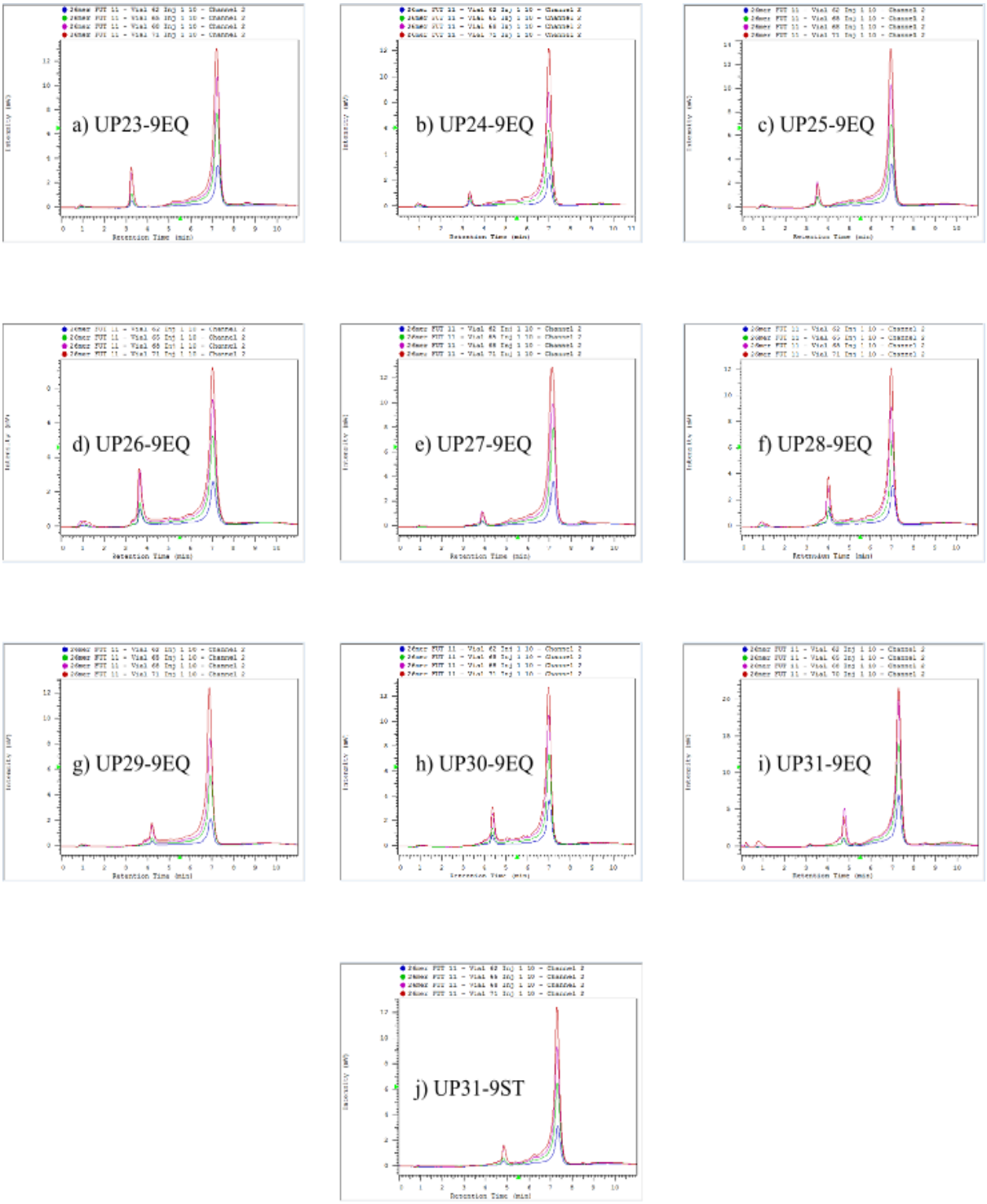

Figure 3.9: *Multiple turnover analyses of the hFEN1 catalysed reactions with 10 different double flap migrating and static substrates. Each of the figure represents; a) for UP23-9EQ, b) for UP24-9EQ, c) for UP25-9EQ, d) for UP26-9EQ, e) for UP27-9EQ, f) for UP28-9EQ, g) for UP29-9EQ, h) for UP30-9EQ, i) for UP31-9EQ and j) for UP31-9ST. The separation by dHPLC at 75 ^{0}C produced two peaks; single-strand products (small peak) and the substrate (the bigger peak).*

Furthermore, the retention time for each single peak of product is proportional with the predicted length 5' single stranded DNA product assuming reaction one nucleotide into the downstream duplex and that the only conformers undergoing reaction all have a single nucleotide 3'flap. **Figure 3.10** illustrates that the retention time (min) for the peak of single-stranded product appeared between 3 to 5 minutes for 10 different sets of double flap substrates. The UP23-9EQ substrate has a one nucleotide 5' flap ssDNA product having the shortest retention times at 3.2 minutes whereas the products from UP31-9EQ and UP31-9ST substrates appeared at 4.8 and 5 minutes respectively. Both substrates have 9 nucleotides 5' flap ssDNA products. The retention times for the products of UP24-9EQ, UP25-9EQ, UP26-9EQ, UP27-9EQ, UP28-9EQ, UP29-9EQ and UP30-9EQ were 3.3, 3.5, 3.7, 3.9, 4.0, 4.2 and 4.4 minutes respectively. Thus the shorter the 5' flap single stranded DNA products detected by the fluorescence detector, the faster the peak will appear in chromatogram. The results show that the number of nucleotides in the 5'-flap is proportional to the length of the ssDNA product and also the number of overlapping base pairs.

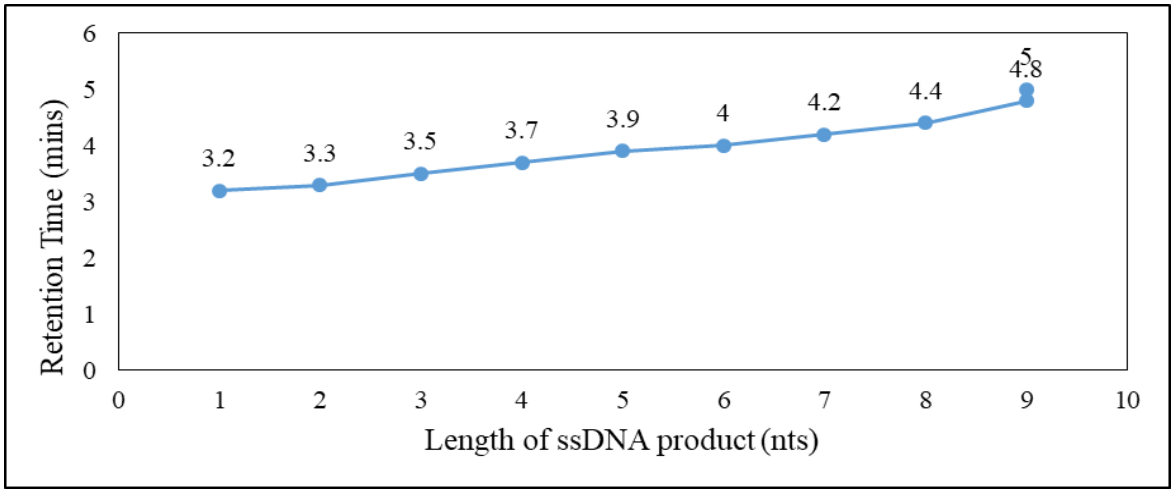

Figure 3.10: *The retention times of the 5'-flap products of the different equilibrating substrates. Analysis by dHPLC show the retention time increases when the length of nucleotides in the 5' flap single stranded increased.*

3.2.4 Studies of the second order rate constant, k_{cat}/K_M for the hFEN1-catalysed reaction of static and equilibrating double flap substrates

These experiments were carried out and initiated as described in section 3.2.3. The $v_0/[E]_0$ min^{-1} versus substrate concentration [S] nM for 10 substrates which have zero to 8 nucleotide 5' flaps in cleaved conformation and increasing numbers of potential conformers were plotted in **Figure 3.11**. The derived second order rate constants were summarized in **Table 3.3**. The graph in **Figure 3.12** plotted k_{cat}/K_M versus the number of potential conformers. **Figure 3.12** shows that, the cleavage by human FEN1 on the migrating flaps occurs at a different rate depending on the amount of potential conformers. Generally, the average k_{cat}/K_M for the 10 substrates were decreased when amount of potential conformers increased reflecting our initial hypothesis.

Table 3.3: *The k_{cat}/K_M for the 10 DNA substrates in these studies. The various length of the 5' flap formed with a 1 nucleotide 3' flap is noted for all DNA substrates. All analysis were made in triplicate with the mean value and standard error are reported.*

DNA Substrates	**Potential Conformers**	**Flap**		**k_{cat}/K_M (nM^{-1} min^{-1})**
		5'	**3'**	
UP31-9ST	1	8	1	6.28 ± 0.59
UP23-9EQ	2	0	1	6.22 ± 0.61
UP24-9EQ	3	1	1	3.70 ± 0.44
UP25-9EQ	4	2	1	0.45 ± 0.19
UP26-9EQ	5	3	1	4.79 ± 0.77
UP27-9EQ	6	4	1	0.58 ± 0.15
UP28-9EQ	7	5	1	2.85 ± 0.47
UP29-9EQ	8	6	1	1.45 ± 0.31
UP30-9EQ	9	7	1	2.25 ± 0.27
UP31-9EQ	10	8	1	0.90 ± 0.11

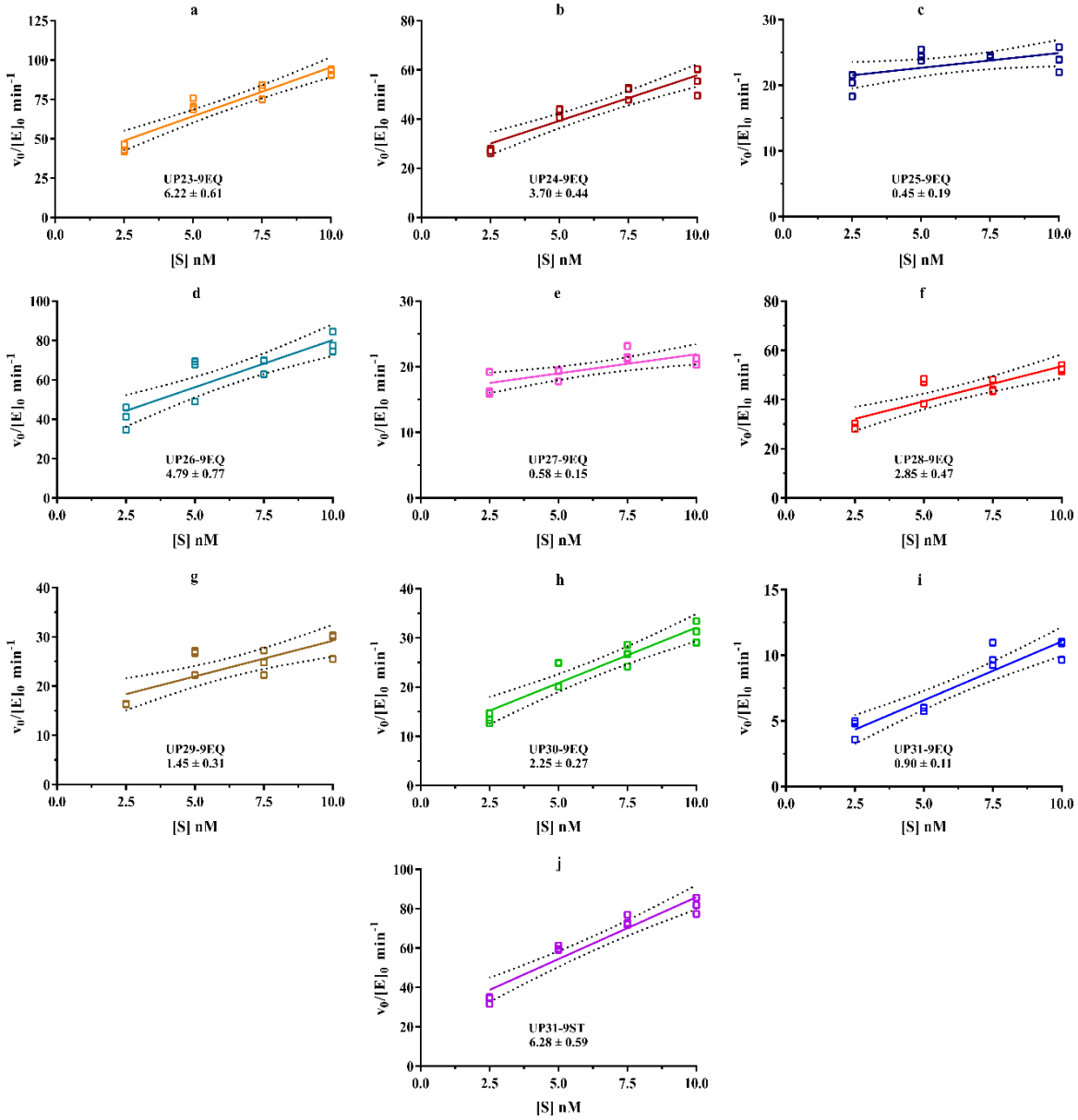

Figure 3.11: *Second-order, k_{cat}/K_M rates of the hFEN1 reaction were determined by plotting by $v_0/[E]_0$ $(min^-)^1$ versus substrate concentration (nM) with the 10 different substrates with differing number of conformers. Each of the figure represents, UP23-9EQ, UP24-9EQ, UP25-9EQ, UP26-9EQ, UP27-9EQ, UP28-9EQ, UP29-9EQ, UP30-9EQ, UP31-9EQ and UP31-9ST.*

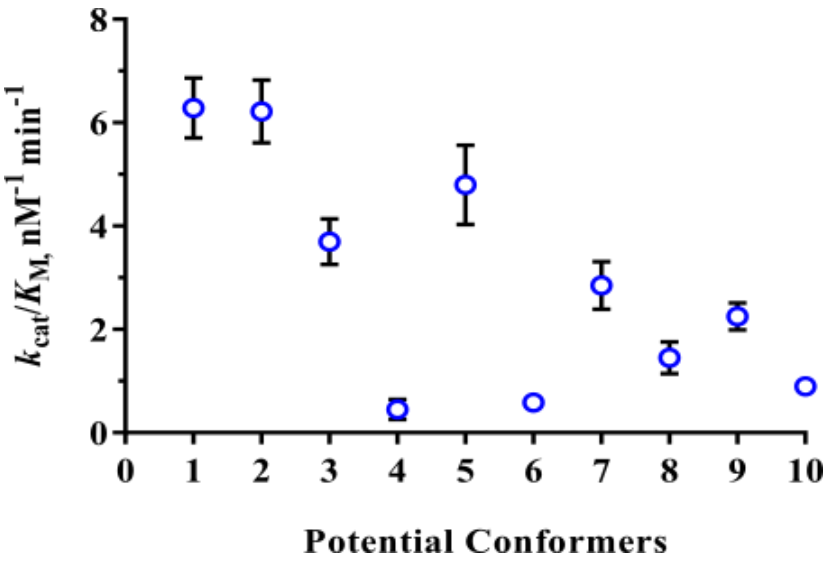

Figure 3.12: *The second order constant; $k_{cat}/K_{M,}$ nM^{-1} min^{-1} versus the number of potential conformers were plotted. Potential conformers 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10 represent of UP31-9ST (static control), UP23-9EQ, UP24-9EQ, UP25-9EQ, UP26-9EQ, UP27-9EQ, UP28-9EQ, UP29-9EQ, UP30-9EQ and UP31-9EQ respectively. The data points (open blue circle) are the average of three replicates and standard errors are shown.*

A linear regression (straight line) and no constrain ('y' not constant equal to zero) was has been used to fit all the data by *GraphPad Prism 7*. Based on the results, three factors probably contribute to the outcome. First, theoretically, assuming the enzyme acts on substrate in one conformational state, the more potential conformers of the migrating flaps substrates the slower the expected k_{cat}/K_M rate constants to form the product. This is because the equilibrating double flap substrates has to undergo a conformational change to form enzyme-substrate complex via interchangeable lengths of 3' and 5' flaps in solution or on enzyme to achieve a 1 nt 3' flap for hydrolysis. For example, **Figure 3.13** demonstrates the 10 potential conformers of UP31-9EQ and since only one conformer reacts its reaction should be the slowest rate when compared to the other equilibrating substrates. However, UP25-9EQ, where only 1 of 4 potential conformers reacts, and UP27-9EQ where only 1 of 6 potential conformers reacts have the slowest rates for second order analysis of 0.45 ± 0.19 nM^{-1} min^{-1} and 0.58 ± 0.15 nM^{-1} min^{-1} respectively. **Figure 3.14** display the static double flaps (UP31-9ST) which has the same length of single stranded 5'-flap as the migrating flap UP31-9EQ but presumably is not undergoing conformational change to make the 3'-flap form. Based on the data, the conformational changes for migrating flaps substrates before the incision of hFEN contribute to the rate of reaction.

Generally, the more possibilities of conformational change of the substrate, the slower the reaction.

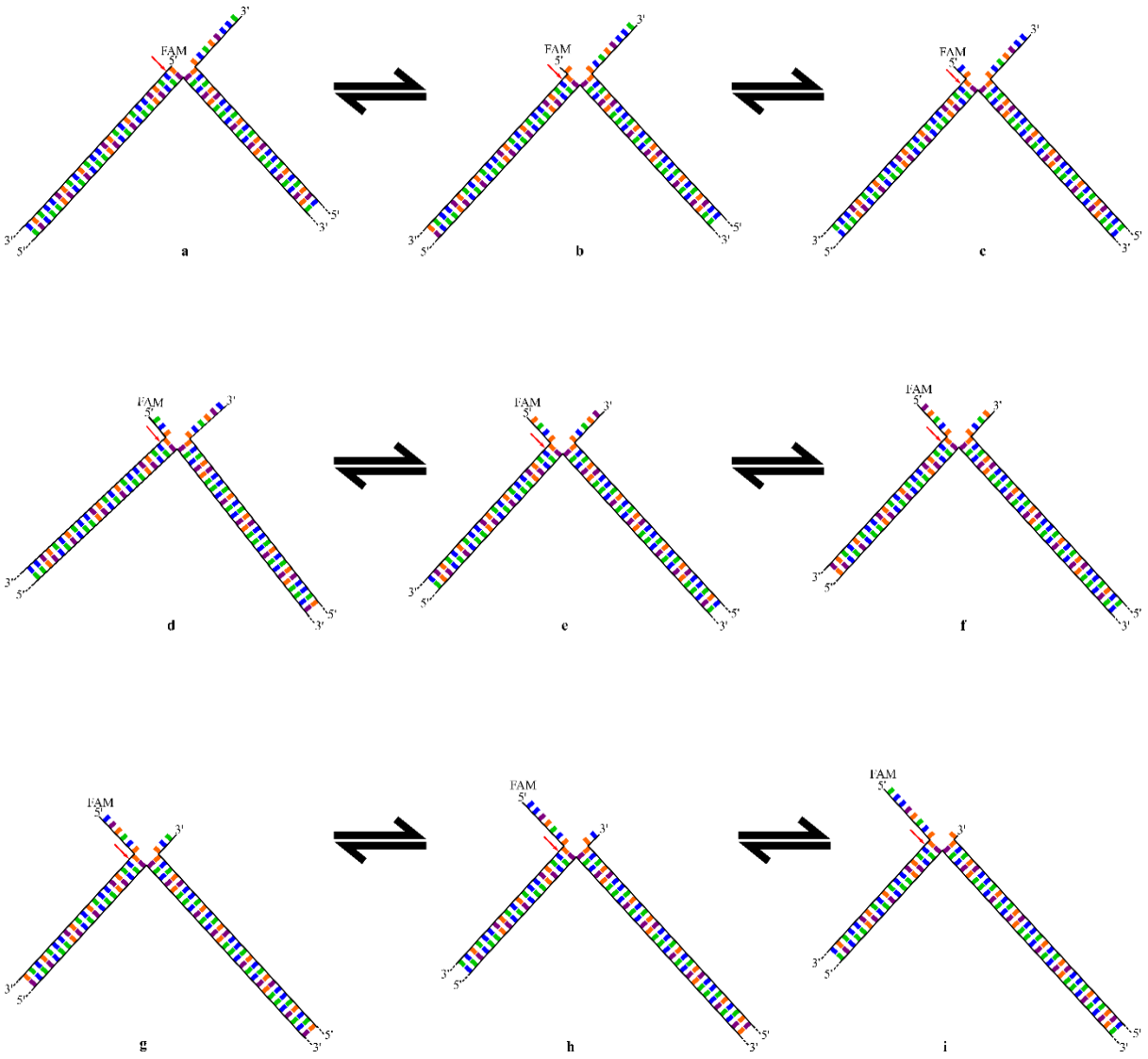

Figure 3.13: *Scheme illustrating 10 potential conformers of UP31-9EQ. The diagram shows 10 potential conformers for the equilibrating flap. FEN1 catalysed incision occurs only when there is a 1 nucleotide at 3' flap. Other potential conformers indirectly contribute to the rate of reaction by reducing the amount of "reactive" substrate. As well as the lengths of the 5'- and 3'-flaps the length of downstream and upstream duplex are changing for each conformer depending on the number of complementary base pairs.*

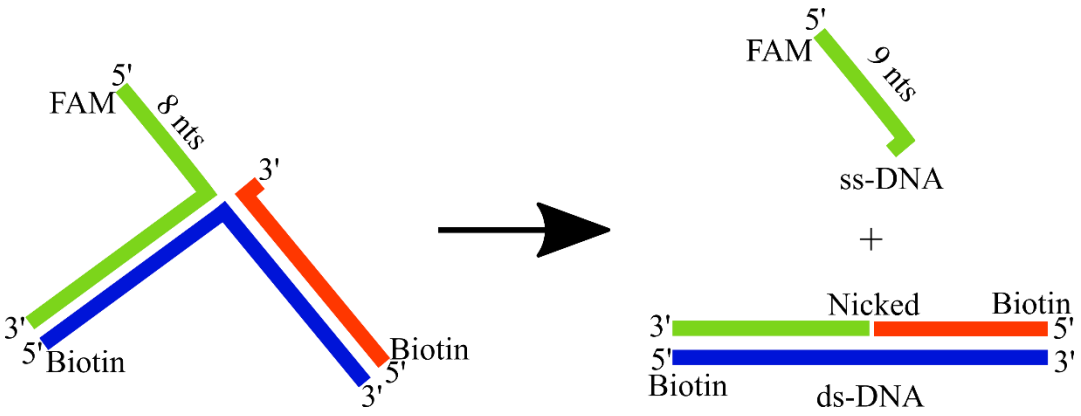

Figure 3.14: *The model of static double flap substrate in-vitro. The direct incision contribute to the fast reaction by human FEN1.*

The second conclusion form this study is that the rate of reaction varies with the length of the ssDNA product produced, which is in turn related to the number of conformers that can be accessed by these substrates. The longer the length of nucleotides on 5' flaps, the slower the k_{cat}/K_M rate constant. **Table 3.3** shows the $k_{cat}/K_{M,}$ for UP31-9EQ (where one of 1 of 10 potential conformers reacts) which has a 9 nucleotide single stranded 5' flap product is 0.90 ±0.11 nM^{-1} min^{-1} compared with UP23-9EQ (where 1 of 2 potential conformers reacts) which was 6 fold higher (6.22 ± 0.61 nM^{-1} min^{-1}) and only has a 1 nucleotide single stranded 5' flap product. This pattern applies generally for all the migrating substrates in this study.

Moreover, there are also additional more subtle effects. Because of the substrate design, which economises on the number of DNAs required, the identity of the base pairs surrounding the scissile phosphodiester bond differ as the number of potential conformers are adjusted. Notably, the nucleotides surrounding the position of the incision (TA or GC), may also have an effect on the rate of reaction. The pattern of second order constants ($k_{cat}/K_{M,}$) (**Figure 3.12**) vary where those with odd numbers of potential conformers have higher rates of reaction compared with those with even numbers of potential conformers. This is probably because of the sequence of the DNA substrates, whereas for 3, 5, 7 and 9 potential conformers the position of the hydrolysis of human FEN1 were on G-C base pairs, while for 2, 4, 6, 8 and 10 potential conformers the phosphate bond were cut off on the T-A base pair.

Alternatively, all the data could be plot with a nonlinear regression Michaelis-Menten ('y' constant equal to zero) by *GraphPad Prism 7*. **Figure 3.15** shows the results of 10 potential conformers via Michaelis-Menten. However, using this method, there are some limitation with the results including a big error for each data.

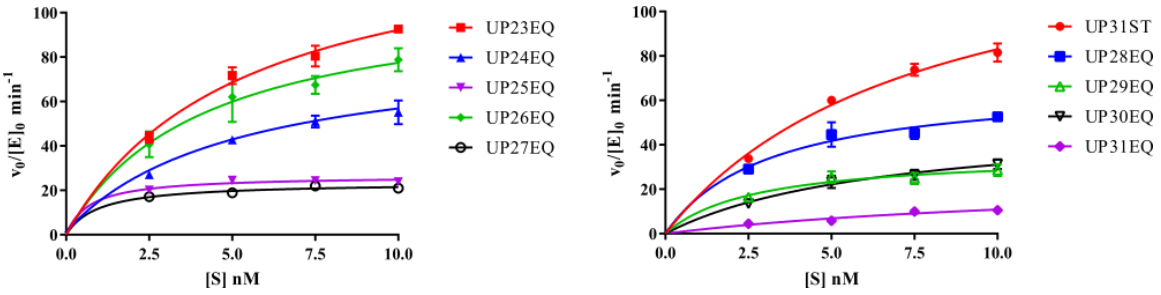

Figure 3.15: *Second-order, k_{cat}/K_M rates of the hFEN1 reaction were determined by plotting by $v_0/[E]_0$ $(min^-)^1$ versus substrate concentration (nM) with the 10 different substrates with differing number of conformers via Michaelis-Menten.*

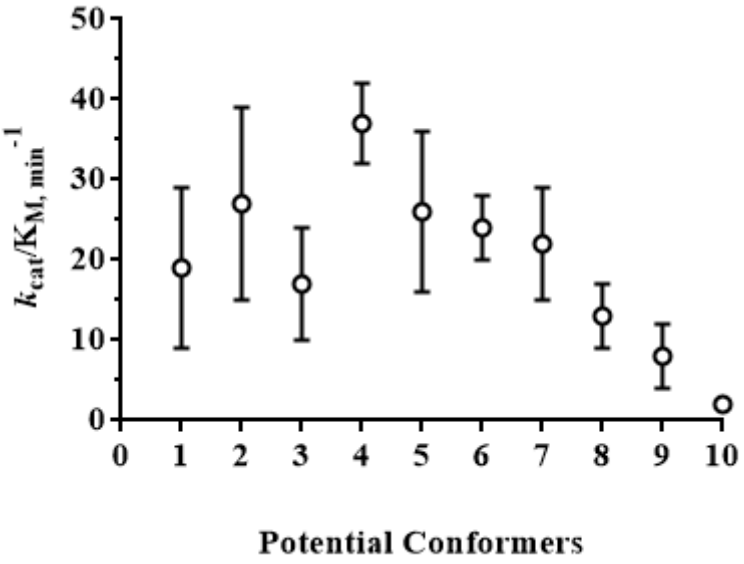

Figure 3.16: *The second order constant; $k_{cat}/K_{M,}$ nM^{-1} min^{-1} versus the number of potential conformers were plotted by Michaelis-Menten. Potential conformers 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10 represent of UP31-9ST (static control), UP23-9EQ, UP24-9EQ, UP25-9EQ, UP26-9EQ, UP27-9EQ, UP28-9EQ, UP29-9EQ, UP30-9EQ and UP31-9EQ respectively. The data points (open black circle) are the average of three replicates and standard errors are shown.*

3.2.5 The effect of lengths of 5’ flaps and the number of potential conformers on static and migrating DNA flap substrates

As our substrate design altered the length of the 5’ flap in the substrates, we decided to investigate the impact of this on the rate of reaction in the range of flap length studies. To investigate this static flaps substrates were used as a control. The equilibrating and static flaps substrates were studied; UP23-9EQ and F18-1ST (F18 = flap strand of 18 nucleotides , 1ST = 1 nt of ssDNA 3’ flap, ST = static) which both have a zero nucleotide 5’ flap in their cleaved conformation, UP27-9EQ and F22-5ST (F = flap strand of 22 nucleotides, 5ST = 5 nts of ssDNA (4 nts 5’flap and 1 nt 3’ flap), ST = static) which both have 4 nucleotide 5’ flaps and UP31-9EQ and UP31-9ST which both have 8 nucleotides 5’ flaps (**Figure 3.17**). 5’-FAM labelled downstream (F26-9ST), and a 5-biotinylated DNA as complementary template (Temp-48) were heating and annealed as mentioned before with a variable upstream 3’-flap strand with 5’ biotin (either F18-1ST or F22-5ST or UP31-9ST individually) to preform a static double flap substrate which has zero nt for F18-1ST or 4 nts for F22-5ST and 8 nucleotides for UP31-9ST 5'-flap length.

Comparison of the catalytic parameters for endonucleoytic cleavage of the three static double flaps, F18-1ST, F22-5ST and UP31-9ST by human FEN1 show that at low[S] here is no statistical difference in the turnover rate (**Table 3.4**), 6.02 ± 0.45 nM^{-1} min^{-1}, 6.40 ± 0.32 nM^{-1} min^{-1} and 6.28 ± 0.59 nM^{-1} min^{-1} respectively, although each of the substrates has different length of nucleotides in its 5’ flap. The k_{cat}/K_M rate constant for UP23-9EQ where 1 of 2 potential conformers reacts is 6.22 ± 0.61 nM^{-1}, close to that of the static flap. However, for UP27-9EQ (1 of 6 potential conformers reacts) and UP31-9EQ (1 of 10 potential conformers reacts), the values of k_{cat}/K_M are 0.58 ± 0.15 nM^{-1} min^{-1} and 0.90 ± 0.11 nM^{-1} min^{-1} respectively, are slower than the k_{cat}/K_M rate constant of their static counterparts (control) with the same length of 5’ flap. This almost certainly arises because both these substrates can adopt conformations that do not present the 1 nucleotide 3’ flap that is preferred by human FEN1. These results signify that the amount of potential conformers influenced to the rate of reaction rather than the 5’-flap length.

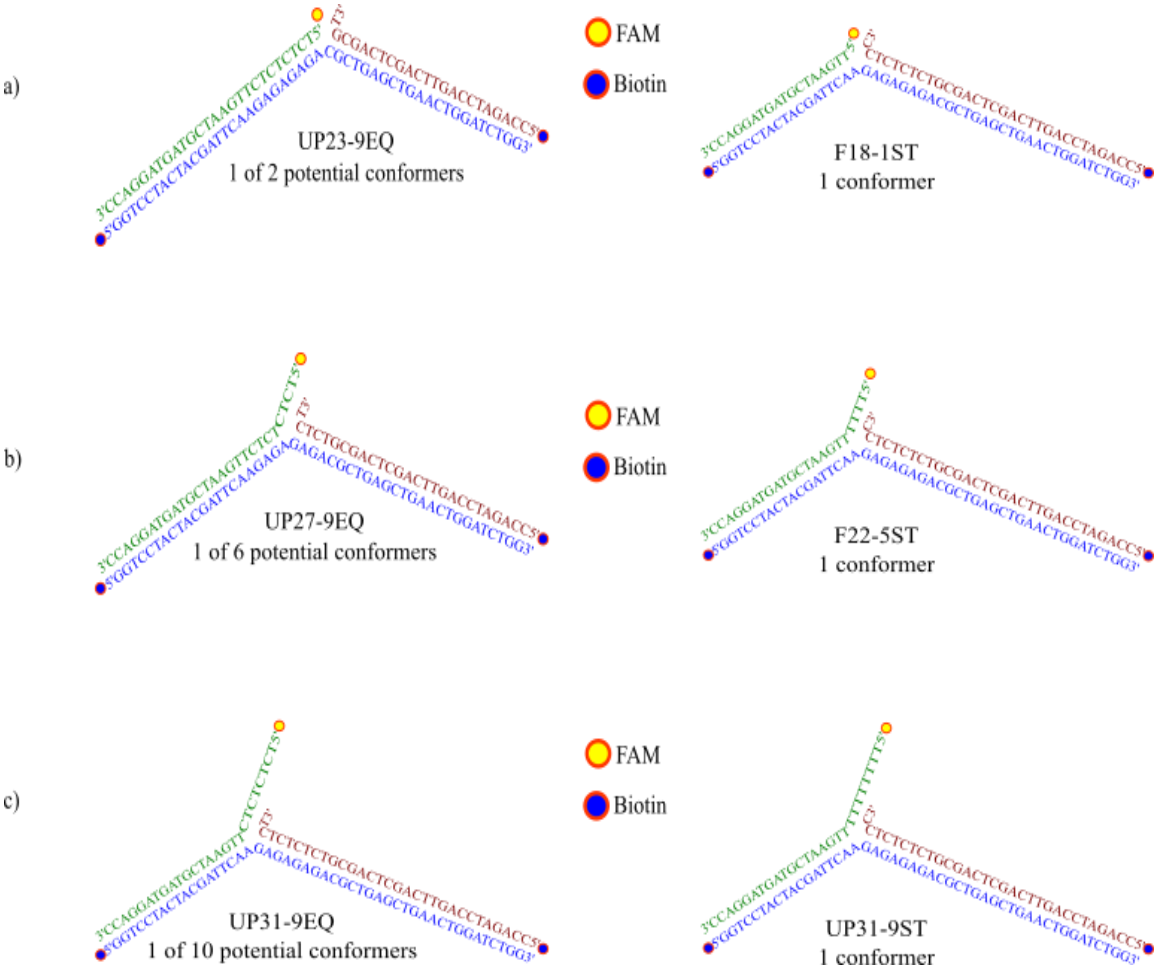

***Figure 3.17**: The sequence of equilibrating (EQ) and static (ST) substrates comprised of 48 nts of template and 26 nts of upstream DNA. Each of the static substrates has 1 conformer, while for the migrating flaps; a) UP23-9EQ has 2 potential conformers with zero nucleotide at 5'-flap in reactive conformation, b) UP27-9EQ has 6 potential conformers with 4 nucleotides of 5'-flap in reactive conformation and c) UP31-9EQ has 10 potential conformers with 8 nucleotides of 5'-flap in reactive conformation.*

***Table 3.4:** The k_{cat}/K_M comparison data between equilibrating and static flaps. Each of the static double flap substrates has one conformer and various length of nucleotides at 5' flap. All measurements were made in triplicate to acquire k_{cat}/K_M rate constants.*

DNA Substrates	Potential Conformers	Flap		k_{cat}/K_M (nM^{-1} min^{-1})
		5'	3'	
UP23-9EQ	2	0	1	6.22 ± 0.61
F18-1ST	1	0	1	6.02 ± 0.45
UP27-9EQ	6	4	1	0.58 ± 0.15
F22-5ST	1	4	1	6.40 ± 0.32

UP31-9EQ	10	8	1	0.90 ± 0.11
UP31-9ST	1	8	1	6.28 ± 0.59

3.2.6 Estimation of k_{cat} for ten different DNA substrates

The steady-state catalytic parameter (k_{cat}) is the measurement of the rate by which an enzyme-substrate complexes [ES] produces the nicked (N) product and free enzyme [E] and is also known as the turnover of the enzyme. The k_{cat} represents the rate of the chemical reaction and the rate of product release and may also include the rate of on-enzyme conformational change steps. Therefore k_{cat} was determined under substrate-saturating multiple turnover conditions to determine what is rate limiting. The analyses have been done for all 10 of the static and migrating double flap substrates at the final concentration of the substrates of 1000 and 3000 nM. For each of the substrate, at least three independent analyses were conducted.

Table 3.5 display that the slope of a plot of $v_0/[E]_0$ against [S] for all the double flap substrates. These are close to zero at high substrate concentration regardless of the numbers of potential conformers and the length of nucleotides at 5' flap. These results indicate that the multiple turnover reaction is saturated with substrate, therefore the rate constants are similar at both concentrations of substrate and are an estimate of k_{cat}. When the average of k_{cat} estimate is plotted against number of conformers (**Figure 3.18**) UP31-9ST is 208 min^{-1} and UP23-9EQ is 228 min^{-1} whereas the other substrates are slower. Generally the data show the substrates with odd numbers of potential conformers have slightly higher average k_{cat} estimates compared with even numbers. Based on these k_{cat} results, 1000 nM substrate concentration is likely well above the K_M for all substrates.

Table 3.5: *The k_{cat} average estimate for 10 DNA substrates in this studies. 1 of 2 until 1 of 10 potential of conformers in all DNA substrates. All analysis were made triplicates at 37^0C.*

DNA Substrates	Potential Conformers	Flap		Slope (nM^{-1} min^{-1})	$v_0/[E]_0$ min^{-1} (Average of k_{cat} estimate)
		5'	3'		
UP31-9ST	1	8	1	0.042 ± 0.006	208.49
UP23-9EQ	2	0	1	0.030 ± 0.005	227.84
UP24-9EQ	3	1	1	2.5×10^{-4} ± 0.001	65.05
UP25-9EQ	4	2	1	2.5×10^{-5} ± 0.001	27.22
UP26-9EQ	5	3	1	-0.004 ± 0.001	63.11
UP27-9EQ	6	4	1	-0.001 ± 0.000	25.37
UP28-9EQ	7	5	1	0.005 ± 0.002	91.74
UP29-9EQ	8	6	1	-0.002 ± 0.001	61.85
UP30-9EQ	9	7	1	-0.006 ± 0.002	46.20
UP31-9EQ	10	8	1	-0.002 ± 0.001	13.01

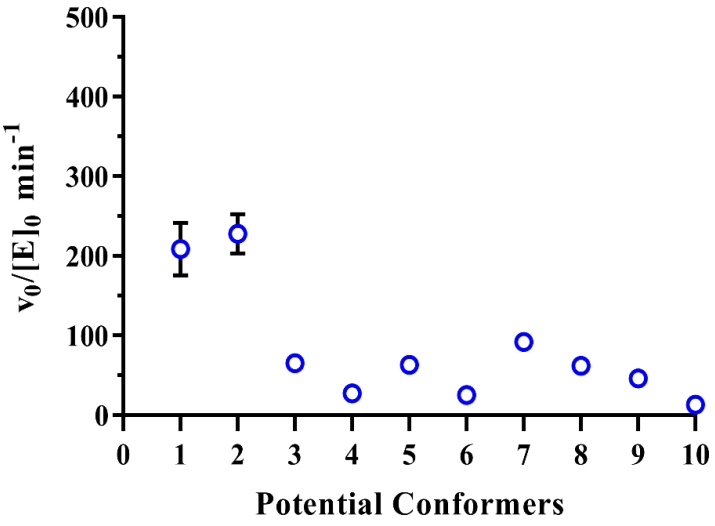

Figure 3.18: *The multiple turnover analysis the average of k_{cat} estimate min^{-1} versus the number of potential conformers were plotted by GraphPad Prism 6 software. Potential conformers 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10 represent of UP31-9ST (static control), UP23-9EQ, UP24-9EQ, UP25-9EQ, UP26-9EQ, UP27-9EQ, UP28-9EQ, UP29-9EQ, UP30-9EQ and UP31-9EQ respectively. The data points (open red circle) are the average of three replicates and standard errors are shown.*

3.2.7 The investigation of single turnover rate, k_{st} on static and migrating flaps DNA substrates

Single turnover rates were measured to complement the study of steady state parameters of the 10 substrates and also to explore whether product release affected the rate of reaction or not. The single turnover rate was determined at higher concentration of enzyme where [E] >> [S]. With these conditions, it is workable to measure the rate of decay of enzyme-substrate complex [ES] before nicked product (N) release. Previous FEN studies observed that the rates of hFEN1-catalysed reactions performed under maximal single turnover (k_{st}) analysis were higher compared with maximal multiple turnover (k_{cat}) conditions. During replication, the diffusion between enzyme and substrate to form enzyme-substrate complex is probably a small factor. As a consequences, it has been proposed that the single turnover analysis *in-vitro* is more relevant to the *in-vivo* situation.

All 10 double flap DNA substrates were studied at 37 ^{0}C using rapid handling apparatus (quenched flow). Under single turnover conditions, it was assumed the reaction started with enzyme-substrate complex when enzyme and substrate were instantaneously mixed. Previous studies have measured identical single turnover rates of reaction under maximal rate conditions for a pre-mixed ES complex to those reactions initiated by mixing E and S[13]. In all experiments for the 10 DNA substrates, the final concentration of hFEN1 and each substrate were 1000 nM and 2.5 nM respectively. Reactions were repeated with at least three replicates to ensure reliability and data was collected in the range of 4.5 ms to 51241 ms in 55 mM HEPES (pH 7.5), 8 mM $MgCl_2$, 1 mM DTT, 110 mM KCl and 0.1 µg/µl BSA. The percent of product versus time were plotted as shown in **Figure 3.19.** The data were fitted with **Equation 2.2** in section 2.3.2 by non-linear regression fitting (*GraphPad Prism 7 software*). The plotted data in **Figure 3.19** show each of the DNA substrates display a two phase exponential relationship. The preferable fit to all the data was two phase exponential in all cases with 99% probability.

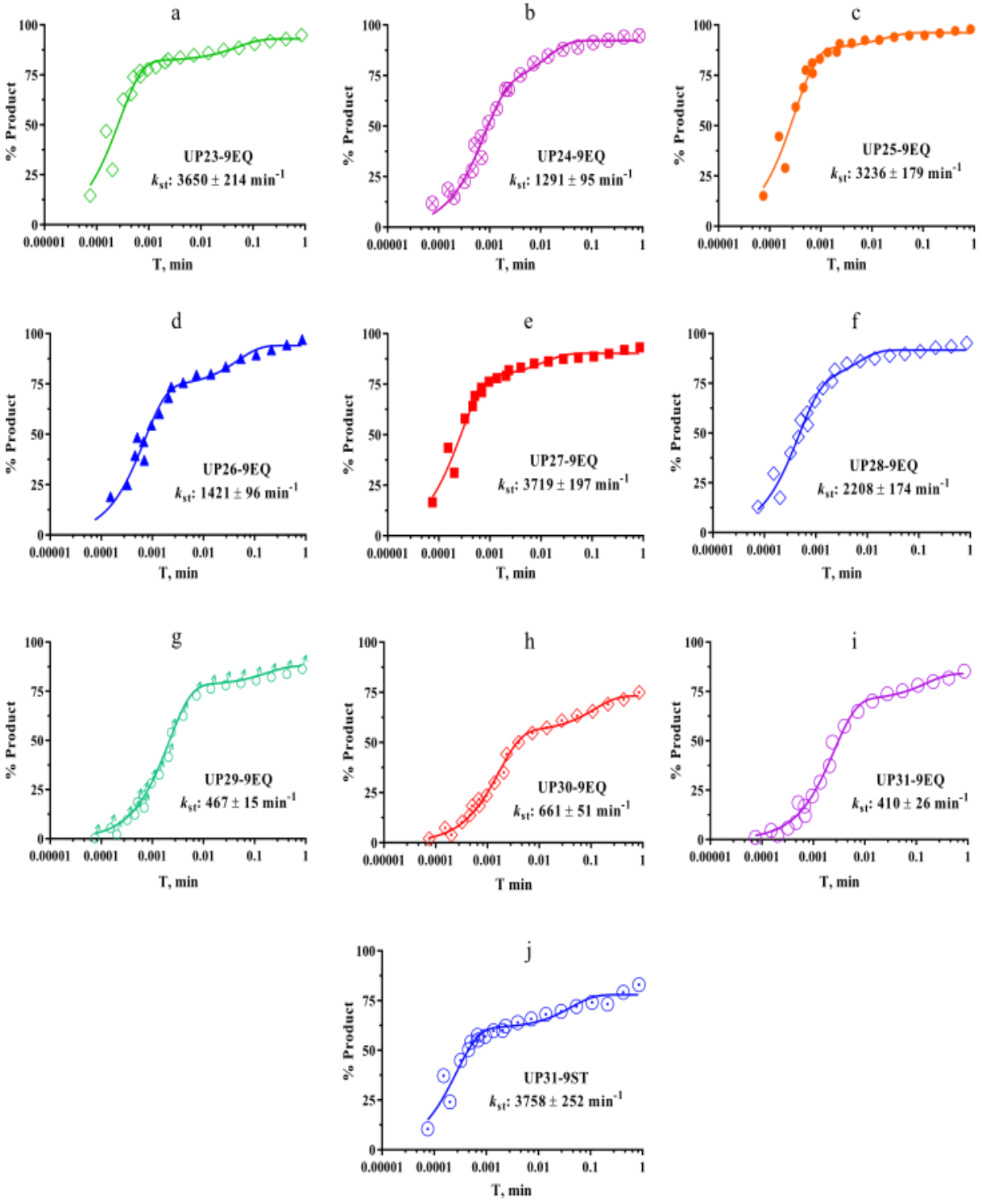

Figure 3.19: *First-order rate data were plot by percentage of the product (%) versus time (minute) with 10 different DNA substrates. Each of the figures represents; a for UP23-9EQ, b for UP24-9EQ, c for UP25-9EQ, d for UP26-9EQ, e for UP27-9EQ, f for UP28-9EQ, g for UP29-9EQ, h for UP30-9EQ, i for UP31-9EQ and j for UP31-9ST. The aliquots were quenched by RQF-63 quench flow device in 1.5 M NaOH and 80 mM EDTA and analysed by dHPLC at 75 ^{0}C.*

In each case data was fit to a double exponential with the rate of the fast phase of the reaction quoted. As shown in **Figure 3.20** and **Table 3.6** the k_{fast} rate for static double flaps (UP31-9ST), 3758 ± 252 min^{-1} is the fastest rate to produce a 9 nucleotide 5' flap product compare with other migrating flaps substrates. However, UP23-9EQ and UP27-9EQ have similar rates of single turnover to the static substrate of 3650 ± 214 min^{-1} and 3719 ± 197 min^{-1} respectively and UP25-9EQ shows a slightly slower k_{st} rate 3236 ± 179 min^{-1}. For UP28-9EQ , UP26-9EQ, UP24-9EQ DNA substrates the single turnover rate are 1.7-fold (2208 ± 174 min^{-1}), 2.6-fold (1421 ± 96 min^{-1}) and 2.9-fold (1291 ± 95 min^{-1}) respectively slower than the UP31-9ST substrate. Another three migrating flaps have the slowest k_{fast}, of 467 ± 15 min^{-1} for UP29-9EQ, 598 ± 24 min^{-1} for UP30-9EQ and 395 ± 17 min^{-1} for UP31-9EQ. Generally, the average of single turnover rate, k_{st} from each DNA static and migrating substrates were faster than that measured under multiple turnover conditions. Moreover, the general trend is that as the number of conformers increases the single turnover rate decreases.

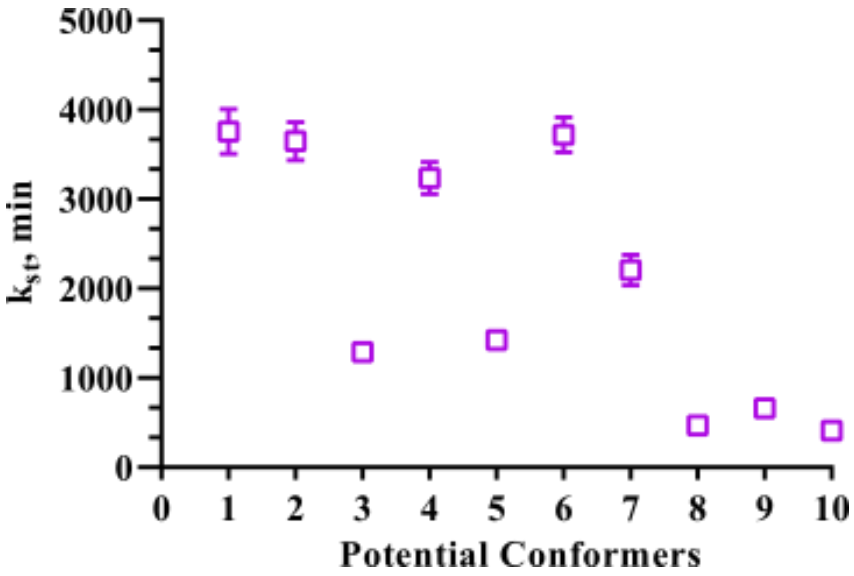

Figure 3.20: *The first order constant; k_{st} min^{-1} versus the number of conformers were plotted. Potential conformers 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10 represent of UP31-9ST (static control), UP23-9EQ, UP24-9EQ, UP25-9EQ, UP26-9EQ, UP27-9EQ, UP28-9EQ, UP29-9EQ, UP30-9EQ and UP31-9EQ respectively. The data points (open purple square) are the average of three replicates and standard errors are shown.*

It is important to note that in all cases the amount of species decaying with a fast rate constant is relatively high (> 74%) and there is no relationship between the percentage fast and the number of conformers with stochastic variation observed between 74-92%. Decreasing proportions of fast decaying species would be expected if increasing number of conformers creates increasing proportions of "inactive" substrate that took significant time periods to rearrange, but this does not seem to be the case. Instead the formation of non-productive [ES] complex at high enzyme concentration, which seems to occur in all cases, could contribute to why the reaction failed to all react in the fast phase. It is likely that non-productive complexes need to dissociate and reassemble in before completing reaction.

The increasing the number of potential conformers in each substrate appears to slow the single turnover rate of reaction. UP29-9EQ, UP30-9EQ and UP31-9EQ are three migrating substrates where 1 of 8, 1 of 9 and 1 of 10 potential conformers respectively, and produce the slowest reactions with 1000 nM enzyme. In one model, where the enzyme can only bind substrate in "reactive" (1 nt 3' flap) conformation, increasing the number of conformers would raise the enzyme-substrate dissociation constant. Therefore a control was carried out with UP30-9EQ (1 of 9 potential conformers reacts) at a higher concentration of enzyme (3000 nM). The k_{st} was 497 ± 42 min^{-1} which is similar to the rate where [E] = 1000 nM (598 ± 24 min^{-1}). This result demonstrated that the rate constants all seem likely to represent the maximal k_{st} value. Thus the maximal single turnover rate is decreased with increasing number of conformers. This is likely to be because the enzyme and substrate (ES) complex needs more time (multiple conformations as discussed before) to form the productive ES complex with the substrate in the correct configuration.

Table 3.6: *The k_{st} for 10 DNA substrates in this studies. Various length of nucleotides 5' flap and 1 nucleotide 3' flaps in all DNA substrates. All analysis were made triplicates and measurements by dHPLC at 75 ^{0}C and plotted by GraphPad Prism 7 software. Errors represent standard errors of experiments performed in at least triplicate.*

DNA Substrate	Potential Conformers	k_{fast} min^{-1}	Percent Fast (A_1)	k_{slow} min^{-1}	Plateau
UP31-9ST	1	3758 ± 252	78.33 ± 1.87	23.11 ± 7.60	77.90 ± 1.36
UP23-9EQ	2	3650 ± 214	88.18 ± 2.13	19.56 ± 12.43	93.01 ± 1.84
UP24-9EQ	3	1291 ± 95	74.87 ± 3.37	76.07 ± 23.04	92.33 ± 1.06
UP25-9EQ	4	3236 ± 179	91.75 ± 2.07	58.74 ± 48.24	96.15 ± 1.28
UP26-9EQ	5	1421 ± 96	78.48 ± 2.43	21.22 ± 7.92	94.06 ± 1.73
UP27-9EQ	6	3719 ± 197	86.72 ± 1.83	88.38 ± 38.64	90.22 ± 0.96
UP28-9EQ	7	2208 ± 174	80.04 ± 4.02	165.30 ± 67.83	91.72 ± 0.97
UP29-9EQ	8	467 ± 15	88.73 ± 1.61	5.66 ± 2.62	87.98 ± 1.41
UP30-9EQ	9	598 ± 24	75.50 ± 1.52	8.50 ± 1.78	73.23 ± 1.01
UP31-9EQ	10	395 ± 17	83.86 ± 2.00	6.40 ± 17.38	84.11 ± 1.57

The results show a very fast reactions (k_{st} above 3000 min^{-1}) for the double flap substrates; UP31-9ST, UP23-9EQ, UP25-9EQ and UP27-9EQ. This suggests, there are small differences in the stability of enzyme-substrates complexes. Most of the even numbers of potential conformers, UP23-9EQ, UP25-9EQ, UP27-9EQ, UP29-9EQ and UP31-9EQ have percent fast 88 ± 2, 92 ± 2, 87 ± 2, 89 ± 2 and 84 ± 2 respectively which are higher than the odd numbers of potential conformers and also share similar results for the rate of the single turnover k_{st}. Curiously, this situation is the opposite of that with multiple turnover analysis.

3.3 Discussion and summary

During replication the bifurcated structures that are the result of discontinuous DNA synthesis on the lagging strand are likely to be able to adopt multiple structures due to sequence complementarity. In contrast, *in-vitro* studies of FEN1 frequently use the 3' and 5' static flap substrates, wherein the flaps are not complementary to the template strand, to remove substrate conformational variability that would complicate interpretation of rates of reaction. However, a complete understanding of the human FEN1 mechanism requires the study of the effect of flap migration on rates of hFEN1 reaction. Hence, 9 double flap migrating substrates were designed to address the effects of substrate conformational diversity on hFEN1 rates of reaction using multiple and single turnover kinetics.

Multiple turnover analysis showed that human FEN1 produces a single peak 5' flap product even on migrating double flap substrates that have at least 10 potential conformers. This showed that hFEN1 always hydrolyses one nucleotide into the reacting duplex presumably by only reacting rapidly on a single conformer bearing a 3'-single nucleotide flap and a 5'-flap of varying lengths. The k_{cat}/K_M and maximal k_{st} conditions roughly show that the rate of reaction was inversely proportional to the number of potential conformers. Moreover, controls show that the decrease in the rate is not due to change in 5'- flap length. However, the data also showed that the rate is not simply inversely proportional to the number of conformers, as the data points do not fit evenly on a line; instead, some substrates show lower than expected rates.

Based on the decreases in rate of reaction observed with increasing conformational heterogeneity, two models could be proposed. First, the enzyme could only bind the correct conformer. This model predicts that the value of k_{cat}/K_M would decrease as the K_M value increases as there would be increasing amounts of substrate not in the correct conformation at any given time. This model predicts that the rate of the fast phase of $k_{st\ max}$ remains the same with increasing conformational heterogeneity although its proportion of fast reacting species may vary. It is assumed that the DNA substrate is quickly equilibrating among its many conformers into its preferred form (i.e., single nt 3'-flap) thus allowing reaction to eventually proceed to completion. Alternatively, enzyme could bind with all (or most) conformers and then allow the DNA substrate to adopt the correct conformation on the enzyme surface. This model predicts k_{cat}/K_M would decrease and that the $k_{st\ max}$ value would

decrease with increasing number of conformers and these model is more relevant with all the kinetic results. A combination of these two models with various migrating substrates could explain why a smooth trend is not observed. In addition, it is possible that there are conformational preferences for various migrating flaps that are not well understood.

Although a direct measurement of K_M is not available, its value can be derived using **Equation 3.3** in chapter 1 using the k_{cat} estimate and k_{cat}/K_M values. The calculated K_M values are shown in **Table 3.7**. The estimates of K_M are between 10 and 60 nM for all substrates showing that the K_M values do not change drastically with increasing substrate complexity. Therefore, the conditions for ST are maximal single turnover rates ([E]>10* K_M) and the model where substrate can re-equilibrate conformation on the enzyme seems preferable.

Table 3.7: *The calculation for* K_M *estimate for 10 potential conformers.*

DNA Substrate	**Potential Conformers**	**k_{cat} estimated (min^{-1})**	**k_{cat}/K_M (nM^{-1} min^{-1})**	**K_M estimate (nM)**
UP31-9ST	1	208.49	6.28	33.20
UP23-9EQ	2	227.84	6.22	36.63
UP24-9EQ	3	65.05	3.70	17.58
UP25-9EQ	4	27.22	0.45	60.49
UP26-9EQ	5	63.11	4.79	13.18
UP27-9EQ	6	25.37	0.58	43.74
UP28-9EQ	7	91.74	2.85	32.19
UP29-9EQ	8	61.85	1.45	42.66
UP30-9EQ	9	46.20	2.25	20.53
UP31-9EQ	10	13.01	0.90	14.46

Chapter 4: Proliferating cell nuclear antigen

Proliferating cell nuclear antigen (PCNA) is a homotrimeric sliding clamp that encircles the DNA substrate and tethers the polymerase to the template DNA to facilitate processive DNA synthesis. The presence of an inter-domain connector loop (IDCL) of PCNA, allows it to interact with various proteins including FEN1 and DNA Ligase 1. During DNA replication and repair, polymerase, FEN1 and ligase interact with PCNA. Previous studies proposed that the PCNA-FEN1 interaction stabilized FEN1-substrate complex resulting in stimuation of FEN1-catalysed hydrolysis of phosphodiester bonds during the Okazaki fragment maturation processes[46]. For example, previous studies from Hubscher et *al.*, suggested that PCNA can load onto a DNA 5'-single flap substrate via the upstream duplex using biotin-streptavidin methods[123]. Other studies by the Bambara lab, have shown that PCNA could alter the FEN1 binding by changing the K_m value. These studies also suggested that PCNA could enhance the FEN1 cleavage activity due to a small increase in V_{max}. PCNA has been reported to stimulate the FEN1 incision activities from 5 to 50-fold on a variety of DNA flap substrates including single flap and gap substrates, suggesting that catalytic efficiency is improved[46]. In addition, structural studies show that the FEN1 and PCNA form an inter-molecule β-sheet interface between PCNA and the FEN1 extended C-terminal regions in FEN1[22]. In addition, further studies by FRET have suggested that the PCNA assisted the opening of the flap conformation for 5'-flap threading processes and indirectly enhanced FEN1 recognition of the DNA substrate[124]. In summary earlier studies suggested that PCNA could stimulate the FEN1 cleavage efficiency.

Using double-flap substrates, mild PCNA-mediated stimulation of archaeal FEN1 proteins has been demonstrated[125]. PCNA stimulation of human FEN1 acting on the preferred conformer of its double-flap substrates seems unlikely as the reaction of these approach the rate associated with diffusion controlled reaction anyhow. However, investigation of the interaction between FEN1-PCNA in migrating double flaps are interesting to explore because migrating substrates seem to have dropped out of the diffusion control regime. In this chapter, we investigate the ability of PCNA to stimulate human FEN1 hydrolysis of a static double flap (UP31-9ST), a migrating double flap (UP27-9EQ), and a single flap (SF). The aim of study to investigate the role of PCNA with the variety DNA substrates of varying conformational complexity and efficiencies.

4.1 Proliferating cell nuclear antigen purification

Human PCNA bearing a C-terminal $(His)_6$-tag was purified following expression in *Escherichia coli* BL21(DE3)-RILP cells using auto-induction methods as described in section 2.2. Initially, target protein was captured by IMAC, and the purificaion was continued with anion exchange step to remove nucleic acid contamination. The other chromatograpic steps included heparin affinity, phenyl Sepharose hydroxyapatite and desalting columns to obtain pure protein having no contaminating nuclease activity. Using a Nanoview spectrophotometer, the protein concentration was measured by absorbance at 280 nm. The volume of glycerol equal to the volume of protein solution was added to make a 300 µM PCNA stock solution in 50% glycerol. The final concentration will be 300 µM of hPCNA monomer, but as hPCNA is a trimeric protein, the concentration of $(hPCNA)_3$ is 100 µM. Multiple turnover kinetic analyses were carried out without human FEN1 to confirm that the PCNA was free from nuclease activity. The final concentration of the double flap static substrates were 50 nM and various concentrations of PCNA were added; 0 µM (blank), 0.025 µM, 0.05 µM 1 µM and 10 µM in 8 mM $MgCl_2$, 50 mM HEPES pH 7.5, 0.02 % NaN_3, 100 mM KCl, 0.1 mg/mL BSA at 37 °C. The formation of single peak products (**Figure 4.1**) monitored by reverse-phase ion-pairing denaturing HPLC at 75 ^{0}C proved that the PCNA free from any nuclease activity.

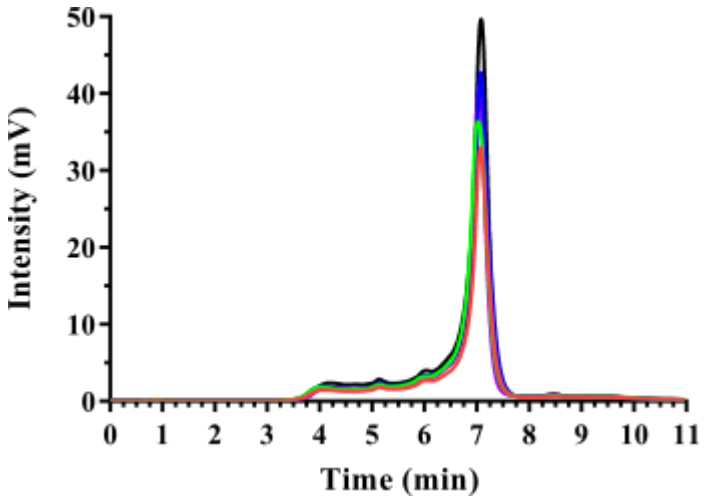

Figure 4.1: *The second order rate for PCNA nucleus activity. A plot of the intensity (mV) of product and substrate chromatogram versus retention time (min) was plotted. Five different concentration of PCNA were analysed with double flaps DNA substrate, UP31-9ST. The red, green, blue, purple and black represents 0, 0.025, 0.05, 1 and 10 µM respectively.*

4.2 PCNA with migrating flaps

The hFEN1-DNA complex structures showed that 12 and 4 base pairs of downstream and upstream duplexes is enough for proper binding between FENs and dsDNA substrate, and substrates used for reaction with FEN1 reflect this. Based on the model from Finger et. *al*., 2012, a substrate capable of being used for PCNA stimulation of FEN1 needed to be designed and was done by Dr. Finger. Although 12 base pairs of downstream duplex is enough for FEN1 interaction, the stability of the 12 base pairs duplex is too low at concentration as low as 1 nM at 37 ^{0}C in assay buffer, thus substrates were designed with a minimum of 18 base pairs for the downstream region[37]. Using a model the FENs-dsDNA-PCNA complex, at least 21 base pairs are needed in the upstream region for FEN1 and PCNA to be able to make all necessary contacts. Therefore, the substrates consists of 18 and 22 base pairs non-migrating downstream and upstream duplex region, respectively, and a nine base pairs region in the middle that migrate to form the variable 5' and 3' flap.

The second order rate of reaction analyses were carried out by heating and annealed the oligonucleotides as mention in chapter 3 to form a migrating double flap substrate, UP27-9EQ. As described before, UP27-9EQ is a double flaps substrate that has 6 potential conformers where the flap structure that undergoes hydrolysis has a 4 nts 5'-flap fluorescein (FAM) labelled, a 1 nt 3'-flap with 5' biotin on the flap and template strand. The PCNA and human FEN1 were added and reactions were quenched with 8 M urea and 80 mM EDTA. The substrate UP27-9EQ (**Figure 4.2**) was chosen because, the multiple turnover analysis from chapter 3 demonstrated that the equilibrating flap UP27-9EQ has the slowest k_{cat}/K_M rate. The impact of PCNA on the rate of reaction was measured using final substrate concentrations of 2.5, 5, 7.5 and 10 nM with 1.5 or 3.5 pM of [WThFEN1] and 50 or 100 nM of $[PCNA]_3$. As before fluorescein-labelled oligonucleotide substrates and reverse-phase dHPLC will be used to monitor DNA cleavage site product and product formation concomitantly.

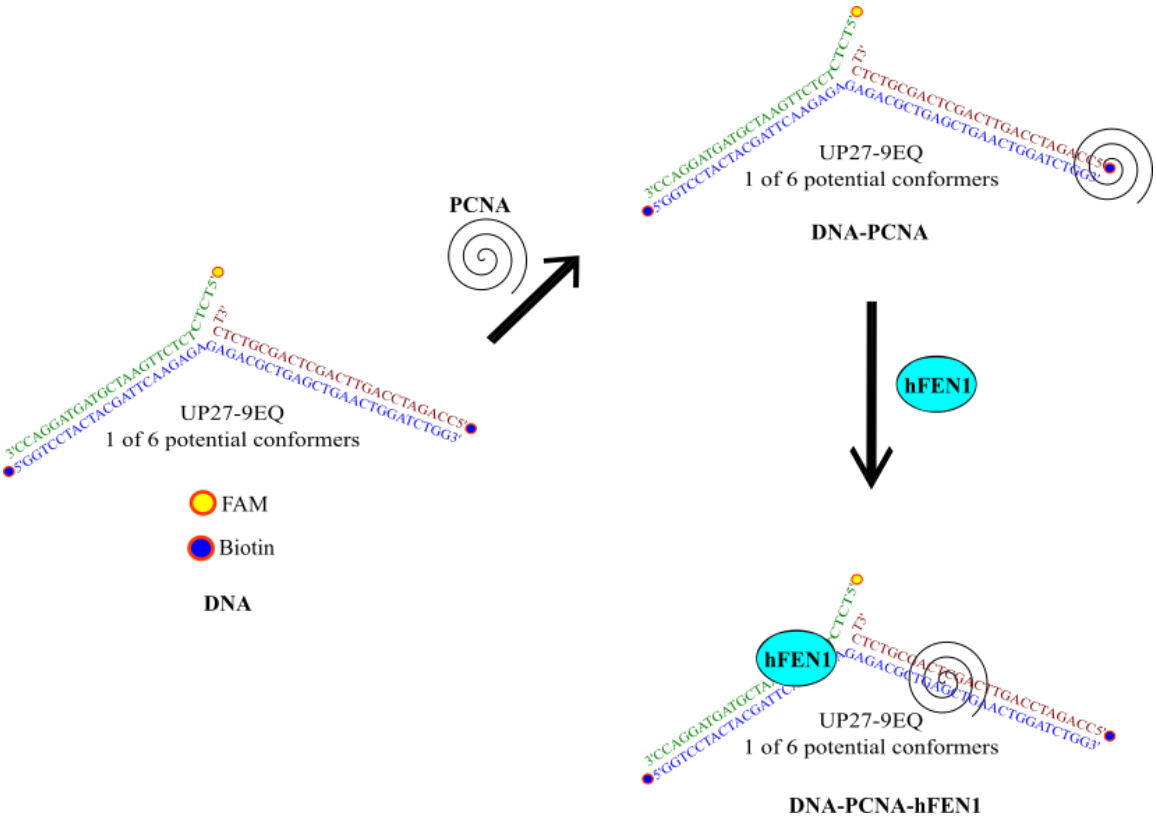

Figure 4.2: *The diagram represents DNA-PCNA-hFEN1 complexes during the experiment with UP27-9EQ substrate.*

Figure 4.3 and **Table 4.1** show the rate of reactions were 0.58 ± 0.15 nM^{-1} min^{-1} for UP27-9EQ, 0.36 ± 0.07 nM^{-1} min^{-1} for UP27-9EQ in the presence of 50 nM $[PCNA]_3$ and 0.32 ± 0.08 nM^{-1} min^{-1} for 100 nM $[PCNA]_3$. It was hypothesised that because UP27-9EQ has 6 potential conformers and therefore reduced rate of reaction stabilising the DNA protein complex with PCNA could assist with formation of productive [ES] complex and therefore increase the k_{cat}/K_M rate. Instead the k_{cat}/K_M rate of FEN1 catalysed reaction of UP27-9EQ without PCNA is roughly two-fold higher than with PCNA. These results suggest that PCNA does not help to stabilize the human FEN1-DNA complex. This is probably because the reaction of the equilibrating double flap substrate approaches diffusion control. Furthermore, the high concentration of PCNA in this study, probably modestly inhibited human FEN1 activity by competing for binding of DNA with FEN1. Other than that, the presence of biotin at two positions; the template and upstream could also contribute to these results by inhibiting loading of the PCNA onto FEN1.

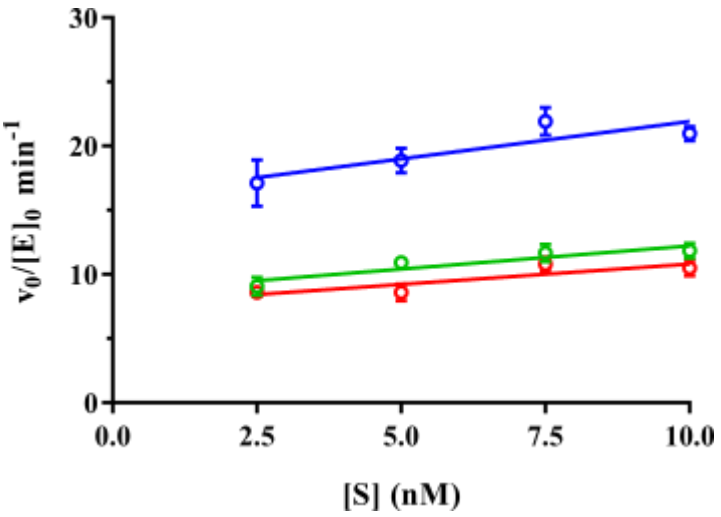

Figure 4.3: *Determination of the second order rate constant (k_{cat}/K_M) from a plot of $v_0/[E]_0$ against [S] for hFEN1 catalysed reaction of UP27-9EQ substrate in the presence of PCNA. The open blue circles represent UP27-9EQ substrate without PCNA, the open green circles represent UP27-9EQ with 50 nM of $[PCNA]_3$ and the open red circle data represent UP27-9EQ with 100 nM of $[PCNA]_3$. The average of three replicates and standard errors are shown.*

Table 4.1: *The* second order *rate analysis of UP27-9EQ with PCNA. The triplicates data were carried out by plotted the normalised rate of reaction $v_0/[E]_0$ versus [S] using GraphPad Prism 7 software to acquire k_{cat}/K_M rate constants.*

DNA Substrate (UP27-9EQ)	k_{cat}/K_M **(nM^{-1} min^{-1})**
Without PCNA	0.58 ± 0.15
50 nM PCNA	0.36 ± 0.07
100 nM PCNA	0.32 ± 0.08

4.3 PCNA with single flap substrates

Single flap substrate was created by heating and annealing F26-9ST downstream, UP30-9EQ upstream and Temp-48 template in 1:1.5:1.5 ratio, respectively in buffer to form a static single flap substrate with an 8 nts 5'-flap strand with 5’-fluorescein (FAM) and a zero nt 3'-flap strand with 5’ biotin and a 5’-biotinylated template (**Figure 4.4).**

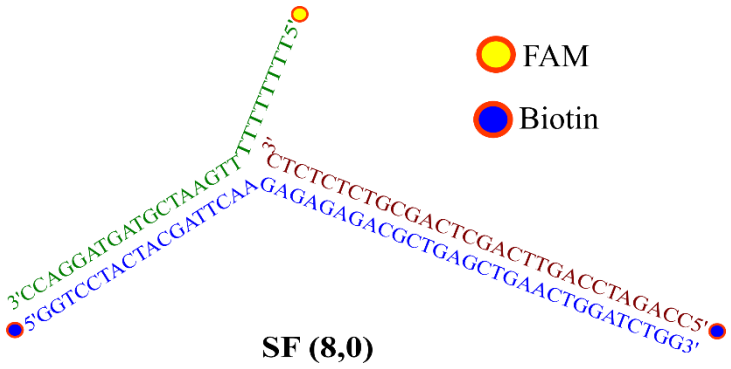

Figure 4.4: *The static single flap substrates were constructed from F26-9ST (green), UP30-9EQ (red) and Temp-48 (blue).*

The stimulation of hFEN1 acting on single flap substrates by PCNA has been reported earlier[46]. We therefore undertook further analyses of the hFEN1 reaction in the presence of PCNA and with single flap (SF) substrate. Reactions were analysed under standard k_{cat}/K_M conditions (2.5, 5, 7.5 and 10 nM SF) with varying ratios of PCNA to substrate including 1[SF]:1$[PCNA]_3$, 1[SF]:3$[PCNA]_3$ (**Table 4.2 and Figure 4.5 (a)**) and enzyme 1[hFEN]:1$[PCNA]_3$ and 1[hFEN1]:3$[PCNA]_3$ (**Table 4.3 and Figure 4.5 (b)**).

Table 4.2: *The multiple conditions with single flap substrate with the different ratio between [SF] and $[PCNA]_3$. The final hFEN1 concentration was 150 pM. The k_{cat}/K_M rate were assessed by dHPLC and each analysis is representative of three replicates.*

[SF] (nM)	**1$[PCNA]_3$ (nM)**	**k_{cat}/K_M (nM^{-1} min^{-1})**	**[SF] (nM)**	**3$[PCNA]_3$ (nM)**	**k_{cat}/K_M (nM^{-1} min^{-1})**
2.5	2.5		2.5	7.5	
5	5	0.035 ± 0.004	5	15	0.043 ± 0.005
7.5	7.5		7.5	22.5	
10	10		10	30	

Table 4.3: *The second order analysis with single flap substrate with the different ratio between [hFEN1] and [PCNA]$_3$ whereas the final [hFEN1] both cases were 150 pM. The k_{cat}/K_M rate were assessed by dHPLC and each analysis is representative of three replicates.*

[SF] (nM)	1[PCNA]$_3$ (pM)	k_{cat}/K_M (nM^{-1} min^{-1})	[SF] (nM)	3[PCNA]$_3$ (nM)	k_{cat}/K_M (nM^{-1} min^{-1})
2.5	150		2.5	450	
5	150	0.018 ± 0.002	5	450	0.032 ± 0.002
7.5	150		7.5	450	
10	150		10	450	

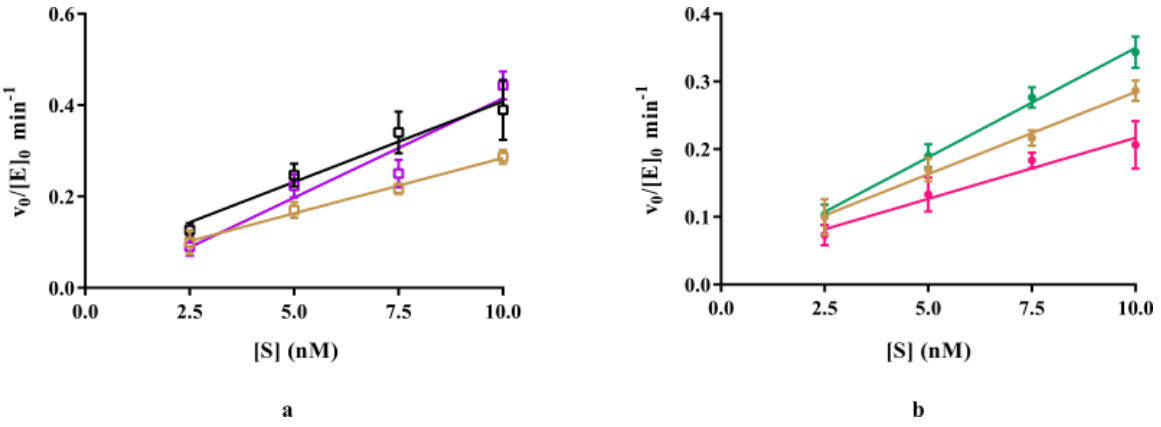

Figure 4.5: *The $v_0/[E]_0$ conditions for single flap substrate with PCNA. a represent the PCNA analysis with different ratio with substrate where, brown for SF without PCNA, black for 1[SF]:1[PCNA]$_3$ and purple for 1[SF]:3[PCNA]$_3$. b represent the PCNA analysis with different ratio with hFEN1 where, brown for SF without PCNA, pink for 1[hFEN]:1[PCNA]$_3$ and green for 1[hFEN1]:3[PCNA]$_3$. The average of three replicates and standard errors are shown.*

The presence of one nucleotide at 3'-flap DNA substrate is required for phosphate diester hydrolysis by FEN1 to produce a nicked DNA for ligation and complete the daughter DNA strand. Therefore, in the absence of 1 nt 3'-flap, the rate of reaction is slow and produces more than one product. In this study, the k_{cat}/K_M rate of SF (8,0) without PCNA was 262-fold slower (0.024 ± 0.002 nM^{-1} min^{-1}) compared with static double flap substrate UP31-9ST (8,1). These results are supported by Finger, et. *al*., 2009 studies where the single flap substrate have 34-fold slower k_{cat}/K_M rate compared with double flap substrates. The different design of substrate between this study and that of Finger et *al.*, probably

contributes to the rate of the reaction. The previous study used a substrate consisting of a 5 nts 5'-flap with the presence of a hairpin turn in the substrate[13].

Figure 4.6 demonstrates that the presence of PCNA had a very small impact on the rate increasing it to 0.035 ± 0.004 nM^{-1} min^{-1} with 1:1 ratio and producing a modest 2-fold increase in the rate of reaction with 1[SF]:3$[PCNA]_3$ ratio (0.043 ± 0.005 nM^{-1} min^{-1}). Similarly, different ratios of hFEN1 and PCNA showed little difference in rate on SF (8,0), with a decrease to 0.018 ± 0.002 nM^{-1} min^{-1} at 1:1 ratio and a very modest increase with 1[hFEN1]:3$[PCNA]_3$, 0.032 ± 0.002 nM^{-1} min^{-1}. The small differences observed are within those associated with reproducibility errors for these experiments and therefore in our hands PCNA does not significantly stimulate the FEN1 catalysed reaction of single flap substrates at the concentrations tested.

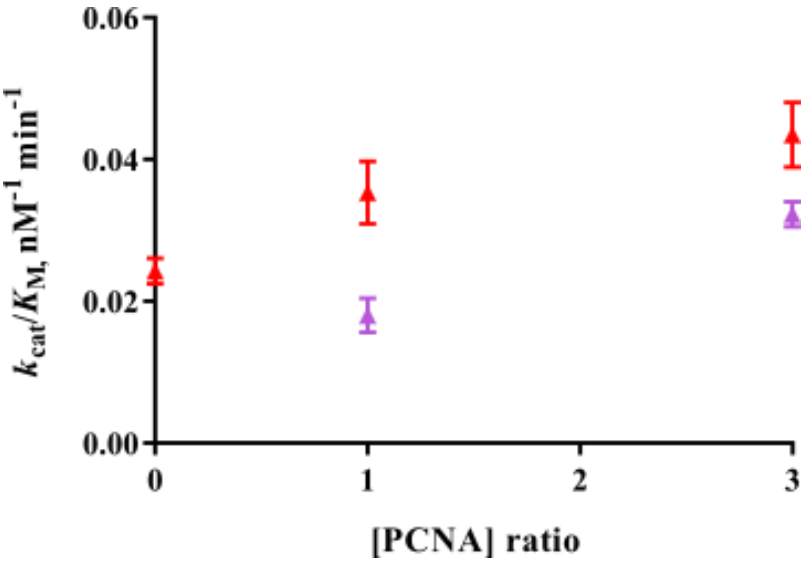

Figure 4.6: *The* k_{cat}/K_M *rate for SF (8,0) and PCNA. The red triangle data represent the conditions with different ratio between substrate and PCNA, while the triangle purple represents with different ration between human FEN1 and PCNA. The average of three replicates and standard errors are shown.*

4.4 PCNA on varying DNA substrates with same flap lengths

In the previous studies on PCNA stimulation the substrates all contained a 5'-biotin on the downstream template and upstream duplex terminus (**Figure 4.2**). This is been added to allow potentially lock PCNA on DNA after loading using clamp loader protein RFC in potential future studies. In the absence of RFC, one possibility that may have lead to a lack of stimulation of FEN1 was that this terminal biotin inhibited the ability of PCNA to load onto the substrate even in the absence of strptavidin. We therefore decided to examine substrates without a 3'-biotin on the upstream region. Therefore the investigation of the impact of PCNA on human FEN1 catalysed reaction of three different substrates with static, migrating and single flaps were analysed to determine whether PCNA could enhance the cleavage efficiency of any of these substrates. All substrates have same lengths at 5' flaps and there are single flaps, SF (8,0), static double flaps, UP31-9ST-NB (8,1) and migrating double flaps, UP31-9EQ-NB (8,1). The final [S] was 5 nM and human FEN1 was 150 pM. The ratio of [S]:[PCNA] were 0, 0.5, 1, 3, 6, 12 and 36 to produce the final concentration of PCNA; 0, 2.5, 5, 15, 30, 60 and 180 nM respectively.

Table 4.4 and **Figure 4.7** illustrates the variation in $v_0/[E]_0$ with the three different DNA substrates with PCNA concentration. For single flaps substrates the gradient of this plot is close to zero, 0.00006 ± 0.00005, while for UP31-9ST-NB and UP31-9EQ-NB both produced a small negative gradient -0.08132 ± 0.01531 and -0.01475 ± 0.00374 respectively. None of the data show increasing rate of reaction when PCNA is present in the reactions. In fact the data show that PCNA does not stimulate the reaction of any of the substrates and at the same time can retard the enzyme activities. This can be seen in the k_{cat}/K_M results, at higher concentration of PCNA (180 pM), where it decreased the rate of reaction. The defects start occur when the concentration of PCNA 6 times higher than the substrates and this observation obviously can be seen with static double flap (UP31-9ST-NB). This is probably because PCNA become as inhibitor in the reaction at higher concentration by competing with human FEN1 to form enzyme-substrate complexes by binding to the downstream DNA duplex. Thus the biotin present on the upstream duplex probably does not affect the human PCNA loading.

Table 4.4: *The k_{cat}/K_M rate analysis of single flap, static flap and equilibrating flap substrates with PCNA. The triplicates data were carried out by plotted the normalised rate of reaction ($v_0/[E]_0$) versus [S]. GraphPad Prism 7 software to acquire k_{cat}/K_M rate constants.*

DNA Substrates	Flap		*Slope of graph of k_{cat}/K_M vs [PCNA] (nM^{-1} min^{-1})*
	5’	3’	
SF	8	0	0.00006 ± 0.00005
UP31-9ST-NB	8	1	-0.08132 ± 0.01531
UP31-9EQ-NB	8	1	-0.01475 ± 0.00374

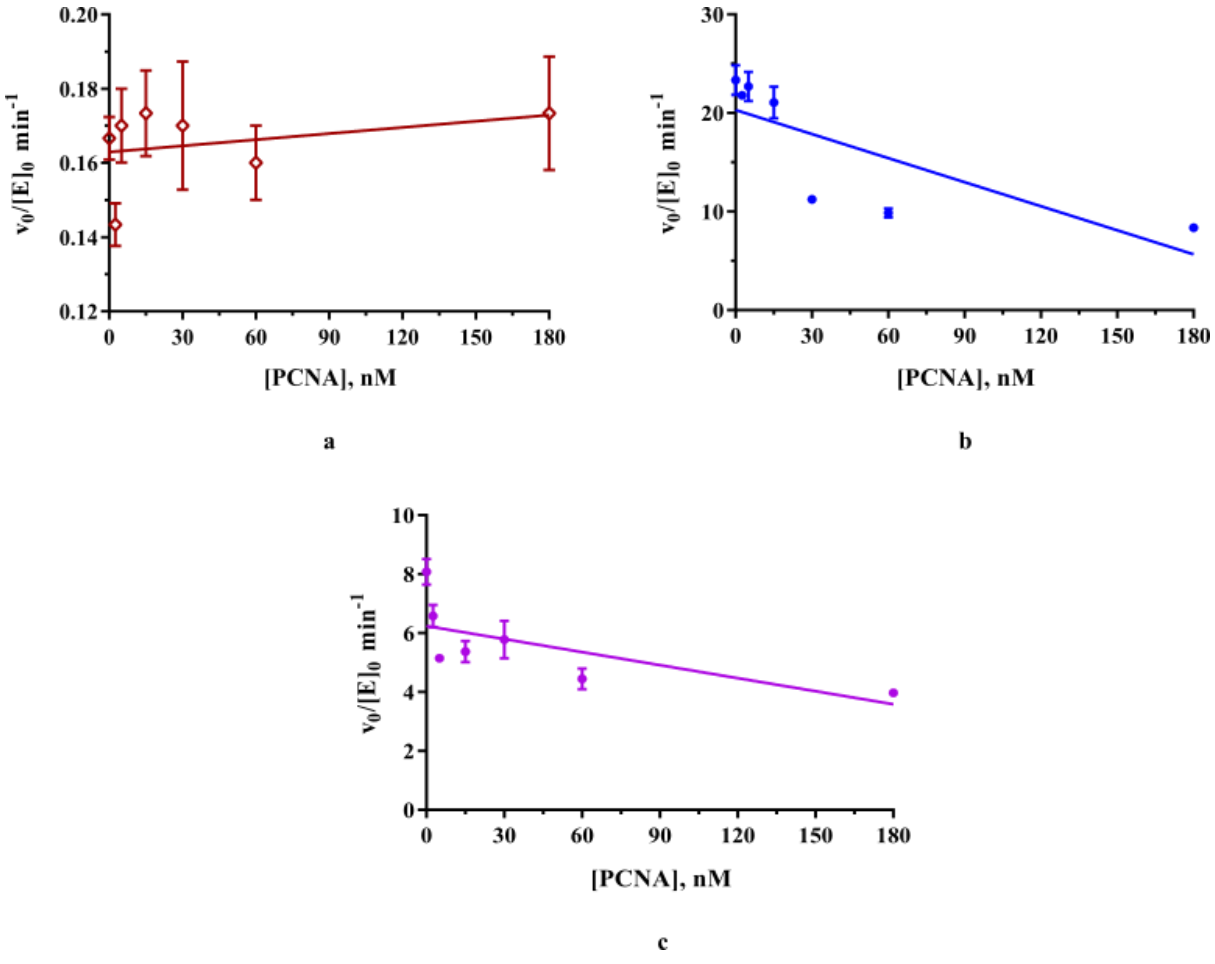

Figure 4.7: *The $v_0/[E]_0$ conditions for single flap, static flap and equilibrating flap substrates without PCNA and 6 different concentration of PCNA. a represent analysis with single flap (8,0) substrate, b represent analysis with static flap (8,1) and c represent analysis with equilibrating flap (8,1). The average of three replicates and standard errors are shown.*

4.5 Discussion and summary

The data presented in this study are contradictory to what have been published. Using 3 different types of DNA substrates, the data consistently show that the rate of reaction is reduced at higher concentrations of PCNA. PCNA can load at either duplex end of the DNA substrate used in this study. Therefore, when a larger excess of PCNA is present, it will compete and prevent hFEN1 binding to the DNA substrate, thereby preventing FEN1 hydrolysis. This type of inhibition is a consequence of our experimental set up in that the substrate possess two duplex ends to which PCNA can add; thus, these results may not reflect what occurs *in-vivo*. Furthermore, this study also revealed that PCNA does not stimulate human FEN1 activities on equilibrating flaps. This could be due to the fact that PCNA is not oriented on the substrate using our experimental setup. Previous studies proposed a few mechanisms that PCNA loaded onto the DNA[46]. Unfortunately, in these studies, none of the experiments follow any of the mechanisms. Furthermore, in cells, PCNA is unidirectionally loaded onto the DNA by RFC. Therefore the absence of clamp loader RFC in the experiments may be a factor in the absence of stimulation of FEN1 by PCNA. However, it should be noted that the majority of earlier studies loaded PCNA in the same way attempted here and report a stimulation that we cannot reproduce. In summary, this study reveals that a more complicated assay involving the loading and trapping of PCNA onto substrates may be necessary to explore more fully how FEN1 and PCNA work together.

Chapter 5: The Effect of human FEN1 Mutants on The Arch

In the human FEN1 crystal structure, a helical arch was discovered consist of two alpha-helices (α4 and α5) and was large enough for 5' ssDNA flap to thread via the hole under it[9]. Based on the observation that hFEN1 could accommodate substrates that contained a short region of duplex within the 5'-flap that could not have threaded through the structured arch, a disordered-to-ordered helical state mechanism was proposed[72]. In the presence of the correct DNA substrate, the arch was believed to adopt an ordered conformation and vice versa without the DNA substrates (**Figure 5.1**). Previous studies showed, the arch contained the conserved gateway (base of α4) residues including Lys-93 and Arg-100[75] while the other residues of the arch located in the helical cap (top of α4 and α5) were just thought necessary to hold these in place and to confine reaction to substrates with 5'-termini[75]. Lys-93 and Arg-100 were believed to facilitate capture of the scissile phosphate and also proposed as electrostatic catalysts during the FEN1 cleavage reaction[9,73]. If the arch was disordered it was assumed that the scissile phosphate was located near but not within the active site and could not interact with metal ions or amino acid residues for hydrolysis. More recent studies have suggested that the helical cap could play a more active role in the FEN1 reaction[75]. In addition to Lys-93 and Arg-100 gateway region (α4), Lys-125, Lys-128 and Arg-129 in cap region (α5) and also Arg-103 and Arg-104 gateway/cap region (α4)[6,9] appear to form substrate interactions in the most recent crystal structures.

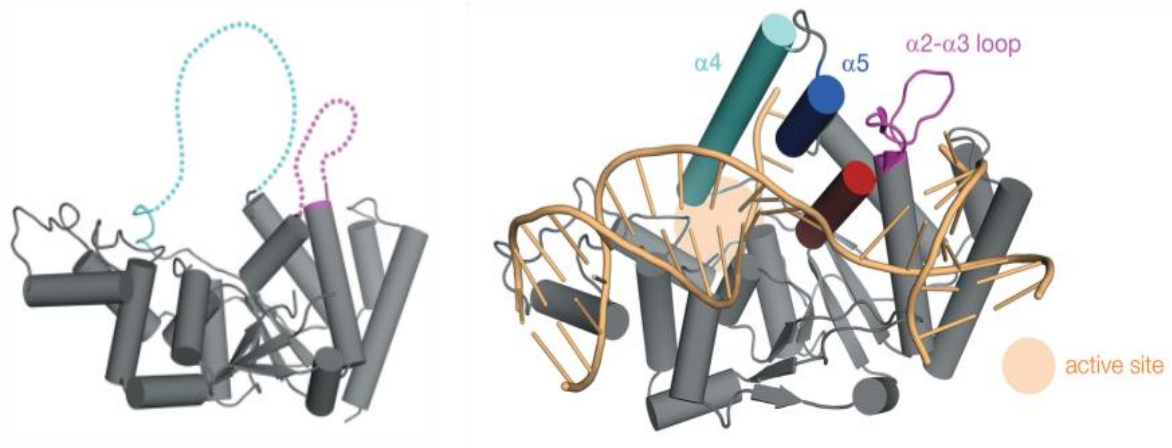

Figure 5.1: *The structure of FEN1s with disordered (left) and ordered (right) arch and α2-α3 loop. Left (PDB code 1A76[121]) is the unstructured helical arch in the absence of DNA substrate (blue dotted lines). Right (PDB code 5UM9[126]) is an ordered structure of α4 and α5 in the presence of DNA substrate[93].*

Investigation into the role of conserved residues in FEN1 family have involved the characterisation of various side-directed mutants, where the helical arch residues (consists of α4 and α5) have been mutated either to alanine (A), glycine (G) or glutamic acid (E). Side-directed mutations are defined by nomenclature such as K125A, where the lysine (K) at position 125 has been mutated to an alanine (A). The 5' ssDNA has been proposed to through the helical arch via threading mechanism. Mutation of the residues in this region will help with better understanding as to how the threading mechanism occurs. Each of the targeted residues was substituted with alanine to remove charge on the helical arch. Within the helical arch of the FEN1 family, there are a number of conserved and semi-conserved basic residues, which are proposed to be important in 5' ssDNA threading[93] including, Arg-103 (R103), Arg-104 (R104), Arg-129 (R129), Lys-132 (K132) and Lys-125 (K125).

5.1 Mutational analyses

To understand more about the importance of helical cap substrate interactions human FEN1 Lys-125 on α5 was targeted for mutagenesis. Lys-125 was suggested to be important interaction between α5 and the template strand of the DNA together with other residues including Lys-128 and Arg-129. Lys-125 is an invariantly conserved residues between human and archaeal FEN1 proteins, its sidechain protrudes from the front of the arch and makes a salt bridge with a phosphate diester of the DNA template (approximately 6.7 Å) (**Figure 5.2**). The targeted residues were substituted with alanine (K125A) to neutralise the charge. Mutagenesis was carried out using the site-directed mutagenesis methodology. The resultant protein was expressed in *Escherichia coli* BL21(DE3)-RILP using auto-induction methods as describe in section 2.1 and the protein was purified using same methods as discussed in section 3.1. The protein purity was assessed by SDS PAGE as shown in **Figure 5.3**. The single band from each of concentrations; 0.1, 0.2, 0.4, 0.6, 0.8, 1.0 and 1.6 μg/μL proved that the K125A mutant was pure and not contaminated. Apart from K125A mutant, mutants of other basic residues have been constructed and purified previously including R103A, R104A, R129A, K132A and double mutants R104AK132. All of these mutations were believed to interact with the DNA substrates, either with the template or with the single strand DNA in the threaded state.

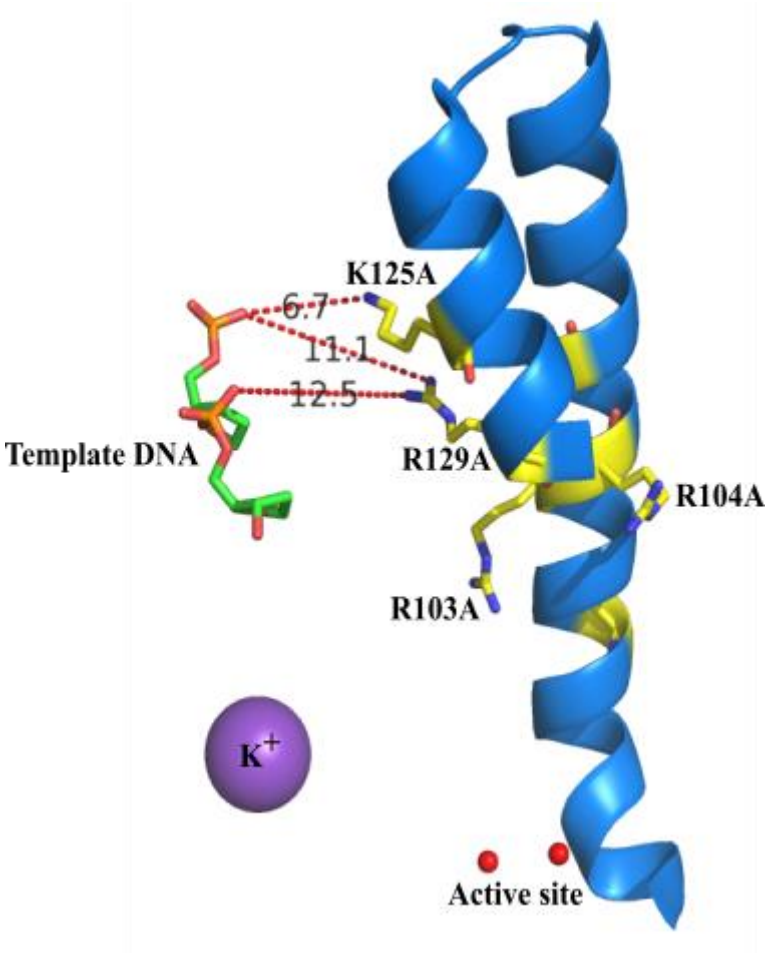

Figure 5.2: *The structures K125 residue with DNA substrate (PDB:3Q8L). The cartoon shows the view of α4 and α5 from the side. The distance between K125 and DNA template substrate approximately 6.7 Å, whereas R129A between 11.1 to 12.5 Å.*

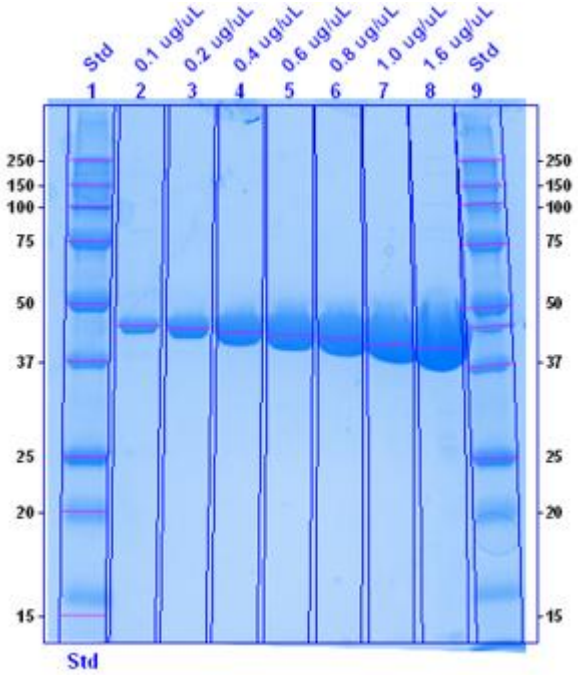

Figure 5.3: SDS-PAGE to show the purity of K125A after purification steps. *Lane 1 and lane 9: Protein ladder (representing 15, 20, 25, 37, 50, 75, 100, 150 and 250 kDa), lane 2: 0.1 µg/µL, lane 3: 0.2 µg/µL, lane 4: 0.4 µg/µL, lane 5: 0.6 µg/µL, lane 6: 0.8 µg/µL, lane 7: 1 µg/µL and lane 8: 1.6 µg/µL.*

Kinetic studies of K125A were undertaken using a static double flap substrate (single conformer) denoted as DF5,1 (DF = double flap with 5 nucleotides 5'-flap and 1 nucleotide 3'-flap). The substrate had a 5'-fluorescein (FAM) labelled at 5'-flap similar to the substrates that have been used for wild type human FEN1 analysis (**Figure 3.5** in chapter 3). The single turnover analysis (enzyme in excess) were conducted with a final concentration of the substrate was 5 nM and the concentration of K125A was 1000 nM. Following determination of the concentration of product with respect to time, the data were fitted by *GraphPad Prism 7 software* with a single-phase exponential. The k_{st} rate of K125A mutant was 14-fold slower (51 ± 2 min^{-1}) compare with the wild type human FEN1 rates of 788 ± 39 min^{-1} as shown in **Figure 5.4 and Table 5.1**. The slower k_{st} rate is because of the absence positive charge and as a consequence the DNA template interacts or binds more weakly with the enzyme.

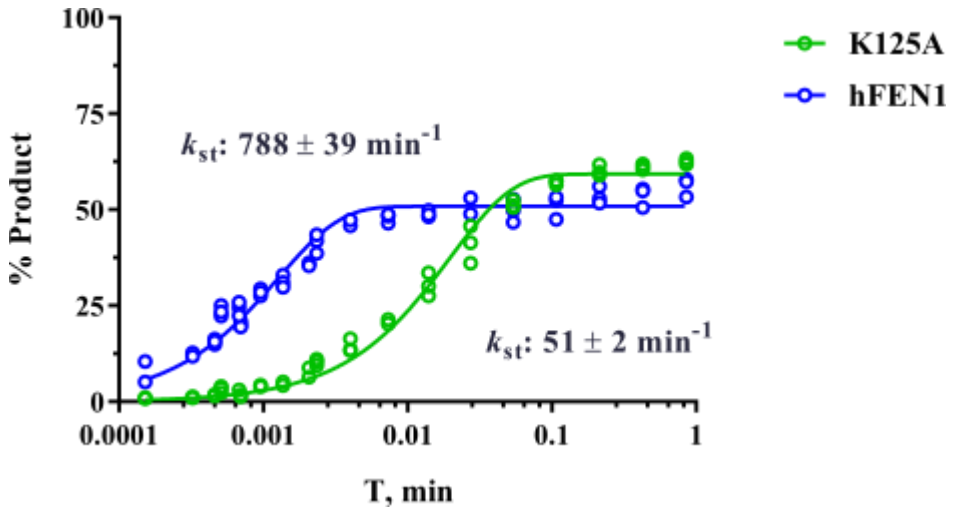

Figure 5.4: *The single turnover rate for DF5,1 cleavage by hFEN1 and K125A mutant. The data points open blue circle represent the human FEN1 and open blue green represents the K125A mutant. The mean of three replicates and standard errors are shown. Data has been fit to a single-phase exponential.*

Table 5.1: *The single turnover rate for WThFEN1 and K125A. All data were plotted with single-phase exponential by GraphPad Prism 7 software.*

Parameters	**Enzyme**	
	WThFEN1	**K125A**
Plateau	51 ± 2	59 ± 1
k_{st}, **min^{-1}**	788 ± 39	51 ± 2

5.2 The role of helical arch in threading mechanism

Previous studies have proposed several mechanisms for FEN1 accommodation of DNA substrates including the threading, the tracking, the bind and thread and also the clamp models. Lyamichev et. *al.*, 1993 first proposed that FEN1 specificity for 5' flaps could originate via a threading mechanism[16]. It was proposed that the 5' flap DNA substrate was threaded through a hole of appropriate size[16] and later the helical arch became a candidate for this. For tracking models, the hypothesis was, that FEN1 initially recognised the 5' ssDNA by clamping or threading, followed by sliding down the 5' termini to the bifurcation region where cleavage occurs[90–92]. In contrast in the bind and thread mechanism, FEN1 initially binds with the double strand DNA substrate at the bifurcated

region and threads the 5' flap ssDNA through the helical arch. A further study proposed, that the helical arch undergoes a disorder-order transition during the threading of 5' flaps ssDNA[9,71,75]. The last hypothesis was the bind and then clamp model by Orans et. *al*., 2011. They suggested that FEN1 bound to the double strand DNA substrate initially at bifurcated region, and continued by clamping the 5' flap ssDNA via either side of the helical arch[17]. However, when structures of FENs bound to full substrates were finally solved, in agreement with the biochemical data, the single stranded flap was found to be threaded through the helical arch[26,72].

The helices are amphipathic in character, where positive residues pointing inward and outwards (Lys-125) to provide binding surface for DNA substrate. The 5' ssDNA flap of the substrate was seen to interact with the arch region. As discussed before, in the structure of human FEN1 without DNA the helical arch region is disordered therefore, the binding between helical arch and DNA substrate may promote the ordering of the arch. In this study, we aim to investigate each mutation in kinetic experiments to allow effective quantitative comparison of catalytic contributions by measuring the k_{st} rate.

Previous studies revealed that addition of a 5'-streptavidin to the 5' ssDNA flap of a double flap substrate is inhibitory to cleavage[75]. Streptavidin is a large stable protein that creates a steric road "block" to the threading of flaps through the helical arch when it is bound the 5'-biotinylated substrates. As shown in **Figure 5.5 b**, to create a "blocked" complex the streptavidin (SA) and DNA substrate are pre-incubated first to form a SA-conjugated substrate, and then human FEN1 is added. In contrast if human FEN1 and 5'-biotinylated substrate are pre-incubated to form enzyme-substrate complex (**Figure 5.5 a**), addition of streptavidin would "trap" the threaded 5'flap substrate on the protein. Previous experiments have demonstrated that the maximum proportion of trapped complex was obtained with human FEN1 in the presence of the inert cofactor calcium ions. The properties of these complexes were compared with control experiments where a 'premix' complex was obtained via the same procedures by adding Ca^{2+} buffer instead of streptavidin. Earlier, it has been shown that when reaction is initiated with Mg^{2+} the trapped complex decays at a similar rate to the premixed one.

The trapping and blocking data were accumulated independently from manual sampling (on the bench) and rapid quench-flow experiments and quenched with urea and EDTA. A two-phase exponential has been used to fit all the data (trapped and premixed) by *GraphPad Prism 7* except for blocked experiment where a single-phase exponential was preferred. The two-phase exponential was preferred[13] because, the 'fast' phase was considered as the rate of reaction (k_{st} fast), while the second phase (k_{st} slow) has been proposed to result from the initial formation of unproductive complex formation[93]. These unproductive complexes would need to dissociate and re-associate to form productive complexes to react, thereby giving rise to the slower second phase. The maximal k_{st} rate were determined and compared to the wild type human FEN1. Under premixed and trapped conditions, the cleavage would be expected occur very fast. Whereas, for blocked condition, the hydrolysis occurs very slow because the DNA substrate is blocked sterically by streptavidin and therefore cannot thread through the helical arch. If a mutant enzyme was deficient in threading the substrate through the helical arch, it would be expected to form both trapped (threaded) and blocked (unthreaded) complexes when streptavidin was added to a complex with 5'-biotinylated substrate.

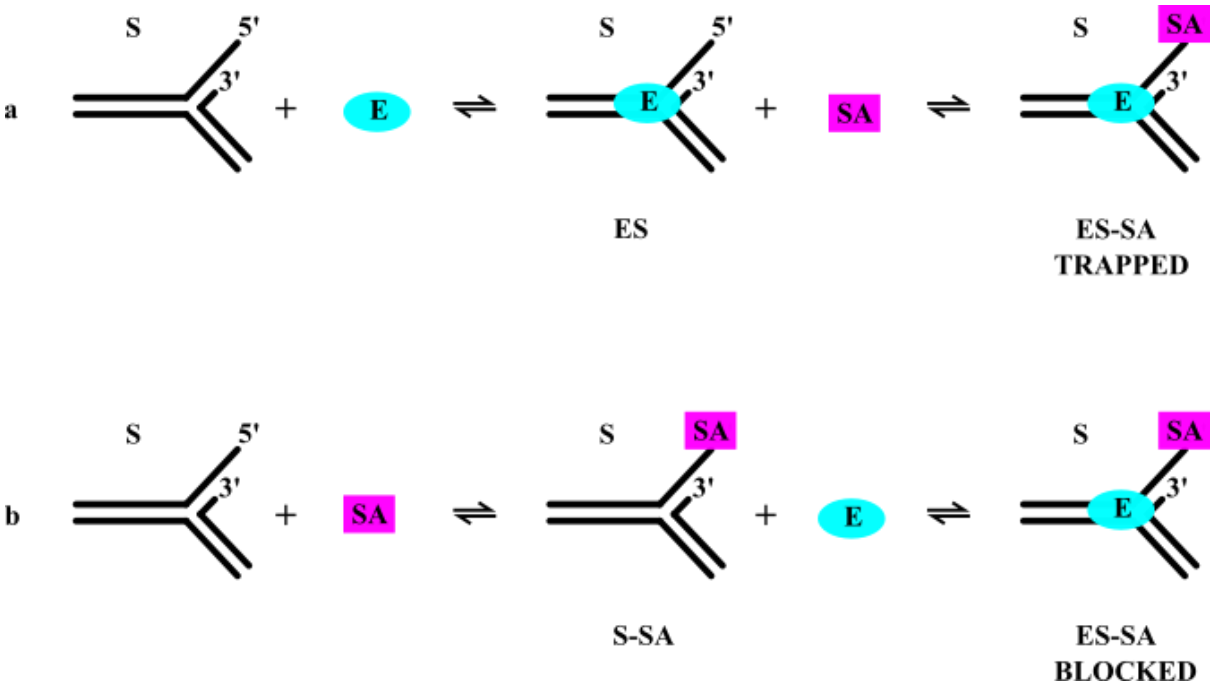

Figure 5.5: *Structure of double flap DNA and schematic of threading assay. a illustrating the formation of ES-SA under trapped conditions and b showing the formation of ES-SA under blocked conditions. Complexes were assembled in Ca^{2+} buffer with Mg^{2+} added to initiate the reaction. S, E and SA represent substrate, enzyme and streptavidin respectively.*

5.2.1 Investigating the effect of mutations of basic residues on the human FEN1 mechanism using single turnover kinetics

The ability of the 5' flaps of ssDNA to thread under the helical arch was investigated via blocking and trapping experiments. The static double flaps substrate SB5,1 had a 5 nts 5' flap and 1 nt 3' flap and was FAM labelled and biotinylated. The final concentration of SB5,1 and enzyme were 5 nM and 500 nM respectively. Four single mutants of human FEN1 were studied: R103A, R104A, K125A, R129A and double mutant R104AK132A. 5-fold molar excess streptavidin (with respect to substrate) was added in the reaction mixtures to produce blocked and trapped complexes. Since Ca^{2+} is a competitive inhibitor with respect to viable cofactor ions, complexes were assembled in the presence of Ca^{2+} ions. Reaction was initiated at 37 ^{0}C by the addition of Mg^{2+} reaction buffer, and aliquots were removed at various time points and quenched with 8 M urea and 80 mM EDTA for manual analysis and 8 M urea and 300 mM EDTA for experiments performed using RQF-63 quench flow device. The product formation was measured by dHPLC equipped with a fluorescence detector and was plotted to determine the initial rate of reaction.

5.2.2 Studying the temperature effect of threading mechanism

Initial experiments were carried out with two conditions for complex formation, either on ice or at room temperature. When a complex of enzyme R104AK132A and biotinylated substrate was formed and then trapped on ice by adding streptavidin only 16.6% (equivalent to 0.83 nM) product formation was observed as the end point of first phase when reaction was initiated with Mg^{2+} ions. In contrast 35% (equivalent to 1.75 nM) product formation was observed when the complex was formed and trapped at room temperature. In each case and percentage of product formation were calculated via **Equation 5.1**;

$$\%\ of\ product\ formation: \frac{Concentration\ of\ product\ formation\ (nM)}{Concentration\ of\ substrate\ (nM)}\ x\ 100$$

Equation 5.1

Where, [S] was 5 nM in all cases.

Streptavidin trapping is thought to be equally efficient under both temperature conditions suggesting that the stability of the enzyme-substrate complex is likely the factor that varies with temperature. **Table 5.2** and **Figure 5.6** show the results of blocked, trapped and premixed experiment using enzyme R104AK132A on ice and at room temperature. Under trapped experiments, the data displayed two phases. The first phase has a rate which is similar to that of the premixed fast enzyme-substrate complexes whereas the second phase (k_{st} slow) has a similar rate to that of the blocked condition.

Table 5.2: *The results from GraphPad Prism 7 software for double mutants R10AK132A. The results were fitted with one-phase association (for blocked) and two-phase association (trapped and premixed) to show the plateau, percentage of fast and k_{st} rate for two different temperature conditions.*

		R104AK132A	
Parameters	**Analysis**	**On ice**	**Room temperature**
Plateau (nM)	Blocked	3.395 ± 0.1027	3.524 ± 0.06439
	Trapped	3.12 ± 0.05	3.78 ± 0.04
	Premixed	4.40 ± 0.08	4.47 ± 0.06
Percentage fast	Blocked	NA	NA
(%)	Trapped	23.69 ± 1.26	43.65 ± 0.85
	Premixed	62.91 ± 1.67	63.80 ± 1.16
k_{st}, s^{-1}	Blocked	$0.0002012 \pm 2.102 \times 10^{-5}$	0.0002819 ± 0.06439
	Trapped fast	0.034 ± 0.005	0.049 ± 0.002
	Trapped slow	$0.0003119 \pm 2.352 \times 10^{-5}$	$0.0002342 \pm 1.542 \times 10^{-5}$
	Premixed fast	0.261 ± 0.022	0.187 ± 0.011
	Premixed slow	0.00156 ± 0.000285	0.0009364 ± 0.0001199

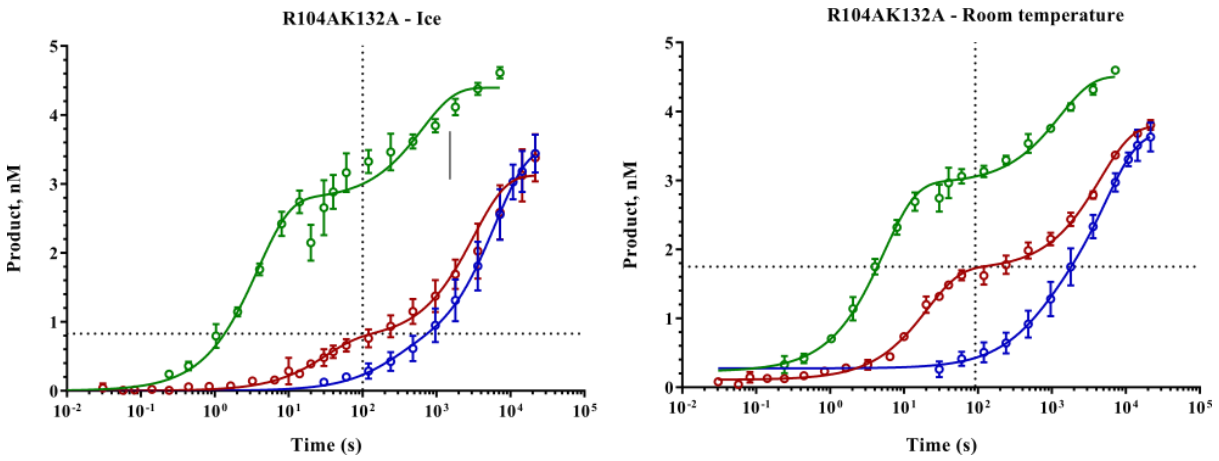

Figure 5.6: *The threading analysis of double flaps substrate (SB5,1) with double mutants R104AK132A. The experiment on the a) ice, b) room temperature whereas the green, red and blue colours are represent premixed, trapped and blocked respectively. The dotted lines represent the end point of the first phase for threading conditions in each cases.*

5.2.3 Exploring the function of basic residues of the helical arch of human FEN1 via threading analysis at room temperature

The investigation for other mutants via threading analysis were continued with the same conditions as described above with complexes formed at room temperature. Arg-103, Arg-104, Lys-125, Arg-129 and Lys-132 are five basic residues that conserved and semi-conserved in human and archaea FEN1 proteins. A previous study proposed that Arg-103, Arg-104, Arg-129 and Lys-132 could help to stabilize the inverted 5' ssDNA flap and steer the phosphodiester backbone during the threading process[126]. All the basic amino acids were mutated to alanine (R103A, R104A, K125A, R129A, R104AK132A) to remove the positive charge and purified in a similar way to wild type hFEN1. For each mutant, except K125A, the multiple turnover rates of reaction were analysed by other members of the Grasby group. Compared with wild type human FEN1, the rate of reaction of R103A was slightly slower, while R104A and R129A both showed 20-fold decreases. The double mutants R104AK132A showed the worst effect with a 200-fold reduction in the $v_0/[E]_0$ rate.

A slower "premixed" rate is expected during trapping/blocking experiments compared with the rate of normal single turnover reaction. For direct comparison with trapped and blocked experiments, here the premixed complex is formed in the presence of calcium ions as these were previously found to be necessary for full threading to occur. Thus a 'delay effect' is expected when mixing in the excess of Mg^{2+} to initiate reaction as Ca^{2+} ions have to be displaced from the active site[93] (**Figure 5.7**). This 'delay effect' would be expected to be greatest when the reaction rates are fastest and can be seen by comparison of results for wild type hFEN1. The maximal single turnover rate of reaction of the static double flap (SB5,1) form previous study[93], the k_{st} rate was 2430 ± 30 min^{-1} whereas in the presence of Ca^{2+} in threading experiment the rate is 67 ± 6 min^{-1}. These result showed the 'delay effect' slower the reaction by approximately 40-fold[93].

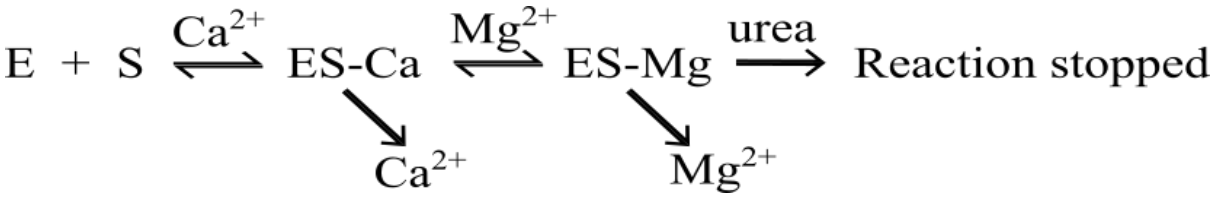

Figure 5.7: *The illustration of Ca^{2+} as competitive inhibitor of reaction in a "premixed" experiment.*

The presence of streptavidin-biotin complexes under threading experiments, two conditions could happen. First, either 'blocked' or 'trapped' depending on the sequence of addition of the streptavidin. **Table 5.3** and **Figure 5.8** show the threading analysis for WThFEN1 and five mutants. The data can be explained by comparing each mutant with end point of premixed (control) condition and end point of trapped conditions. Based on **Figure 5.8**, the wild type human FEN1 showed that the end point of premixed and trapped condition were similar. These results demonstrated that the substrate-enzyme complexes are fully threaded at equilibrium[93], and we observed the same results for R103A. For other mutants (R104A, K125A, R129A and R104AK132A), the product formation at the end point of first phase are lower compared with the premixed. This results like arose because the protein failed maintain the threaded state at equilibrium.[93].

Table 5.3: *Trapping and blocking experiments for five mutants. The results are given as the average of at least four replicates and errors. The results were fitted with one-phase association (for blocked) and two-phase association (trapped and premixed) to show the plateau, percentage of fast and k_{st} rate under Ca^{2+} buffer at room temperature.*

		Enzyme					
Parameters	**Analysis**	**hFEN1**	**R103A**	**R104A**	**K125A**	**R129A**	**R104A K132A**
Plateau (nM)	Blocked	4.18 ± 0.06	4.70 ± 0.03	3.56 ± 0.06	4.70 ± 0.03	4.05 ± 0.05	3.52 ± 0.06
	Trapped	4.47 ± 0.06	4.68 ± 0.08	3.87 ± 0.07	4.74 ± 0.05	4.31 ± 0.08	3.79 ± 0.03
	Premixed	4.71 ± 0.05	4.62 ± 0.05	4.59 ± 0.05	4.74 ± 0.05	4.38 ± 0.05	4.51 ± 0.05
Percentage fast (%)	Blocked	NA	NA	NA	NA	NA	NA
	Trapped	63.48 ± 1.40	59.96 ± 1.88	49.03 ± 1.15	51.70 ± 1.06	51.02 ± 1.54	43.25 ± 0.69
	Premixed	62.87 ± 1.36	60.24 ± 1.96	63.06 ± 1.14	75.56 ± 1.08	62.41 ± 1.34	63.57 ± 1.01
k_{st}, s^{-1}	Blocked	0.005954 ± 0.0004544	0.003628 ± 0.0001361	0.0001106 ± 3.882×10^{-6}	0.001292 ± 4.214×10^{-5}	0.0002681 ± 1.372×10^{-5}	0.0002819 ± 0.06439
	Trapped fast	0.388 ± 0.033	0.429 ± 0.053	0.617 ± 0.047	0.265 ± 0.023	0.243 ± 0.025	0.049 ± 0.002
	Trapped slow	0.0005366 ± 8.319×10^{-5}	0.0005982 ± 0.0001146	0.0001769 ± 1.869×10^{-5}	0.0006007 ± 5.67×10^{-5}	0.0002545 ± 3.598×10^{-5}	0.0002208 ± 1.133×10^{-5}
	Premixed fast	1.123 ± 0.097	2.340 ± 0.278	0.378 ± 0.023	1.607 ± 0.099	1.300 ± 0.105	0.187 ± 0.011
	Premixed slow	0.005338 ± 0.0007322	0.01184 ± 0.001971	0.001552 ± 0.0001868	0.003219 ± 0.000546	0.003638 ± 0.0005056	0.0007763 ± 8.189×10^{-5}
Threading efficiency		1.00	0.99	0.78	0.68	0.82	0.68

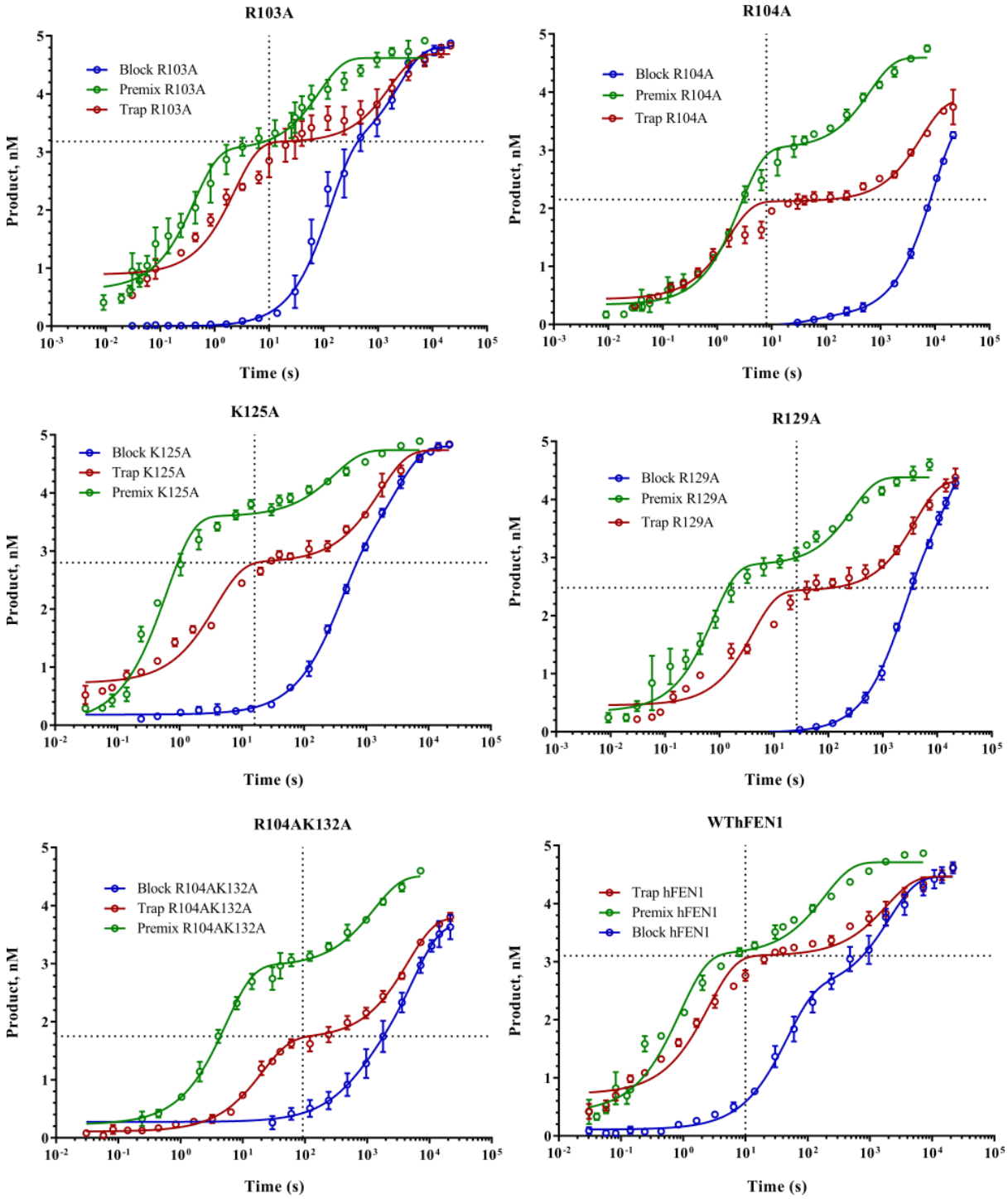

Figure 5.8: *Single turnover analysis for threading analysis for R103A, R104A, K125, R129A, R104AK132A and WThFEN1. The mean points open green, red and blue circle data represent premixed, trapped and blocked experiments respectively. Each analysis were done four replicates and standard errors are shown. The dotted lines represent the end point of the first phase for threading conditions in each cases. The results were fitted with one-phase association (for blocked) and two-phase association (trapped and premixed) to show the plateau, percentage of fast and k_{st} rate under Ca^{2+} buffer at room temperature.*

As we can see from **Figure 5.8**, a second phase that reacts as the 'blocked' mechanism, and those represent the DNA substrates in the equilibrium that are not threaded or probably unproductive products. The blocked reaction in all cases were slow as expected because of the presence of streptavidin on the 5'-flap. The rate of reaction for R104A, K125A and R129A were decreased roughly 1000-fold above than control complexes after initiation of Mg^{2+}. On the other hand, R103A and R104AK132A substrate complexes decay in blocked complexes were approximately 600-fold slower than premixed complexes. The WThFEN1 blocked reaction only is only around 200-fold reduced compared with trapped fast phase.

With the premixed reactions, rapid product formation results (an exponential curve) as single-strand 5'-flap threading is uninhibited and substrate-enzyme complexes are fully threaded. Without streptavidin, the reaction appears to result in the fastest threading and product formation. On the other hand, the slow phase on trapped rate in each case were approximately similar or close with blocked rate. This correspondent to the assumption that the substrate was not bound and threaded at equilibrium because of the presence streptavidin and this data parallel with previous study[93] (**Figure 5.9**).

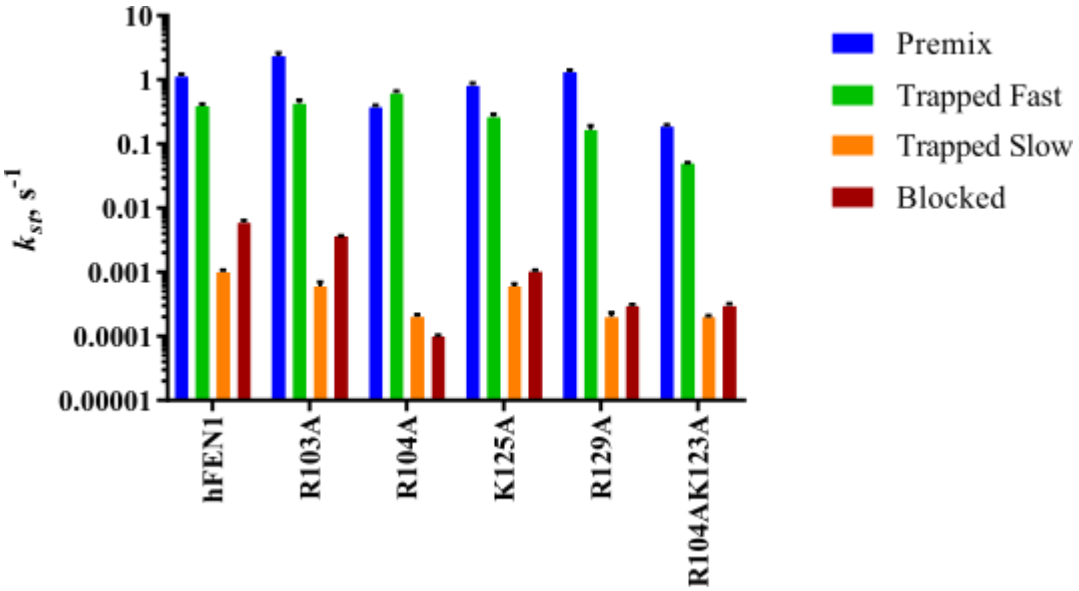

Figure 5.9: *The k_{st} rate of single mutant; R103A, R104A, R129A, K125A and double mutant R104AK132A compared with human FEN1. The results were fitted with one-phase association (for blocked) and two-phase association (trapped and premixed) by GraphPad Prism 7 software have been used and error bars show.*

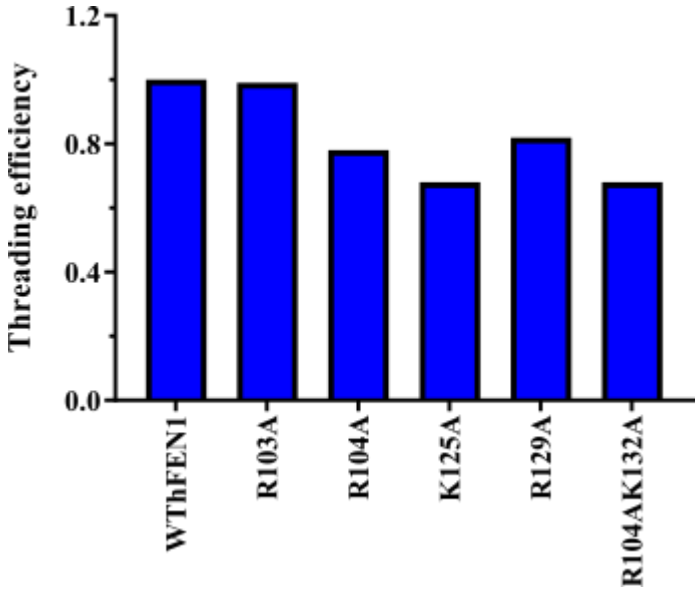

Figure 5.10: Threading efficiency derived for R103A, R104A, K125A, R129A and R104AK132A compared with wild type human FEN1. The data were derived as the ratio of fast values from trapped and premixed reactions[93].

Figure 5.10 and **Table 5.3** represent the threading efficiency (between 0, unthreaded; and 1 fully threaded[93]) of WThFEN1 and five mutants residues. These bar graph were calculated by **Equation 5.2**,

$$Threading\ efficiency = \frac{Percentage\ of\ fast\ trapped}{Percentage\ of\ fast\ premixed}$$

Equation 5.2

The residues on the helical arch mutants, R104A, K125A, R129A and R104AK132A all revealed threading defects except for R103A. This suggesting that each of these residue play a role during the threading processes either to stabilise the ordered helical arch or contact with the substrate. While regarding with Tainer et *al.*, 2017, they suggested basic residues are required to facilitate threading and protect the flap from reacting by interacting with the phosphates in the 5'-flap. Therefore, in our assay, the stability of 5'-flap threaded state depends on rate of threading and rate of un-threading.

5.2.4 Discussion and summary

Tainer et. *al.*, 2017 proposed Arg-103, Arg-104, Arg-129 and Lys-132 were 'phosphate steering' residues[126]. Moreover, these residues helped 'steer' the 5'-flap portion of substrates away from the active site to prevent inadvertent reaction in the flap. In addition, some of these residues were also thought to be important for positioning the scissile phosphate through a twisting or rolling type of motion on the active site metal ions[126]. Thus, phosphate steering was define as 'electrostatic interaction that can dynamically position the phosphate diester backbone[126].

The similarity of Arg-103 and wild type hFEN1 results proved this residue is not important for phosphate steering. It is possible that, the other mutants impair the threading rate by destabilizing the ordered threaded state. The results of Lys-125 and Arg-129 indicated an important role for contacts made from wild type human FEN1 to the template strand phosphate diesters. Even though both residues show strong conservation across eukaryotes and archaea, the functional significance of these contacts has not been fully appreciated. Arg-104, Lys-125, Arg-129 and Lys-132 results proved that all these residues have specific contacts with DNA substrate and play important role in promoting the succession of protein and DNA conformational changes leading to catalysis.

An interesting observation under trapped conditions with streptavidin resulted in two output: one phase that reacted rapidly known as trapped fast, whereas another phase reacted on the timescale of blocked complex known as trapped slow. Generally, the rate of reaction of trapped fast enzyme-substrate complexes are approximately similar to the premixed rate, showing all substrates are accommodated identically in a threaded state at equilibrium. Under trapped substrate conditions, we observed the biphasic correlation. This suggest that the reaction is not perfectly trapped, probably due to unproductive binding (i.e., incorrect binding between enzyme and substrate). In contrast, 5'-flap blocked substrates have drastically decreased the rate of reaction because of the substrate cannot be optimal bound to the enzyme and cleavage does not occur on a biologically relevant timescale. These results are predicted because of a steric hindrance of streptavidin on the 5'-flap dsDNA substrate probably affecting the efficient position of helical arch residues. In addition, the 5'-flap threading hypothesis was successfully verified by the streptavidin binding experiments, with the data regarding the difference order of the adding of

streptavidin. The reaction for 5’-blocked is at least 100-fold slower than trapped fast phase indicating the importance of the threading. However, in reality, the reaction were 1000-fold slower because of delay effect.

Chapter 6: Observation Local Conformation Changes of Nucleotides in The human FEN1 by Exciton Coupled Circular Dichroism

Although the ground state of DNA is a double helix, it is still a dynamic biomolecule that is flexible and has the ability to adopt alternative excited states[127]. Previous studies revealed that human FEN1 could discriminate between the structures of potential DNA substrates at specific regions by various mechanism. Firstly, a feature of the discontinuous bifurcated structures such as flaps is that they have enhanced conformational flexibility at the duplex–duplex junction. They can be bent at dsDNA-dsDNA junction removing the coaxial stack, or alternatively the two duplex arms can coaxially stack with each other[9,22,128]. Secondly, FEN1 recognizes the one nucleotide 3'-flap of the DNA substrate[13,85,86]. Thirdly, the single-strand 5'-flap of the DNA has to pass through the hole under the helical arch (between α4 and α5) for the enzyme-catalysed reaction to occur at a significant rate[72,129]. Finally, the DNA substrate has to untwist to allow the scissile phosphodiester bond to enter the active site for hydrolysis[126]. Understanding the requirements for and order of these events will aid understanding of the complex process of substrate recognition by FEN1. DNA untwisting is accompanied by substantial helical distortion and adjustment to the stacking environments within the duplex. We therefore, chose to investigate the ability of FEN1 proteins to bring about this step using the 2-aminopurine Exciton Coupled Circular Dichroism (ECCD) signal at 326 nm as this signal is very sensitive to duplex environment.

ECCD is a sensitive and established technique that can be applied to the study of macromolecular structure. ECCD represents an interaction of two chromophores in a chiral macromolecule. When the two chromophores are in position to 'be coupled', their electronic transition dipole moments interact and generate two distinct CD bands, which have opposite signs due to Cotton effects. ECCD analysis could help to understanding the structural changes at specific site of adjacent bases at the ss/ds junction of human FEN1 substrates if tandem modified bases with a distinctive ECCD signal are introduced into the duplex. Changes in these spectra on addition of protein and/or cofactors demonstrate a change in nucleobase orientation has occurred. The circumstances under which the changes occurred could lead to information about what the conformation could be and what the requirements are.

6.1 The role of 2-aminopurine

Nucleobase cross-linking studies by Beddows et *al.*, 2012 first suggested the necessity of for DNA conformational change, which at that point was thought to be double nucleotide unpairing (DNU), for the FEN1 reaction[38]. These studies revealed that the nucleotides at ss/dsDNA junction must be capable of movement for reaction to occur. To understand the nucleotide dynamic, the spectral probe 2-aminopurine (2-AP) was substituted within a DNA construct at the base of the 5'-flap in two adjacent sites. 2-AP is an adenine analogue that which still pairs to thymine and absorbs and emits light in a region where other nucleobases are transparent. The replacement of two neighbouring nucleotides (**Figure 6.1**) with 2-AP molecules will affect the energy levels and also excitation coupling of the electronic fields[130]. The substitution of 2-AP for adenine will reposition the exocyclic amino group but does not significantly alter the structure of stability of DNA[131,132]. Furthermore, 2-AP in tandem at the end of the flap strand provides a characteristic spectrum when contained within duplex DNA and when separated from one another. By substituting the nucleotides with 2-AP which has an electric dipole transition moment, the base stacking and conformational changes can be monitored.

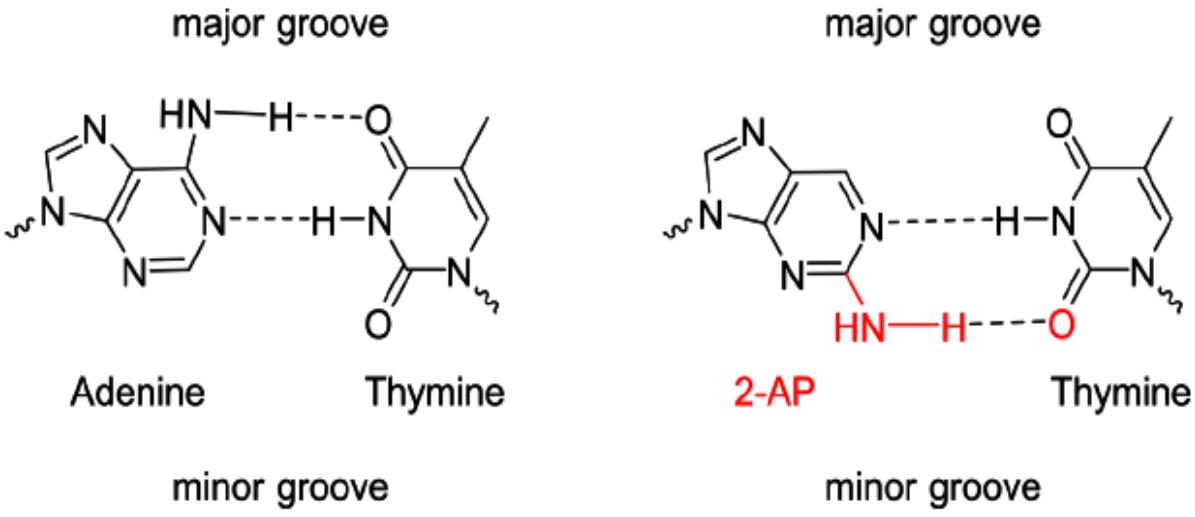

Figure 6.1: *The hydrogen bond between adenine:thymine (left) and 2-aminopurine:thymine (right) is altered in the 2-AP substituted base pair*[130].

6.2 Substrate design for ECCD experiment using 2-aminopurine

Previous studies proposed that the eukaryotic FEN1 proteins could recognize specific junction known as a 5'-3' double flap which can have a single nucleotide 3'-flap and any 5'-flap length including zero. Based on **Figure 6.2**, the nicked product (P) and single stranded product (Q) are formed after substrate hydrolysis resulting in a one nucleotide 3'-flap and a one nucleotide gap[133]. In addition, the distortion or untwisting of the double strand DNA near the site of cleavage is important for hydrolysis because it places the scissile phosphate on the active site metals. 2AP ECCD spectra of substrate alone typically show a maxima around 326 nm. ECCD spectra were measured from 300 nm to 480 nm with 0.4 seconds per data point and recorded in a 5 mm path length at 20 ^{0}C. The analyses were performed in a Ca^{2+} buffer, with and without 25 mM EDTA with final concentrations of enzyme and substrate of 12.5 μM and 10 μM respectively. The data were processed by Pro-Data Viewer and all spectra had blank subtraction for baseline-correction. Under these conditions, we believed that all the DNA was fully bound to the protein during the ECCD measurements. The ECCD traces and the ellipticity differences of each complex were monitored in the presence and absence of divalent metal ions and calculated from three independent repetitions.

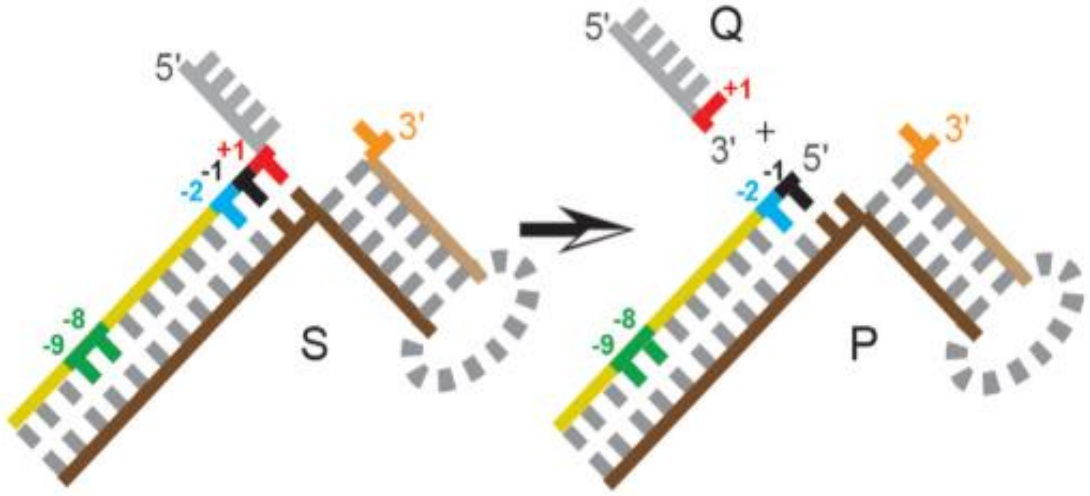

Figure 6.2: *The reaction of non-complementary double flap DNA substrate. S is static double flap, P is nicked product and Q is single stranded product. The pictures shows the FEN1-catalyzed reaction occurs between +1 and -1 nucleotides*[133].

Initially, the ECCD experiments were conducted with tandem 2-APs at the -1 and -2 positions, which are the two terminal nucleotides in the hFEN:product complex. Earlier work has shown that the position of the -1 nucleotide is altering relative to the -2 nucleotide and unstacked when bound with human FEN1 in the presence of Ca^{2+} ions[9,133]. Further investigation at +1 and -1 nucleotides positions have been done to understand the relative positioning of these two nucleotides. These studies were performed with a static double flap comprised of a displaced 5 nts 5' flap and a 1 nt 3' flap containing 2-AP dimers (downstream of substrate) at +1-1 and -1-2 relative to the scissile phosphate. The numbering refers the position of the 2-AP dimer position as shown in **Figure 6.3** and **Figure 6.4**. The static substrate was important to restrict the nucleic acid structure to only one conformer. The bimolecular DNA substrate were constructed from a 2AP containing single stranded flap strand and a template strand that also provided the upstream duplex. Due to low sensitivity, the substrate concentration is kept at 10 μM, and at these concentrations a shorter downstream duplex is sufficient to maintain stability of the construct.

The aim of this chapter to study the specific conformational changes of the substrate at ss/dsDNA junction with mutated residues of human FEN1. In this study, we used two types of oligonucleotides, FEC1 annealed with Temp-1 and FEC2 annealed with Temp-2 to form the static double strand DNA substrate construct $dsFEC1_{+1-1}$ and $dsFEC2_{-1-2}$ respectively (the subscript numbers denote the place of 2AP substitution). The static substrates were used to allow the measurement by ECCD to be precise at the specific position of 2-AP dimer within DNA and also to eliminate flap migration.

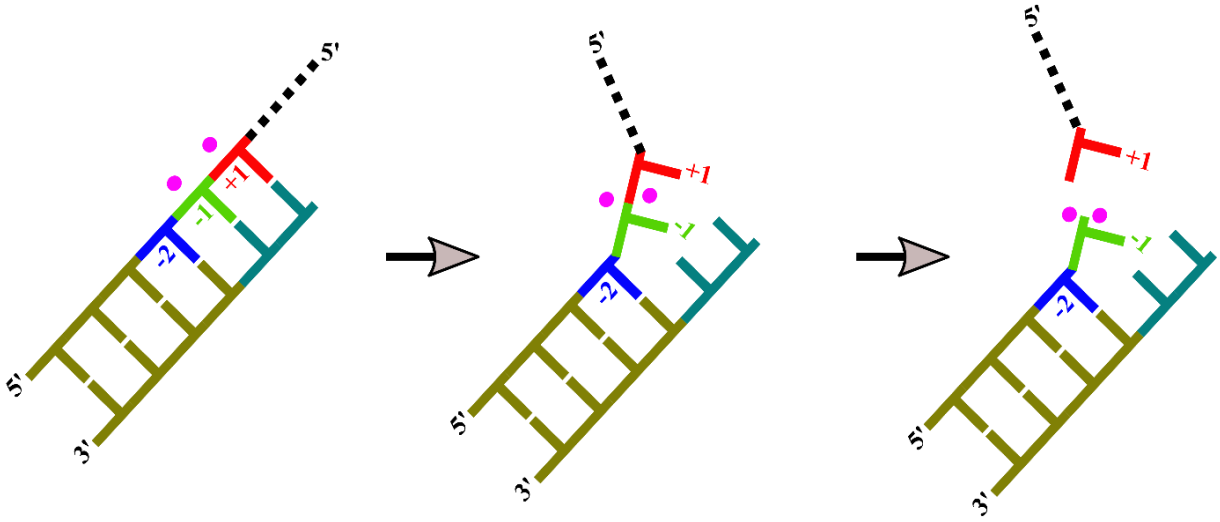

Figure 6.3: *2-aminopurine positions on static dsFEC1$_{+1-1}$ or dsFEC2$_{-1-2}$ of downstream DNA substrate.*

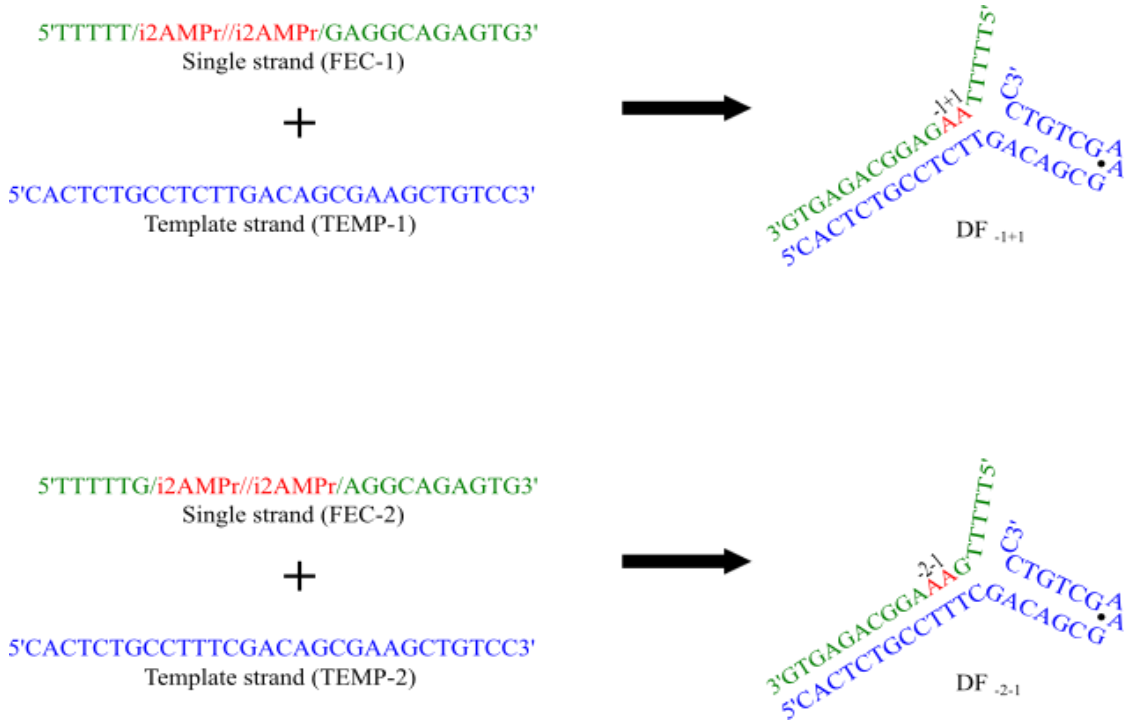

Figure 6.4: *A schematic representation of the oligonucleotides used to study the conformational changes of hFEN1, R103A, K125A and R129A with DNA substrates. Static double flaps contain substitutions at -1+1 and -1-2 (numbering relative to cleavage site) and AA represents the position of the 2-aminopurine nucleotides.*

6.3 ECCD spectra of free static double strand FEC1$_{+1-1}$ and FEC1$_{-1-2}$ substrates

The replacement of the nucleotides on 5' flap strand at +1-1 or -1-2 with a 2-AP dimer in dsFEC1$_{+1-1}$ and dsFEC1$_{-1-2}$ respectively produced a unique ECCD signal with a maxima close to 326 nm. The +1-1 and -1-2 2AP DNAs were analysed as free ssDNAs and dsDNAs and also bound with the each of the mutants and wild type human FEN1. Although the ECCD signals were recorded from 300 – 480 nm to have a large baseline for spectral normalization, only spectra between 300 – 360 nm were presented (no absorbance was detected above 360 nm).

To begin, studies were performed with ssDNA and dsDNA in the presence and absence of divalent metal ions without enzyme **(Figure 6.5)**. The results show, the intensity of dsFEC1$_{+1-1}$ and dsFEC2$_{-1-2}$ was approximately 2-fold higher than ssFEC1$_{+1-1}$ and ssFEC2$_{-1-2}$ DNA oligonucleotides. These observations indicated that the 2-AP nucleobase are restricted to fewer conformations in dsDNA compared to ssDNA. This spectroscopic behaviour also suggests that in the duplex environment, there are greater stacking interactions of the two spectral probes. The weaker electronic transitions in the ssDNA are the result of reduced the interaction between the two adjacent 2-APs. On the other hand, the FEC1$_{+1-1}$ and FEC2$_{-1-2}$ DNA substrates showed the same intensity at 326 nm under Ca^{2+} and EDTA conditions. This implies similar interactions occur in the presence and the absence of divalent metal ions without FEN1 proteins present.

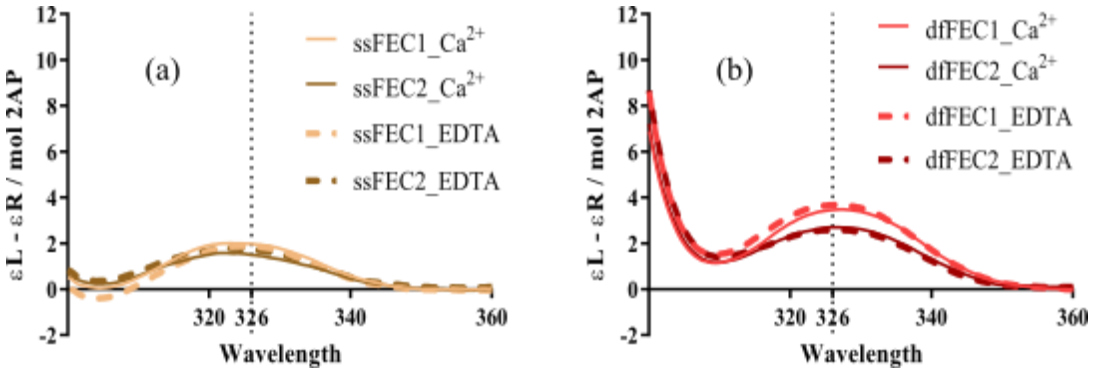

***Figure 6.5**: ECCD spectra of free single strands (a) and double strands (b) of DNA substrates containing 2-AP dimer. FEC1 and FEC2 represent +1-1 and -1-2 positions of 2-AP at the downstream substrate respectively. The ECCD traces in Ca^{2+} are represented as a solid line, while in EDTA represented as dashed lines. The experiments were done at 20 ^{0}C, n = 2, time per point 0.4 s.*

6.4 ECCD spectra of static dsFEC1$_{+1-1}$:hFEN1 and dsFEC1$_{-1-2}$:hFEN1 complexes

As shown in **Figure 6.6**, for both dsFEC1$_{+1-1}$ and dsFEC2$_{-1-2}$, a reduction in ECCD intensity was related with binding of human FEN1 in the presence of Ca^{2+} ions. The reduction in energy of the exciton-coupling at 326 nm was similar to that obtained in a previous study by Finger et. *al*., 2013, Patel et. *al*., 2013 and Algasaier et *al*., 2016. This indicated the ability of hFEN1:Ca^{2+} to distort the DNA substrate. However, the addition of hFEN1:EDTA to dsFEC1$_{+1-1}$ and dsFEC2$_{-1-2}$ caused the ECCD signal to be higher than that with hFEN1:Ca^{2+}. These results parallel those of a previous study showing that the 2-APs remain stacked in the double-stranded DNA substrates when bound to FEN1 in the absence of divalent metal ions[75,128,133]. These results demonstrate that the presence of divalent metal ions within the active site is a pre-requisite for the substrate to distort and transfer the scissile phosphate to the active site. In addition, this data also proved that DNA conformational changes occurred when induced by the enzyme.

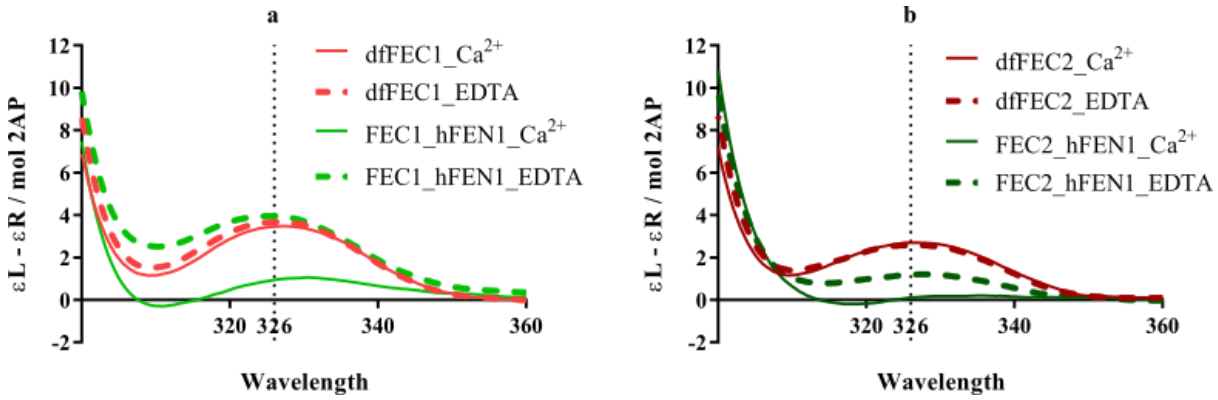

***Figure 6.6**: The comparative ECCD spectra between free dsDNA and dsDNA:hFEN1. Two different positions of 2-AP, (a) FEC1 (+1-1) and (b) FEC2 (-1-2).*

6.5 ECCD spectra of static dsFEC1$_{+1-1}$ and dsFEC2$_{-1-2}$ substrates and mutated FEN1 complexes

To understand the importance and requirements of the ideal local conformational changes, the effects of helical arch mutations R103A, R104A, K125A, R129A, A107G, R104AK132A and A98GA107G were studied (**Figure 6.7**). The results were compared to wild type human FEN1 to indicate any differences, thereby identifying important residues to facilitate DNA conformational changes. All mutants were expressed, purified and kinetically characterised by Dr Mark Thompson (member of the Grasby group) except K125A. The mutation of basic residues arginine (R) and lysine (K) were designed to study the electrostatic interaction between helical arch and phosphate diesters of the threaded 5'-flap DNA substrate. Arg-103, Arg-104 and Lys-132 were basic residues proposed to be responsible to pilot the 5'-flap phosphate diester backbone through the helical arch and to play a role in stabilizing the ssDNA orientation[126]. The α5 residues Lys-125 and Arg-129 appear positioned to interact with the DNA template strand especially when the DNA is active site positioned and the arch has moved forward as discussed in chapter 5. Whereas mutations Ala-107 to glycine could inhibit structuring of α4, while double mutant A98GA107G was expected to impair structuring of enzyme. Generally, all mutations were done to investigate the impact of decreased helical propensity to formation of the helical arch structure.

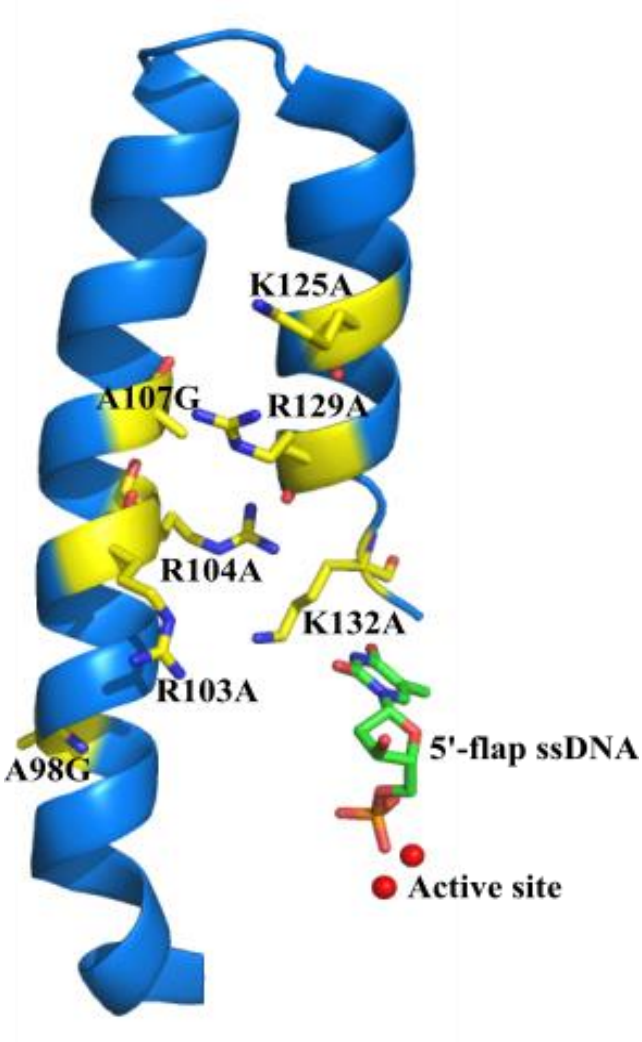

Figure 6.7: *Cartoon represents the position of mutated residues on the helical arch.*

The single turnover rates of reaction of R104A, R107G, A98GA107G and R104AK132A were measured by Dr Mark Thompson, whereas the result for K125A were described in the preceding chapter. These experiments used the static double flap SB5,1. **Table 6.1** represents the maximal single turnover (k_{st}) rates of some of the mutants. The back of arch mutation R104A and the corresponding double mutants R104AK132A decreased the rate of reaction by approximately 44- and 227-fold. The A107G mutation slowed the rate of reaction 16-fold and the double mutant A98GA107G decreased the rate by 105-fold compared with WThFEN1 underscoring the importance of the arch to undergo a disorder to order transition. As discussed previously K125A decreased the rate by 14-fold. Previously, the multiple turnover rates of reaction of R103A and R129A data were determined by Tsutakawa et *al*., 2017 and the $v_0/[E]_0$ (min^{-1}) rate is reduced 3-fold for R103A and 20-fold for R129A relative to WThFEN1.

Table 6.1: *The single turnover rate and the rate decrease by fold for each mutant relative to WT human FEN1. The experiments were done using static double flap DNA substrates.*

Enzyme	k_{st}, min^{-1}	Rate decrease by fold
Human FEN1	722.5 ± 58.8	-
R104A	16.5 ± 0.4	44
A107G	44.0 ± 2.8	16
K125A	51 ± 2	14
A98GA107G	6.9 ± 1.12	105
R104AK132A	3.2 ± 0.12	227

To monitor the ability of mutant proteins to bring about DNA conformational change ECCD measurements of the protein-DNA complexes were performed in the presence of divalent metal ions or EDTA using $dsFEC1_{+1-1}$ and $dsFEC2_{-1-2}$ substrates. **Figure 6.8** shows the CD spectrum at 326 nm for five mutants comparatively with wild type human FEN1. **Figure 6.9** displays the ellipticity changes at 326 nm for seven mutants (R104A and R104AK132A data were from Dr. Mark Thompson). Because the absolute magnitude of the data can vary on a day to day basis, we normalised all data with dfFEC2 measured at a similar time in every case. **Figure 6.9 a** and **c** represent the original data before normalisation, whereas **Figure 6.9 b** and **d** show the data after normalisation with dfFEC2 under +1-1 and -1-2 conditions.

ECCD signals for wild type human FEN1 and all mutant complexes were higher under EDTA compared to Ca^{2+} conditions. As observed previously when the hFEN1:$dsFEC1_{+1-1}$ complex was formed in the presence of Ca^{2+} ions the ECCD signal was decreased with respect to the free substrate and the protein-DNA complex in EDTA[128]. This indicates that the +1 and -1 nucleotides change their relative positions to some extent when the substrate is positioned within the active site. Compared with hFEN1:$dsFEC1_{+1-1}$ complex, the seven mutants affected the ECCD traces of their complexes with $dsFEC1_{+1-1}$ substrate. R103A:Ca^{2+} and R103A:EDTA complexes displayed comparable magnitude to the ECCD signal of wild type human FEN1 with $dsFEC1_{+1-1}$ and $dsFEC2_{-1-2}$ substrates (**Figure 6.8 a, b** and **Figure 6.9 a, b, c, d**). The equivalence in exciton-coupling of R103A and wild type human FEN1 in all conditions suggests that the positions of the nucleotide are similar in both enzyme-substrate complexes. The identical ECCD spectrum for R103A-

$(2AP)_2$DNAs in comparison with wild type human FEN1 complexes suggested Arg-103 does not important for orientating of the nucleotides flanking the scissile phosphate of DNA. These results are consistent with enzyme activities in multiple turnover experiments, where the mutant R103A is only reduce by 3-fold in comparison wild type to human FEN1. Although Arg-103 is semi-conserved, it does not show any significance impact on the WThFEN1 activity.

The altered protein-DNA complexes R104A:dsFEC1$_{+1-1}$ (**Figure 6.9 a, b**). and A107G:dsFEC1$_{+1-1}$ (**Figure 6.8 c** and **Figure 6.9 a, b)** complexes produced modestly increased ECCD signals relative to the wild type protein in Ca^{2+} ions. This indicated that the relative orientations of +1 nucleotide with the -1 nucleotide was only modestly altered regardless being bound to R104A or A107G which have Ca^{2+} ions bound in the active site. Another four mutant complexes, K125A:dsFEC1$_{+1-1}$ (**Figure 6.8 e** and **Figure 6.9 a, b**), R129A:dsFEC1$_{+1-1}$ (**Figure 6.8 g** and **Figure 6.9 a, b**), R104AK132A:dsFEC1$_{+1-1}$ (**Figure 6.9 a, b**) and A98GA107G:dsFEC1$_{+1-1}$ (**Figure 6.8 i** and **Figure 6.9 a, b**) produced approximately 4-fold increases in the magnitude of the CD spectrum than the comparable wild type human FEN1 complex (**Figure 6.8 e, g, i and Figure 6.9 a, b**) suggesting that these mutants cannot induce the same magnitude of conformational change in the substrate or alters the partitioning of to change the equilibrium position. Further experiments were performed by adding EDTA and overall the calculated ECCD intensity in Ca^{2+} increased slightly at 326 nm for all mutants to approach the human FEN1 signal (**Figure 6.9 a, b**).

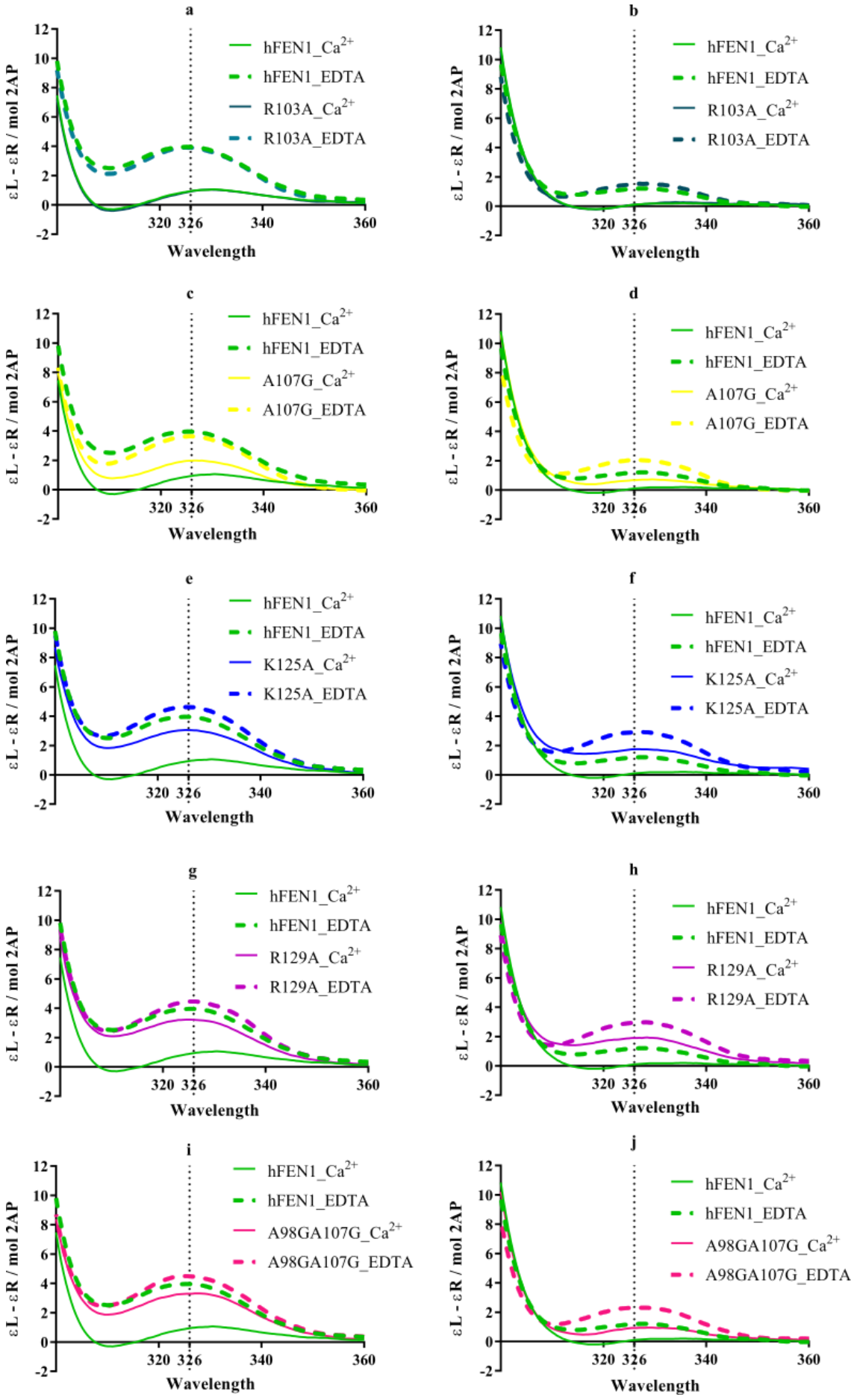
a
hFEN1_Ca2+
hFEN1_EDTA
R103A_Ca2+
R103A_EDTA
εL - εR / mol 2AP
Wavelength
b
hFEN1_Ca2+
hFEN1_EDTA
R103A_Ca2+
R103A_EDTA
εL - εR / mol 2AP
Wavelength
c
hFEN1_Ca2+
hFEN1_EDTA
A107G_Ca2+
A107G_EDTA
εL - εR / mol 2AP
Wavelength
d
hFEN1_Ca2+
hFEN1_EDTA
A107G_Ca2+
A107G_EDTA
εL - εR / mol 2AP
Wavelength
e
hFEN1_Ca2+
hFEN1_EDTA
K125A_Ca2+
K125A_EDTA
εL - εR / mol 2AP
Wavelength
f
hFEN1_Ca2+
hFEN1_EDTA
K125A_Ca2+
K125A_EDTA
εL - εR / mol 2AP
Wavelength
g
hFEN1_Ca2+
hFEN1_EDTA
R129A_Ca2+
R129A_EDTA
εL - εR / mol 2AP
Wavelength
h
hFEN1_Ca2+
hFEN1_EDTA
R129A_Ca2+
R129A_EDTA
εL - εR / mol 2AP
Wavelength
i
hFEN1_Ca2+
hFEN1_EDTA
A98GA107G_Ca2+
A98GA107G_EDTA
εL - εR / mol 2AP
Wavelength
j
hFEN1_Ca2+
hFEN1_EDTA
A98GA107G_Ca2+
A98GA107G_EDTA
εL - εR / mol 2AP
Wavelength

Figure 6.8: *ECCD spectra of dsFEC1$_{+1-1}$ and dsFEC2$_{-1-2}$ containing of (2-AP)$_2$. a, c, e, g and i represent the ECCD traces of R103A, A107G, K125A, R129A and A98GA107G for dsFEC1$_{+1-1}$ respectively, whereas b, d, f, h and j represent the ECCD signal of R103A, A107G, K125A, R129A and A98GA107G for dsFEC2$_{-1-2}$ respectively. Solid lines displayed the ECCD spectrum in Ca^{2+}, while dashed lines in EDTA. All mutants were plotted with human FEN1 (green solid and dashed lines).*

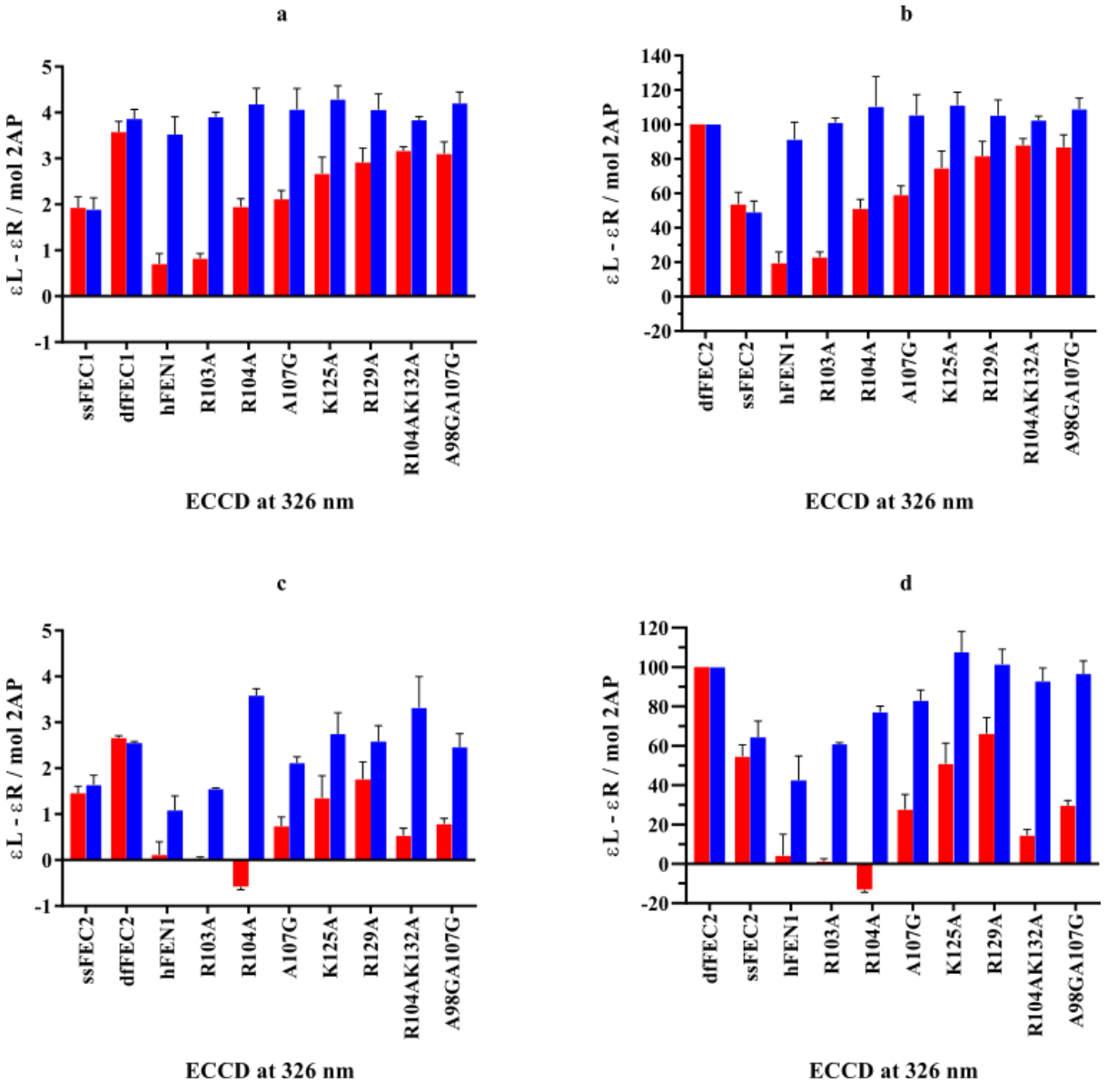

Figure 6.9: *Ellipticity changes at 326 nm for free substrates, and complexes with human FEN1 and all mutant enzymes. Red and blue represent the experiments in Ca^{2+} and EDTA respectively. **a** and **b** displayed the bar chart for grouped data and normalised grouped data for dsFEC1$_{+1-1}$ substrates, whereas **c** and **d** showed the ECCD results at 326 nm for grouped data and normalised grouped data for dsFEC2$_{-1-2}$. All data sets were collected from three replicates and standard errors are shown.*

Studies have also been carried out with 2-AP at positions -1 and -2 using dsFEC2$_{-1-2}$ DNA substrate. As shown previously, in the presence of Ca^{2+} ions the ECCD signal associated with hFEN1:dsFEC2$_{-1-2}$ is near zero at 326 nm indicative of considerable change of relative positioning of the 2-AP in the active site. Similar to hFEN1:dsFEC1$_{+1-1}$ complexes, further variations were observed with the mutant enzymes:dsFEC2$_{-1-2}$ complexes compared to the corresponding wild type hFEN1:dsFEC2$_{-1-2}$ complex. The signal at 326 nm associated with R104A decreases 3-fold further than human FEN1 in the presence of Ca^{2+} ions (**Figure 6.9 c, d**). This probably reflects a different orientation of the 2-AP nucleobases at -1 and -2 positions or/and maybe a difference in a partition between active site and non-active site positioned forms in the R104A:dsFEC2$_{-1-2}$ complexes. Interestingly, the A98GA107G:Ca^{2+} complex produced a change in the exciton pair signal (**Figure 6.8 j** and **Figure 6.9 c, d**), which showed that the dsFEC2$_{-1-2}$ substrate had undergone a comparable conformational change to that when bound to A107G:Ca^{2+} **(Figure 6.8 d** and **Figure 6.9 c, d**). This implied that when bound to A98GA107G the -1 and -2 nucleotides were in the same position as with A107G but they failed to reproduce the positioning observed with the wild type enzyme. In Ca^{2+}, more variations of ECCD spectrum for α5 single mutants K125A:dsFEC2$_{-1-2}$ (**Figure 6.8 f**) and R129A:dsFEC2$_{-1-2}$ (**Figure 6.8 h**) complexes were observed with 11.3-fold and 15.8-fold higher signals respectively (**Figure 6.9 c, d**). A more modest increase was observed for the double mutant R104AK132A:dsFEC2$_{-1-2}$, which was increased by 3.5-fold relative to hFEN1:dsFEC2$_{-1-2}$ complex (**Figure 6.9 c, d)**. This suggests that the adjacent 2-APs at -1 and -2 do not adopt the same position as in the WT protein or the partition between forms is adjusted. All mutants R103A, R104A, A107G, K125A, R129A, R104AK132A and A98GA107G complexes produced slightly increased between 1.4-fold to 2.5-fold of ECCD signal by adding EDTA to approach the corresponding human FEN1 complex signal (**Figure 6.9 c, d**).

6.6 Discussion and summary

The ECCD experiments in this chapter were carried out to understand the active site positioning mechanism of seven mutant enzymes compared to human FEN1 using a static double flap DNA substrate. By inserting two tandem 2-aminopurines in substrates at specific positions, the ability of the mutant enzymes to bring about conformational changes of nucleotides within the enzyme-substrate complex could be monitored. The residues show differences in the roles played in conformational change of DNA substrate according to their positions within the helical arch. For example, Arg-103 at in front of α4 behaved similarly to WThFEN1. This indicates that Arg-103 does not play a crucial role in positioning the DNA nucleotides for reaction. In contrast Arg-104 and Lys-132, both positioned at the back of the arch, displayed a modest effect when tested at both positions (+1-1 and -1-2 condition). Based on these data, we can propose that these residues are important to rearrange and locate the nucleotides towards the active site. Similarly, when Ala-98 and Ala-107 (top and bottom of α4) were mutated to glycine to decrease helical propensity, they displayed similar effects to the back of arch mutations. This suggested that for precise positioning of the scissile phosphate on the active site metals depends on α4 being able to readily form a α-helix. Additionally, α5 Lys-125 and Arg-129 which are believed to interact with the template DNA are important to transfer the DNA substrate to the active sites.

Regarding 2-APs position on 5'-flap dsDNA substrates at +1-1 and -1-2 positions, all of the seven human FEN1 mutant complexes allows identification of the importance of the active site metal ions for the local conformational changes of the DNA reacting duplex. The ECCD spectrum intensity at 326 nm were increased to some degree in all cases when Ca^{2+} ions were removed upon adding EDTA. These data proved that the active site metal ions are important for the local conformational changes. Without Ca^{2+} at the active site, the nucleotides could not produce the ideal conformational changes to position the scissile phosphate for hydrolysis. According to each mutant results with dsDNA substrates, some differences of ECCD signal from the corresponding human FEN1 complexes were observed and summarized in **Table 6.2**.

Table 6.2: *The summary of electronic interaction for two adjacent 2APs of +1-1 and -1-2 positions of the seven mutated human FEN1 complexes comparing with WThFEN1. The symbols; (=) is similar interaction, (-) is less interaction and (+), (++), (+++) are higher interaction with different level of increasing the CD signal. The data were summarised from* ***Figure 6.8****.*

DNA Substrate	2-APs position	R103A	R104A	A107G	K125A	R129A	R104A K132A	A98G A107G
dsFEC1	+1-1	=	+	+	++	++	+++	+++
dsFEC2	-1-2	=	-	++	+++	+++	+	++

7. Conclusion

Flap endonucleases (FENs) catalyse the essential removal of single-stranded 5'-DNA or RNA protrusions known as flaps that occur during DNA replication and repair. Thus, FEN1 is essential for Okazaki fragment maturation, long-patch base excision repair, ribonucleotide excision repair and is purportedly involved a newly discovered DNA repair process (α-segment editing). When flaps occur *in-vivo* they are migrating structures that can theoretically adopt a number of conformations, but it is reported that FEN1 will only act on one conformer. Using migrating flap structures, this study proved that despite the existence of multiple potential conformations of the substrates the specificity of reaction one nucleotide into duplex with a one nucleotide 3'-flap was still maintained. Studies of FEN1 are generally confined to using static substrates that present DNA in the preferred conformation; that is, with a 5'-flap of any length including zero and a single nt 3'-flap. The results in this study showed that the rate of reaction (k_{cat}/K_M) was inversely proportional to the number of potential conformers. Thus, the presence of additional conformers slows the hFEN1-catalysed reaction, as might be expected if only one of many conformers can undergo reaction. Roughly, the presence of additional conformers slows the hFEN1-catalysed multiple-turnover reaction by 10-fold, as might be expected if only one of many conformers can undergo reaction. Substrate conformational heterogeneity also decreases the single turnover rate of reaction by 10-fold, suggestive of the potential for on-enzyme rearrangement to the preferred conformation. Furthermore, under first and second order kinetic studies revealed that, the catalytic activity of human FEN1 is not affected by the length of 5'-flaps that we have studied, but rather the number of conformers influences to the rate of reaction.

In terms of the data and experiments discussed in chapter 3, there are several minor issues that need to be addressed for completeness. For a better understanding as to how the human FEN1 binds migrating substrate, a measurement with fluorescence resonance energy transfer (FRET) would be helpful. This technique involves a non-radiative shift of energy from an excited state donor fluorophore to a nearby acceptor[134]. The insertion of donor-acceptor pair inside DNA substrate construct could afford dynamic data of any conformation changes under equilibrium state[135]. Besides that, further investigation using substrates of similar sequence except for the junction region will be studied, and FRET could measure the bending effects for each substrate. This will reveal whether stacking

effects alter the proportions of various conformers or whether different sequences around the site of reaction alter the rate.

The interaction between wild type human FEN1 and PCNA was tested to determine whether PCNA stimulation would be an attractive way of regulating hFEN1 when it acts on substrates that can adopt multiple conformers. Flap endonucleases (FENs), along with the PCNA catalyse the essential removal of single-stranded 5'-DNA or RNA protrusions known as flaps that occur during DNA replication and repair. Interestingly in chapter 4, we found that PCNA did not alter or stimulate the reactions of FEN1 with static or migrating double or single flaps under the conditions tested. The addition of clamp loader RFC could also possibly give better understanding how the PCNA interact with the DNA substrate and enzyme proteins. In addition, use of the clamp loader could ensure that PCNA is oriented correctly on the substrate.

The capability of FENs to accommodate the single strand 5'-flap DNA substrate has been studied for some time from diverse groups with a number of mechanisms[9,17,27,72,90]. Dahlberg and colleagues first came up with the theory of FEN1 specificity for removal of ssDNA 5'-flap by passing flaps through a hole in the protein. The work reported in this thesis in chapter 5 has sought to understand the mechanism of threading DNA through the helical arch (α4 and α5). The 5'-flap of DNA was presumed to be like a needle threading via a small hole in wild type human FEN1. We proposed the FENs interact with the double strand DNA first and then thread the single stranded 5'-flap through a disordered arch. By mutation of the conserved residues at helical arch, we revealed not all the residues are important for the threading such as Arg-103. However, α5 Lys-125 and Arg-129 both interact directly with template DNA and were observed to play a role in stabilising the threaded state.

The objective of chapter 6 was to understand the active site transfer mechanism of the FEN1-catalysed reaction. The replacement of nucleotides by 2-aminopurine in a static double flap DNA substrate allowed the conformational changes of nucleotides to be measured by Exciton Coupled Circular Dichroism (ECCD) when bound with FEN1 and mutants. Low-energy CD of $(2\text{-AP})_2$ DNAs is an established technique to understanding the dynamics of nucleotides in different nucleic acid structures[136]. ECCD of 2-AP residues inserted into DNA also can be exploited on a large enzyme-substrate complexes to

investigate the conformation of specific nucleotides[137–139]. In this study the comparison between single strand and double strand spectra and between wild type human FEN1 and mutated human FEN1 were collated. With substitution of 2-AP on two different positions (+1-1 and -1-2) in the DNA substrate, this study demonstrated that, the positions of the residues either in-front of the arch or at the back will contribute and indicate how important each residue is to placing the scissile phosphate near or close to active site. The ECCD signals suggested that each residue (except Arg-103) participated in the active site transfer mechanism. In addition, the existence of divalent metal ions at the active site are required for binding between WThFEN1 and active site positioned nucleotides of the DNA substrate.

It has been noted for some time that the helical arch above the active site and the α2- α3 loop appear unstructured in the absence of DNA substrate, as recently demonstrated by NMR studies[74]. Moreover, a recent publication shows that, at least in part, recognition of 3'-flaps is a property of the enzyme-substrate complex with the absence of a 3'-flap having large impacts on the maximal single turnover rate of reaction[93]. Thus communication between the 3'-flap binding pocket and the active site must occur. The juxtaposition of helical arch α5 and the α2- α3 loop, which pack against one another when structured, suggests a potential mechanism for this where the substrate induced ordering of one of these regions might impact on the other. The work presented here suggests that this would occur by controlling the ability of substrate to transfer to the active site, as key residues controlling this process reside in α5.

References

(1) Watson, J. D.; Crick, F. H. Molecular Structure of Nucleic Acids; a Structure for Deoxyribose Nucleic Acid. *Nature* **1953**, *171*, 737–738.

(2) Zhang, T. B.; Zhang, C. L.; Dong, Z. L.; Guan, Y. F. Determination of Base Binding Strength and Base Stacking Interaction of DNA Duplex Using Atomic Force Microscope. *Sci. Rep.* **2015**, *5*, 1–7.

(3) lumenlearning. Structure and Function of DNA | Microbiology. *Web Page*.

(4) Alberts, B.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P. *Molecular Biology of the Cell*; 2002.

(5) Griffiths, A.; Miller, J.; Suzuki, D. *Introduction to Genetic Analysis. 7th Edition.*; 2000.

(6) Grasby, J. A.; Finger, L. D.; Tsutakawa, S. E.; Atack, J. M.; Tainer, J. A. Unpairing and Gating: Sequence-Independent Substrate Recognition by FEN Superfamily Nucleases. *Trends Biochem. Sci.* **2012**, *37* (2), 74–84.

(7) Tomlinson, C. G.; Atack, J. M.; Chapados, B.; Tainer, J. A.; Grasby, J. A. Substrate Recognition and Catalysis by Flap Endonucleases and Related Enzymes. *Biochem Soc Trans* **2010**, *38* (2), 433–437.

(8) Lieber, M. R. The FEN-1 Family of Structure-Specific Nucleases in Eukaryotic DNA Replication, Recombination and Repair. *BioEssays* **1969**, *19* (3), 233–240.

(9) Tsutakawa, S. E.; Classen, S.; Chapados, B. R.; Arvai, A. S.; Finger, L. D.; Guenther, G.; Tomlinson, C. G.; Thompson, P.; Sarker, A. H.; Shen, B.; et al. Human Flap Endonuclease Structures, DNA Double-Base Flipping, and a Unified Understanding of the FEN1 Superfamily. *Cell* **2011**, *145* (2), 198–211.

(10) Schroeder, G. K.; Lad, C.; Wyman, P.; Williams, N. H.; Wolfenden, R. The Time Required for Water Attack at the Phosphorus Atom of Simple Phosphodiesters and of DNA. *Proc. Natl. Acad. Sci.* **2006**, *103* (11), 4052–4055.

(11) Rossi, M. L.; Purohit, V.; Brandt, P. D.; Bambara, R. A. Lagging Strand Replication Proteins in Genome Stability and DNA Repair. *Chem. Rev.* **2006**, *106* (2), 453–473.

(12) Yang, W. *Nucleases: Diversity of Structure, Function and Mechanism*; 2011; Vol. 44.

(13) Finger, L. D.; Blanchard, M. S.; Theimer, C. A.; Sengerová, B.; Singh, P.; Chavez, V.; Liu, F.; Grasby, J. A.; Shen, B. The 3′-Flap Pocket of Human Flap

Endonuclease 1 Is Critical for Substrate Binding and Catalysis. *J. Biol. Chem.* **2009**, *284* (33), 22184–22194.

(14) Mitsunobu, H.; Zhu, B.; Lee, S.-J.; Tabor, S.; Richardson, C. C. Flap Endonuclease of Bacteriophage T7: Possible Roles in RNA Primer Removal, Recombination and Host DNA Breakdown. *Bacteriophage* **2014**, *4*, e28507.

(15) Friedberg, E. C.; Goldthwait, D. A. Endonuclease II of E. Coli. I. Isolation and Purification. *Proc. Natl. Acad. Sci. U. S. A.* **1969**, *62* (3), 934–940.

(16) Lyamichev, V.; Brow, M. A. D.; Dahlberg, J. E. Structure-Specific Endonucleolytic Cleavage of Nucleic-Acids by Eubacterial DNA-Polymerases. *Science (80-.).* **1993**, *260* (5109), 778–783.

(17) Orans, J.; A McSweeney, E.; R. Iyer, R.; A. Hast, M.; W. Hellonga, H.; Modrich, P.; S. Beese, L. Structures of Human Exonuclease 1 DNA Complexes Suggest a Unified Mechanism for Nuclease Family. *Cell* **2011**, *15* (145(2)), 212–223.

(18) Waga, S.; Bauer, G.; Stillman, B. Reconstitution of Complete SV40 DNA Replication with Purified Replication Factors. *J. Biol. Chem.* **1994**, *269* (14), 10923–10934.

(19) Beese, L. S.; Steitz, T. A. Structural Basis for the 3 ’ 5 ’ Exonuclease Activity of Escherichia Coli DNA Polymerase 1 : A Two Metal Ion Mechanism. *EMBO J.* **1991**, *10* (1), 25–33.

(20) Ayyagari, R.; Gomes, X. V; Gordenin, D. A.; Burgers, P. M. J. Okazaki Fragment Maturation in Yeast: I. Distribution of Functions between FEN1 and DNA2. *J. Biol. Chem.* **2003**, *278* (3), 1618–1625.

(21) Jaiswal, A. S.; Balusu, R.; Armas, M. L.; Kundu, C. N.; Narayan, S. Mechanism of Adenomatous Polyposis Coli (APC)-Mediated Blockage of Longpatch Base Excision Repair. In *NIH Public Access*; 2006; Vol. 45, pp 15903–15914.

(22) Chapados, B. R.; Hosfield, D. J.; Han, S.; Qiu, J.; Yelent, B.; Shen, B.; Tainer, J. A. Structural Basis for FEN-1 Substrate Specificity and PCNA-Mediated Activation in DNA Replication and Repair. *Cell* **2004**, *116* (1), 39–50.

(23) Sakurai, S.; Kitano, K.; Yamaguchi, H.; Hamada, K.; Okada, K.; Fukuda, K.; Uchida, M.; Ohtsuka, E.; Morioka, H.; Hakoshima, T. Structural Basis for Recruitment of Human Flap Endonuclease 1 to PCNA. *EMBO J.* **2005**, *24* (4), 683–693.

(24) Kim, Y.; Eom, S. H.; Wang, J.; Lee, D.-S.; et al. Crystal Structure of Thermus Aquaticus DNA Polymerase. *Nature* **1995**, *376* (6541), 612–616.

(25) Harrington, J. J.; Lieber, M. R. The Characterization of a Mammalian DNA Structure-Specific Endonuclease. *EMBO J.* **1994**, *13* (5), 1235–1246.

(26) Ceska, T. A.; Sayers, J. R.; Stier, G.; Suck, D. A Helical Arch Allowing Single-Stranded DNA to Thread through T5 5'- Exonuclease. *Nature* **1996**, *382*, 90–93.

(27) Devos, J. M.; Tomanicek, S. J.; Jones, C. E.; Nossal, N. G.; Mueser, T. C. Crystal Structure of Bacteriophage T4 5′ Nuclease in Complex with a Branched DNA Reveals How Flap Endonuclease-1 Family Nucleases Bind Their Substrates. *J. Biol. Chem.* **2007**, *282* (43), 31713–31724.

(28) Mueser, T. C.; Nossal, N. G.; Hyde, C. C. Structure of Bacteriophage T4 RNase H, a 5' to 3' RNA-DNA and DNA-DNA Exonuclease with Sequence Similarity to the RAD2 Family of Eukaryotic Proteins. *Cell*. 1996, pp 1101–1112.

(29) Hwang, K. Y.; Baek, K.; Kim, H. Y.; Cho, Y. The Crystal Structure of Flap Endonuclease-1 from Methanococcus Jannaschii. *Nat. Struct. Mol. Biol.* **1998**, *5* (8), 707–713.

(30) Hosfield, D. J.; Mol, C. D.; Shen, B.; Tainer, J. A. Structure of the DNA Repair and Replication Endonuclease and Exonuclease FEN-1. *Cell* **1998**, *95* (1), 135–146.

(31) Spiro, C.; Pelletier, R.; Rolfsmeier, M. L.; Dixon, M. J.; Lahue, R. S.; Gupta, G.; Park, M. S.; Chen, X.; Mariappan, S. V. S.; McMurray, C. T. Inhibition of FEN-1 Processing by DNA Secondary Structure at Trinucleotide Repeats. *Mol. Cell* **1999**, *4* (6), 1079–1085.

(32) Lieber, M. R. The FEN-1 Family of Structure-Specific Nucleases in Eukaryotic DNA Replication, Recombination and Repair. *BioEssays* **1997**, *19* (3), 233–240.

(33) Shen, B.; Singh, P.; Liu, R.; Qiu, J.; Zheng, L.; Finger, L. D.; Alas, S. Multiple but Dissectible Functions of FEN-1 Nucleases in Nucleic Acid Processing, Genome Stability and Diseases. *BioEssays* **2005**, *27* (7), 717–729.

(34) Burgers, P. M. J. Polymerase Dynamics at the Eukaryotic DNA Replication Fork. *Journal of Biological Chemistry*. 2009, pp 4041–4045.

(35) Balakrishnan, L.; Bambara, R. A. Flap Endonuclease 1. *Annu. Rev. Biochem.* **2013**, *82*, 119–138.

(36) Bambara, R. A.; Murante, R. S.; Henricksen, L. A. Enzymes and Reactions at the Eukaryotic DNA Replication Fork. *J. Biol. Chem.* **1997**, *272* (8), 4647–4650.

(37) Finger, L. D.; Atack, J. M.; Tsutakawa, S.; Classen, S.; Tainer, J.; Grasby, J.; Shen, B. The Wonders of Flap Endonucleases: Structure, Function, Mechanism

and Regulation. In *Eukaryotic Replisome A Guide to Protein Structure Function*; 2012; Vol. 62, pp 301–326.

(38) Beddows, A.; Patel, N.; Finger, L. D.; Atack, J. M.; Williams, D. M.; Grasby, J. A. Interstrand Disulfide Crosslinking of DNA Bases Supports a Double Nucleotide Unpairing Mechanism for Flap Endonucleases. *Chem. Commun.* **2012**, *48*, 8895–8897.

(39) Ayyagari, R.; Gomes, X. V.; Gordenin, D. A.; Burgers, P. M. J. Okazaki Fragment Maturation in Yeast: I. Distribution of Functions between FEN1 and DNA2. *J. Biol. Chem.* **2003**, *278* (3), 1618–1625.

(40) Zaher, M. S.; Rashid, F.; Song, B.; Joudeh, L. I.; Sobhy, M. A.; Tehseen, M.; Hingorani, M. M.; Hamdan, S. M. Missed Cleavage Opportunities by FEN1 Lead to Okazaki Fragment Maturation via the Long-Flap Pathway. *Nucleic Acids Res.* **2018**, *46* (6), 2956–2974.

(41) Liu, Y.; Kao, H.-I.; Bambara, R. A. Flap Endonuclease 1: A Central Component of DNA Metabolism. *Annu. Rev. Biochem.* **2004**, *73*, 589–615.

(42) Krokan, H. E.; Bjoras, M. Base Excision Repair. In *Advances in Protein Chemistry*; 2013; pp 1–41.

(43) Liu, Y.; Wilson, S. H. DNA Base Excision Repair : A Mechanism of Trinucleotide Repeat Expansion. *Trends Biochem. Sci.* **2012**, *37* (4), 162–172.

(44) Dianov, G.; Lindahl, T. Reconstitution of the DNA Base Excision-Repair Pathway. *Curr. Biol.* **1994**, *4* (12), 1069–1076.

(45) Schermerhorn, K. M.; Delaney, S. A Chemical and Kinetic Perspective on Base Excision Repair of DNA. *Acc. Chem. Res.* **2014**, *47* (4), 1238–1246.

(46) Tom, S.; Henricksen, L. A.; Bambara, R. A. Mechanism Whereby Proliferating Cell Nuclear Antigen Stimulates Flap Endonuclease 1. *J. Biol. Chem.* **2000**, *(14) 275* (April 7), 10498–10505.

(47) Kao, H. I.; Veeraraghavan, J.; Polaczek, P.; Campbell, J. L.; Bambara, R. A. On the Roles of Saccharomyces Cerevisiae Dna2p and Flap Endonuclease 1 in Okazaki Fragment Processing. *J. Biol. Chem.* **2004**, *279* (15), 15014–15024.

(48) Sun, H.; He, L.; Wu, H.; Pan, F.; Wu, X.; Zhao, J.; Hu, Z.; Sekhar, C.; Li, H.; Zheng, L.; et al. The FEN1 L209P Mutation Interferes with Long-Patch Base Excision Repair and Induces Cellular Transformation. *Oncogene* **2017**, *36* (2), 194–207.

(49) Nazarkina, J. K.; Petrousseva, I. O.; Safronov, I. V; Lavrik, O. I.; Khodyreva, S.

N. Interaction of Flap Endonuclease 1 and Replication Protein A with Photoreactive Intermediates of DNA Repair. *Biochemistry* **2003**, *68* (8), 934–942.

(50) Brosh, R. M.; Driscoll, H. C.; Dianov, G. L.; Sommers, J. A. Biochemical Characterization of the WRN-FEN-1 Functional Interaction. *Biochemistry* **2002**, *41* (40), 12204–12216.

(51) Brosh, R. M.; Kobbe, C. Von; Sommers, J. A.; Karmakar, P.; Opresko, P. L.; Piotrowski, J.; Dianova, I.; Dianov, G. L.; Bohr, V. A. Werner Syndrome Protein Interacts with Human Ap Endonuclease 1 and Stimulates Its Cleavage Activity. *EMBO J.* **2001**, *20* (20), 5791–5801.

(52) Hasan, S.; Stucki, M.; Hassa, P. O.; Imhof, R.; Gehrig, P.; Hunziker, P.; Hu, U.; Hottiger, M. O. Regulation of Human Flap Endonuclease-1 Activity by Acetylation through the Transcriptional Coactivator P300. *Mol. Cell* **2001**, *7*, 1221–1231.

(53) Hitomi, K.; Iwai, S.; Tainer, J. A. The Intricate Structural Chemistry of Base Excision Repair Machinery: Implications for DNA Damage Recognition, Removal, and Repair. *DNA Repair (Amst).* **2007**, *6* (4), 410–428.

(54) Kucherlapati, M.; Yang, K.; Kuraguchi, M.; Zhao, J.; Lia, M.; Heyer, J.; Kane, M. F.; Fan, K.; Russell, R.; Brown, A. M. C.; et al. Haploinsufficiency of Flap Endonuclease (Fen1) Leads to Rapid Tumor Progression. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (15), 9924–9929.

(55) Matsuzaki, Y.; Adachi, N.; Koyama, H. Vertebrate Cells Lacking FEN-1 Endonuclease Are Viable but Hypersensitive to Methylating Agents and H2O2. *Nucleic Acids Res.* **2002**, *30* (14), 3273–3277.

(56) Sommers, C. H.; Miller, E. J.; Dujon, B.; Prakash, S.; Prakash, L. Conditional Lethality of Null Mutations in RTH That Encodes the Yeast Counterpart of a Mammalian 5’- to 3’-Exonuclease Required for Lagging Strand DNA Synthesis in Reconstituted System. *J. Biol. Chem.* **1995**, *270* (9), 4193–4196.

(57) Gary, R.; Ludwig, D. L.; Cornelius, H. L.; MacInnes, M. A.; Park, M. S. The DNA Repair Endonuclease XPG Binds to Proliferating Cell Nuclear Antigen (PCNA) and Shares Sequence Elements with the PCNA-Binding Regions of FEN-1 and Cyclin-Dependent Kinase Inhibitor P21. *J. Biol. Chem.* **1997**, *272* (39), 24522–24529.

(58) Kunkel, T. A.; Resnick, M. A.; Gordenin, D. A. Mutator Specificity and Disease: Looking over the FENce. *Cell* **1997**, *88* (2), 155–158.

(59) Chai, Q.; Zheng, L.; Zhou, M.; Turchi, J. J.; Shen, B. Interaction and Stimulation of Human FEN-1 Nuclease Activities by Heterogeneous Nuclear Ribonucleoprotein A1 in ??-Segment Processing during Okazaki Fragment Maturation. *Biochemistry* **2003**, *42* (51), 15045–15052.

(60) Burgers, P. M. J. Polymerase Dynamics at the Eukaryotic DNA Replication Fork. *J. Biol. Chem.* **2009**, *284* (7), 4041–4045.

(61) Lee, S. H.; Princz, L. N.; Klugel, M. F.; Habermann, B.; Pfander, B.; Biertumpfel, C. Human Holliday Junction Resolvase GEN1 Uses a Chromodomain for Efficient DNA Recognition and Cleavage. *Elife* **2015**, *4*, 1–24.

(62) Sobhy, M.; Joudeh, L.; Huang, X.; Takahashi, M.; Hamdan, S. Sequential and Multistep Substrate Interrogation Provides the Scaffold for Specificity in Human Flap Endonuclease 1. *Cell Rep.* **2013**, *3* (6), 1785–1794.

(63) Kanai, Y.; Ishikawa, G.; Takeuchi, R.; Ruike, T.; Nakamura, R. I.; Ihara, A.; Ohashi, T.; Takata, K. I.; Kimura, S.; Sakaguchi, K. DmGEN Shows a Flap Endonuclease Activity, Cleaving the Blocked-Flap Structure and Model Replication Fork. *FEBS J.* **2007**, *274* (15), 3914–3927.

(64) Liu, Y.; Freeman, A. D. J.; Déclais, A. C.; Wilson, T. J.; Gartner, A.; Lilley, D. M. J. Crystal Structure of a Eukaryotic GEN1 Resolving Enzyme Bound to DNA. *Cell Rep.* **2015**, *13* (11), 2565–2575.

(65) O'Brien, P. J.; Herschlag, D. Catalytic Promiscuity and the Evolution of New Enzymatic Activities. *Chem. Biol.* **1999**, *6* (4).

(66) Mietus, M.; Nowak, E.; Jaciuk, M.; Kustosz, P.; Studnicka, J.; Nowotny, M. Crystal Structure of the Catalytic Core of Rad2: Insights into the Mechanism of Substrate Binding. *Nucleic Acids Res.* **2014**, *42* (16), 10762–10775.

(67) Hohl, M.; Thorel, F.; Clarkson, S. G.; Schärer, O. D. Structural Determinants for Substrate Binding and Catalysis by the Structure-Specific Endonuclease XPG. *J. Biol. Chem.* **2003**, *278* (21), 19500–19508.

(68) Hohl, M.; Dunand-Sauthier, I.; Staresincic, L.; Jaquier-Gubler, P.; Thorel, F.; Modesti, M.; Clarkson, S. G.; Schärer, O. D. Domain Swapping between FEN-1 and XPG Defines Regions in XPG That Mediate Nucleotide Excision Repair Activity and Substrate Specificity. *Nucleic Acids Res.* **2007**, *35* (9), 3053–3063.

(69) Koch, S. C.; Simon, N.; Ebert, C.; Carell, T. Molecular Mechanisms of Xeroderma Pigmentosum (XP) Proteins. *Q. Rev. Biophys.* **2016**, *49*, 1–32.

(70) Feng, M.; Patel, D.; Dervan, J. J.; Ceska, T.; Suck, D.; Haq, I.; Sayers, J. R. Roles

of Divalent Metal Ions in Flap Endonuclease-Substrate Interactions. *Nat. Struct. Mol. Biol.* **2004**, *11* (5), 450–456.

(71) Devos, J. M.; Tomanicek, S. J.; Jones, C. E.; Nossal, N. G.; Mueser, T. C. Crystal Structure of Bacteriophage T4 5′ Nuclease in Complex with a Branched DNA Reveals How Flap Endonuclease-1 Family Nucleases Bind Their Substrates. *J. Biol. Chem.* **2007**, *282* (43), 31713–31724.

(72) Patel, N.; Atack, J. M.; Finger, L. D.; Exell, J. C.; Thompson, P.; Tsutakawa, S.; Tainer, J. A.; Williams, D. M.; Grasby, J. A. Flap Endonucleases Pass 5′-Flaps through a Flexible Arch Using a Disorder-Thread-Order Mechanism to Confer Specificity for Free 5′-Ends. *Nucleic Acids Res.* **2012**, *40* (10), 4507–4519.

(73) Sengerová, B.; Tomlinson, C.; Atack, J. M.; Williams, R.; Sayers, J. R.; Williams, N. H.; Grasby, J. A. Brønsted Analysis and Rate-Limiting Steps for the T5 Flap Endonuclease Catalyzed Hydrolysis of Exonucleolytic Substrates. *Biochemistry* **2010**, *49*, 8085–8093.

(74) Bennet, I. A.; Finger, L. D.; Baxter, N. J.; Ambrose, B.; Hounslow, A. M.; Thompson, M. J.; Exell, J. C.; Shahari, N. N. B. M.; Craggs, T. D.; Waltho, J. P.; et al. Regional Conformational Flexibility Couples Substrate Specificity and Scissile Phosphate Diester Selectivity in Human Flap Endonuclease 1. *Nucleic Acids Res.* **2018**, *46* (11), 5618–5633.

(75) Patel, N.; Exell, J. C.; Jardine, E.; Ombler, B.; Finger, L. D.; Ciani, B.; Grasby, J. A. Proline Scanning Mutagenesis Reveals a Role for the Flap Endonuclease-1 Helical Cap in Substrate Unpairing. *J. Biol. Chem.* **2013**, *288* (47), 34239–34248.

(76) Tock, M. R.; Frary, E.; Sayers, J. R.; Grasby, J. A. Dynamic Evidence for Metal Ion Catalysis in the Reaction Mediated by a Flap Endonuclease. *EMBO J.* **2003**, *22* (5), 995–1004.

(77) Molina, R.; Stella, S.; Redondo, P.; Gomez, H.; Marcaida, M. J.; Orozco, M.; Prieto, J.; Montoya, G. Visualizing Phosphodiester-Bond Hydrolysis by an Endonuclease. *Nat. Struct. Mol. Biol.* **2015**, *22* (1), 65–72.

(78) Beese, L. S.; Steitz, T. A. Structural Basis for the 3'-5' Exonuclease Activity of Escherichia Coli DNA Polymerase I: A Two Metal Ion Mechanism. *Embo j* **1991**, *10* (1), 25–33.

(79) Kim, C. Y.; Shen, B.; Park, M. S.; Olah, G. A. Structural Changes Measured by X-Ray Scattering from Human Flap Endonuclease-1 Complexed with Mg2+ and Flap DNA Substrate. *J. Biol. Chem.* **1999**, *274* (3), 1233–1239.

(80) Kim, C. Y.; Park, M. S.; Brian Dyer, R. Human Flap Endonuclease-1: Conformational Change upon Binding to the Flap DNA Substrate and Location of the Mg2+ Binding Site. *Biochemistry* **2001**, *40* (10), 3208–3214.

(81) Dervan, J. J.; Feng, M.; Patel, D.; Grasby, J. a; Artymiuk, P. J.; Ceska, T. a; Sayers, J. R. Interactions of Mutant and Wild-Type Flap Endonucleases with Oligonucleotide Substrates Suggest an Alternative Model of DNA Binding. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (13), 8542–8547.

(82) Liu, R.; Qiu, J.; Finger, L. D.; Zheng, L.; Shen, B. The DNA-Protein Interaction Modes of FEN-1 with Gap Substrates and Their Implication in Preventing Duplication Mutations. *Nucleic Acids Res.* **2006**, *34* (6), 1772–1784.

(83) Pelletier, H.; Sawaya, M. R. Characterization of the Metal Ion Binding Helix-Hairpin-Helix Motifs in Human DNA Polymerase β by x-Ray Structural Analysis. *Biochemistry* **1996**, *35* (39), 12778–12787.

(84) Friedrich-Heineken, E.; Hübscher, U. The Fen1 Extrahelical 3′-Flap Pocket Is Conserved from Archaea to Human and Regulates DNA Substrate Specificity. *Nucleic Acids Res.* **2004**, *32* (8), 2520–2528.

(85) Kao, H. I.; Henricksen, L. A.; Liu, Y.; Bambara, R. A. Cleavage Specificity of Saccharomyces Cerevisiae Flap Endonuclease 1 Suggests a Double-Flap Structure as the Cellular Substrate. *J. Biol. Chem.* **2002**, *277* (17), 14379–14389.

(86) Lyamichev, V.; Brow, M. a; Varvel, V. E.; Dahlberg, J. E. Comparison of the 5’ Nuclease Activities of Taq DNA Polymerase and Its Isolated Nuclease Domain. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 6143–6148.

(87) Xu, Y.; Derbyshire, V.; Ng, K.; Chen Sun, X.; Grindley, N. D. f.; Joyce, C. M. Biochemical and Mutational Studies of the 5’-3’ Exonuclease of DNA Polymerase I of Escherichia Coli. *J. Mol. Biol.* **1997**, *268* (2), 284–302.

(88) Williams, R.; Sengerová, B.; Osborne, S.; Syson, K.; Ault, S.; Kilgour, A.; Chapados, B. R.; Tainer, J. A.; Sayers, J. R.; Grasby, J. A. Comparison of the Catalytic Parameters and Reaction Specificities of a Phage and an Archaeal Flap Endonuclease. *J. Mol. Biol.* **2007**, *371* (1), 34–48.

(89) Henricksen, L. A.; Tom, S.; Liu, Y.; Bambara, R. A. Inhibition of Flap Endonuclease 1 by Flap Secondary Structure and Relevance to Repeat Sequence Expansion. *J. Biol. Chem.* **2000**, *275* (22), 16420–16427.

(90) Murante, R. S.; Rust, L.; Bambara, R. A. Calf 5′ to 3′ Exo/Endonuclease Must Slide from a 5′ End of the Substrate to Perform Structure-Specific Cleavage. *J.*

Biol. Chem. **1995**, *270* (51), 30377–30383.

(91) Barnes, C. J.; Wahl, A. F.; Shen, B.; Park, M. S.; Bambara, R. A. 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.* **1996**, *271* (47), 29624–29631.

(92) Bornarth, C. J.; Ranalli, T. A.; Henricksen, L. A.; Wahl, A. F.; Bambara, R. A. Effect of Flap Modifications on Human FEN1 Cleavage. *Biochemistry* **1999**, *38* (40), 13347–13354.

(93) Thompson, M. J.; Gotham, V. J. B.; Ciani, B.; Grasby, J. A. A Conserved Loop–wedge Motif Moderates Reaction Site Search and Recognition by FEN1. *Nucleic Acids Res.* **2018**, No. June, 1–15.

(94) Dieckman, L. M.; Freudenthal, B. D.; Washington, T. M. PCNA Structure and Function: Insights from Structures of PCNA Complexes and Post-Translationally Modified PCNA. In *The Eukaryotic Replisome: a Guide to Protein Structure and Function*; 2012; Vol. 62, pp 281–299.

(95) Hedglin, M.; Benkovic, S. J. Eukaryotic Translesion DNA Synthesis on the Leading and Lagging Strands: Unique Detours Around the Same Obstacle. In *Chem Rev*; 2017; Vol. 117, pp 7857–7877.

(96) Kelman, L. M.; Kelman, Z. Archaeal DNA Replication. *Annu. Rev. Genet.* **2014**, *48* (1), 71–97.

(97) Strzalka, W.; Ziemienowicz, A. Proliferating Cell Nuclear Antigen (PCNA): A Key Factor in DNA Replication and Cell Cycle Regulation. *Ann. Bot.* **2011**, *107* (7), 1127–1140.

(98) Georgescu, R. E.; Kim, S. S.; Yurieva, O.; Kuriyan, J.; Kong, X. P.; O'Donnell, M. Structure of a Sliding Clamp on DNA. *Cell* **2008**, *132* (1), 43–54.

(99) Naryzhny, S. N. Proliferating Cell Nuclear Antigen : A Proteomics View. *Cell. Mol. Life Scineceses* **2008**, *65* (23), 3789–3808.

(100) Yao, N. Y.; Donnell, M. O. The RFC Clamp Loader: Structure and Function. In *The Eukaryotic Replisome: a Guide to Protein Structure and Function*; 2012; Vol. 62, pp 259–279.

(101) Zhang, Z.; Zhang, S.; Lin, S. H. S.; Wang, X.; Wu, L.; Lee, E. Y. C.; Lee, M. Y. W. T. Structure of Monoubiquitinated PCNA: Implications for DNA Polymerase Switching and Okazaki Fragment Maturation. *Cell Cycle* **2012**, *11* (11), 2128–2136.

(102) Doré, A. S.; Kilkenny, M. L.; Jones, S. A.; Oliver, A. W.; Roe, S. M.; Bell, S. D.; Pearl, L. H. Structure of an Archaeal PCNA1-PCNA2-FEN1 Complex: Elucidating PCNA Subunit and Client Enzyme Specificity. *Nucleic Acids Res.* **2006**, *34* (16), 4515–4526.

(103) Zheng, L.; Dai, H.; Qiu, J.; Huang, Q.; Shen, B. Disruption of the FEN-1/PCNA Interaction Results in DNA Replication Defects, Pulmonary Hypoplasia, Pancytopenia, and Newborn Lethality in Mice. *Mol. Cell. Biol.* **2007**, *27* (8), 3176–3186.

(104) Bruning, J. B.; Shamoo, Y. Structural and Thermodynamic Analysis of Human PCNA with Peptides Derived from DNA Polymerase-δ P66 Subunit and Flap Endonuclease-1. *Structure* **2004**, *12* (12), 2209–2219.

(105) Lau, P. J.; Kolodner, R. D. Transfer of the MSH2 ⅐ MSH6 Complex from Proliferating Cell Nuclear Antigen to Mispaired Bases in DNA *. *J. Biol. Chem.* **2003**, *278* (1), 14–17.

(106) Kleczkowska, H. E.; Marra, G.; Lettieri, T.; Jiricny, J. HMSH3 and HMSH6 Interact with PCNA and Colocalize with It to Replication Foci. *Genes Dev.* **2001**, *15*, 724–736.

(107) Clark, A. B.; Valle, F.; Drotschmann, K.; Gary, R. K.; Kunkel, T. A. Proliferating Cell Nuclear Antigen with MSH2-MSH6 and MSH2-MSH3 Complexes. *J. Biol. Chem.* **2000**, *275* (47), 36498–36501.

(108) Flores-rozas, H.; Clark, D.; Kolodner, R. D. Proliferating Cell Nuclear Antigen and Msh2p-Msh6p Interact to Form an Active Mispair Recognition Complex. *Nat. Genet.* **2000**, *26*, 375–378.

(109) Gary, R.; Ludwig, D. L.; Cornelius, H. L.; Macinnes, M. A.; Park, M. S. The DNA Repair Endonuclease XPG Binds to Proliferating Cell Nuclear Antigen (PCNA) and Shares Sequence Elements with the PCNA-Binding Regions of FEN-1 and Cyclin-Dependent Kinase Inhibitor P21 *. *J. Biol. Chem.* **1997**, *272* (39), 24522–24529.

(110) Pascucci, B.; Stucki, M.; Jo, O.; Dogliotti, E.; Hu, U. Long Patch Base Excision Repair with Purified Human Proteins. *J. Biol. Chem.* **1999**, *274* (47), 33696–33702.

(111) Matsumoto, Y.; Kim, K.; Hurwitz, J.; Gary, R.; Levin, D. S.; Tomkinson, A. E.; Park, M. S. Reconstitution of Proliferating Cell Nuclear Antigen-Dependent Repair of Apurinic / Apyrimidinic Sites with Purified Human Proteins. *J. Biol.*

Chem. **1999**, *274* (47), 33703–33708.

(112) Matsumoto, Y.; Kim, K.; Bogenhagen, D. F. Proliferating Cell Nuclear Antigen-Dependent Abasic Site Repair in Xenopus Laevis Oocytes : An Alternative Pathway of Base Excision DNA Repair. *Mech. Ageing Dev.* **1994**, *14* (9), 6187–6197.

(113) Gary, R.; Kim, K.; Cornelius, H. L.; Park, M. S.; Matsumoto, Y. Proliferating Cell Nuclear Antigen Facilitates Excision in Long-Patch Base Excision Repair. *J. Biol. Chem.* **1999**, *274* (7), 4354–4363.

(114) Fortini, P.; Pascucci, B.; Parlanti, E.; Sobol, R. W.; Wilson, S. H.; Dogliotti, E.; Elena, V. R. Accelerated Publications Different DNA Polymerases Are Involved in the Short- and Long-Patch Base Excision Repair in Mammalian Cells †. *Biochemistry* **1998**, *37* (11), 1–6.

(115) Klungland, A.; Lindahl, T. Second Pathway for Completion of Human DNA Base Excision-Repair : Reconstitution with Purified Proteins and Requirement for DNase IV (FEN1). *EMBO J.* **1997**, *16* (11), 3341–3348.

(116) Shivji, M. K. K.; Kenny, M. K.; Wood, R. D. Proliferating Cell Nuclear Antigen Is Required for DNA Excision Repair. *Cell* **1992**, *69*, 367–374.

(117) Aboussekhra, A.; Biggerstaff, M.; Moncollin, V.; Podust, V. N.; Protid, M.; Hqbscher, I. I. U.; Egly, J.; Wood, R. D. Mammalian DNA Nucleotide Excision Repair Reconstituted with Purified Protein Components. *Cell* **1995**, *80*, 859–868.

(118) Li, R.; Hannon, G. J.; Beach, D.; Stillman, B. Subcellular Distribution of P21 and PCNA in Normal and Repair-Deficient Cells Following DNA Damage. *Curr. Biol.* **1996**, *6* (2), 189–199.

(119) Michels, J.; Vitale, I.; Saparbaev, M.; Castedo, M.; Kroemer, G. Predictive Biomarkers for Cancer Therapy with PARP Inhibitors. *Oncogene* **2014**, *33* (30), 3894–3907.

(120) Waters, L. S.; Minesinger, B. K.; Wiltrout, M. E.; Souza, S. D.; Woodruff, R. V; Walker, G. C. Eukaryotic Translesion Polymerases and Their Roles and Regulation in DNA Damage Tolerance. *Microbiol. Mol. Biol. Rev.* **2009**, *73* (1), 134–154.

(121) Exell, J. C.; Thompson, M. J.; Finger, L. D.; Shaw, S. J.; Debreczeni, J.; Ward, T. A.; McWhirter, C.; Siöberg, C. L. B.; Molina, D. M.; Abbott, W. M.; et al. Cellularly Active N-Hydroxyurea FEN1 Inhibitors Block Substrate Entry to the Active Site. *Nature Chemical Biology*. 2016, pp 815–821.

(122) Exell, J. Spectroscopic Evidence for Catalytically- Required FEN1-Mediated DNA Conformational Change ; a Novel Strategy for FEN1 Inhibition, 2015.

(123) Jónsson, Z. O.; Hindges, R.; Hübscher, U. Regulation of DNA Replication and Repair Proteins through Interaction with the Front Side of Proliferating Cell Nuclear Antigen. *EMBO J.* **1998**, *17* (8), 2412–2425.

(124) Craggs, T. D.; Hutton, R. D.; Brenlla, A.; White, M. F.; Penedo, J. C. Single-Molecule Characterization of Fen1 and Fen1/PCNA Complexes Acting on Flap Substrates. *Nucleic Acids Res.* **2014**, *42* (3), 1857–1872.

(125) Beattie, T. R.; Bell, S. D. The Role of the DNA Sliding Clamp in Okazaki Fragment Maturation in Archaea and Eukaryotes. *Biochem. Soc. Trans.* **2011**, *39*, 70–76.

(126) Tsutakawa, S. E.; Thompson, M. J.; Arvai, A. S.; Neil, A. J.; Shaw, S. J.; Algasaier, S. I.; Kim, J. C.; Finger, L. D.; Jardine, E.; Gotham, V. J. B.; et al. Phosphate Steering by Flap Endonuclease 1 Promotes 5′-Flap Specificity and Incision to Prevent Genome Instability. *Nat. Commun.* **2017**, *8* (May), 1–14.

(127) Sabir, T.; Toulmin, A.; Ma, L.; Jones, A. C.; McGlynn, P.; Schröder, G. F.; Magennis, S. W. Branchpoint Expansion in a Fully Complementary Three-Way DNA Junction. *J. Am. Chem. Soc.* **2012**, *134* (14), 6280–6285.

(128) Algasaier, S. I.; Exell, J. C.; Bennet, I. A.; Thompson, M. J.; Gotham, V. J. B.; Shaw, S. J.; Craggs, T. D.; Finger, L. D.; Grasby, J. A. DNA and Protein Requirements for Substrate Conformational Changes Necessary for Human Flap Endonuclease-1-Catalyzed Reaction. *J. Biol. Chem.* **2016**, *291* (15), 8258–8268.

(129) AlMalki, F. A.; Flemming, C. S.; Zhang, J.; Feng, M.; Sedelnikova, S. E.; Ceska, T.; Rafferty, J. B.; Sayers, J. R.; Artymiuk, P. J. Direct Observation of DNA Threading in Flap Endonuclease Complexes. *Nat. Struct. Mol. Biol.* **2016**, *23* (7), 640–646.

(130) Widom, J. R.; Johnson, N. P.; Von Hippel, P. H.; Marcus, A. H. Solution Conformation of 2-Aminopurine Dinucleotide Determined by Ultraviolet Two-Dimensional Fluorescence Spectroscopy. *New J. Phys.* **2013**, *15*, 1–11.

(131) Eritja, R.; Horowitz, D. M.; Walker, P. A.; Ziehler-Martin, J. P.; Kaplan, B .E., Boosalis, M. S.; Goodman, M. F.; Itakura, K.; Kaplan, B. E. Synthesis and Properties of Oligonucleotides Containing 2'-Deoxynebularine and 2'-Deoxyxanthosine. *Nucleic Acids Res.* **1986**, *14*, 8135–8153.

(132) Sowers, L. C.; Fazakerley, G. V.; Eritja, R.; Kaplan, B. E.; Goodman, M. F. Base

Pairing and Mutagenesis: Observation of a Protonated Base Pair between 2-Aminopurine and Cytosine in an Oligonucleotide by Proton NMR. *Proc. Natl. Acad. Sci.* **1986**, *83* (15), 5434–5438.

(133) Finger, L. D.; Patel, N.; Beddows, A.; Ma, L.; Exell, J. C.; Jardine, E.; Jones, A. C.; Grasby, J. A. Observation of Unpaired Substrate DNA in the Flap Endonuclease-1 Active Site. *Nucleic Acids Res.* **2013**, *41* (21), 9839–9847.

(134) Stryer, L. Fluorescence Energy Transfer as a Spectroscopic Ruler. *Annu. Rev. Biochem.* **1978**, *47* (1), 819–846.

(135) Klostermeier, D.; Millar, D. P. Energetics of Hydrogen Bond Networks in RNA: Hydrogen Bonds Surrounding G+1 and U42 Are the Major Determinants for the Tertiary Structure Stability of the Hairpin Ribozyme. *Biochemistry* **2002**, *41* (48), 14095–14102.

(136) Jose, D.; Datta, K.; Johnson, N. P.; von Hippel, P. H. Spectroscopic Studies of Position-Specific DNA "Breathing" Fluctuations at Replication Forks and Primer-Template Junctions. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (11), 4231–4236.

(137) Datta, K.; Johnson, N. P.; LiCata, V. J.; von Hippel, P. H. Local Conformations and Competitive Binding Affinities of Single-and Double-Stranded Primer-Template DNA at the Polymerization and Editing Active Sites of DNA Polymerases. *J. Biol. Chem.* **2009**, *284* (25), 17180–17193.

(138) Datta, K.; Johnson, N. P.; von Hippel, P. H. DNA Conformational Changes at the Primer-Template Junction Regulate the Fidelity of Replication by DNA Polymerase. *Proc. Natl. Acad. Sci.* **2010**, *107* (42), 17980–17985.

(139) Jose, D.; Weitzel, S. E.; von Hippel, P. H. Breathing Fluctuations in Position-Specific DNA Base Pairs Are Involved in Regulating Helicase Movement into the Replication Fork. *Proc. Natl. Acad. Sci.* **2012**, *109* (36), 14428–14433.

Appendices

Michaelis-Menten experiment template for WThFEN1 with DF5,1 substrate

Make 5X RRB		
Total Vol	1200	uL
10X RB	600	uL
H2O	594	uL
1M DTT	6	uL

Master Mix		
Total Vol	1400	uL
5X RRB	315	uL
H2O	1085	uL
	(1.125X RRB)	

Enzyme Dilutions Stock of 100 Um	
1x RRB	
Total Vol	2500
5X RRB	500
H2O	2000

Enzyme Concs for Experiments (pM)	Concentration pM @ nM	Stock/Previous Dilution	1X RRB
150	1000000/1000	4	396
100	10000/10	4	396
50	1500/1.5	15	85
30	1000/1	10	90
20	500/0.5	20	380
10	300/0.3	60	40
5	200/0.2	40	60
3	100/0.1	40	160
2	50/0.05	50	50
1	30/0.03	30	70
	20/0.02	20	80
	10/0.01	10	90

SUBSTRATE Master Mix

	Stock Conc (nM) --	500	nM @ 0.5µM	
ncs for Experiments (nM)	**Actual Conc (10X)**	**500 nM [S] Stock**	**1x FB**	**Total Vol**
5	50	2	18	20 uL
10	100	4	16	20 uL
25	250	10	10	20 uL
50	500	20	0	20 uL

	Stock Conc (nM) --	5000	nM @ 5µM	
ncs for Experiments (nM)	**Actual Conc (10X)**	**500 nM [S] Stock**	**1x FB**	**Total Vol**
75	750	3	17	20 uL
100	1000	4	16	20 uL
250	2500	2.5	2.5	5 uL

	Stock Conc (nM) --	50000	nM @ 50µM	
ncs for Experiments (nM)	**Actual Conc (10X)**	**500 nM [S] Stock**	**1x FB**	**Total Vol**
500	5000	0.5	4.5	5 uL
750	7500	0.75	4.25	5 uL
1000	10000	1	4	5 uL
2500	25000	2.5	2.5	5 uL
5000	50000	5	0	5 uL

[S] + MM Making		
MM	160	
[S]	20	
Total Vol	180	
X[S] Stock In tube Before adding [E]		After adding [E]
	1.111111111	1
	1.111111111	1
	1.111111111	1

Take out control:
18µL

Final Substrate (nM)	MM (uL) Vol. [E] Injection (uL)	Final [E] (pM)	EDTA Quench in each well (uL)	Vol. to inject onto the WAVE
5	160	18	1	50	65
10	160	18	2	50	65
25	160	18	2	50	65
50	160	18	3	100	77.78
75	160	18	3	100	51.85
100	160	18	5	100	38.89
250	40	4.5	10	100	62.22
500	40	4.5	20	150	46.67
750	40	4.5	30	150	31.11
1000	40	4.5	50	200	31.11
2500	40	4.5	100	200	12.44
5000	40	4.5	150	250	7.78

Dilution 500nM to 250nM EDTA:
10mL EDTA + 10mL H2O

An example multiple turnover experiment template for WThFEN1 and 10 migrating flap substrates

Make 5X RRB		
Total Vol	2400	uL
10X RB	1100	uL
H2O	1089	uL
1M DTT	11	uL

Master Mix		
Total Vol	2800	uL
5X RRB	630	uL
H2O	2170	uL
	(1.125X RRB)	

Enzyme Dilutions Stock of 100 Um		
1x RRB		
Total Vol	4000	uL
5X RRB	800	uL
H2O	3200	uL

Enzyme Concs for Experiments (pM)	Concentration pM @ nM Stock/Previous Dilution		1X RRB
1	1000000/1000	4	396
2.5	10000/10	4	396
4	100/0.1	4	396
3	90/0.09	90	10
6	60/0.06	60	40
9	40/0.04	40	60
	30/0.03	30	70
	2.5/0.025	25	75
	10/0.01	10	90

SUBSTRATE Master Mix

	Stock Conc (nM) -->	500	nM @0.5µM		
Subsr Concs for Experiments (nM)	**Actual Conc (10X)**	**500 nM [S] Stock**		**1x FB**	**Total Vol**
2.5	25	5		95	100 uL
5	50	10		90	100 uL
7.5	75	15		85	100 uL
10	100	20		80	100 uL

[S] + MM Making

MM	160	
[S]	20	
Total Vol	180	
X[S] Stock In tube Before adding [E]		After adding [E]
	1.111111111	1
	1.111111111	1
	1.111111111	1
Take out control:	1.111111111	1
18µL		

Final Substrate (nM)	Substrate MM (uL)	MM (uL)	Volume [E] Injection (uL)	Final [E] (pM)	EDTA Quench in each well (uL)	Volume to inject onto the WAVE
2.5	20	160	18	1, 2.5, 4	50	60
5	20	160	18	1, 2.5, 4	50	60
7.5	20	160	18	3, 6, 9	50	60
10	20	160	18	3, 6, 9	50	60

Dilution 500nM to 250nM EDTA:
5mL EDTA + 5mL H2O

Dilution 50µM - 5µM - 0.5µM »		**50µM [S] Stock**	**1xFB**
	5µM @ 5000nM	5µL	45µL
		(heat & anneal)	
		5µM [S] Stock	**1xFB**
	0.5µM @ 500nM	10µL	90µL

An example single turnover experiment template for WThFEN1 and 10 migrating flap substrates

Reaction	Loop	10X/1X	Mode	Flow rateSwitch Position	Push 1 Steps	Push 2 Steps	Delay Time (s)	Time (ms)
1	1	1X	C	5	160	95	100	4.5
2	1	1X	C	2	160	95	100	12.1
3	2	1X	C	6	180	95	100	9.1
4	2	1X	C	3	180	95	100	19.4
5	2	1X	C	2	180	95	100	27.6
6	2	1X	C	1	180	95	100	41.8
7	4	1X	C	5	255	95	100	30.6
8	4	1X	C	4	255	95	100	40.8
9	4	1X	C	3	255	95	100	57.5
10	4	1X	C	2	255	95	100	82.1
11	4	1X	**C**	**1**	**255**	**95**	**100**	124.2
12	4	1X	I	4	160	95	100	140.8
13	4	1X	I	4	160	95	200	240.8
14	4	1X	I	4	160	95	400	440.8
15	4	1X	I	4	160	95	800	840.8
16	4	1X	I	4	160	95	1600	1640.8
17	4	1X	I	4	160	95	3200	3240.8
18	4	**1X**	I	4	160	95	6400	6440.8
19	4	10X	I	4	160	95	1280	1281
20	4	10X	I	4	160	95	2560	25641
21	4	10X	I	4	160	95	5120	51241

5xRRB (6mL)	Vol. (µl)
10xRB	3000
DTT	30
H2O	2970

Final Conc. For Experiment	[E]2µM (10mL)	µl	[S]5 nM (10mL)	µl
	100 µM [E]	200	500 Nm [UP259EQ]	100
	5xRRB	2000	5xRRB	2000
	1xFB	1000	1xFB	900
	H_2O	6800	H_2O	7000

Dilution 50µM - 5µM - 0.5µM »		**50µM [S] Stock**	**1xFB**
	5µM @ 5000nM	5µL	45µL
		heat & anneal	
		5µM [S] Stock	**1xFB**
	0.5µM @ 500nM	10µL	90µL

Freshly prepare 1XRB
250nM EDTA

An example PCNA experiment template for WThFEN1 with DNA substrates

Make 5X RRB		Enzyme Dilutions Stock of 100 Um	Enzyme Concs for Experiments (pM)	Concentration pM @ nM Stock/Previous Dilution		1X RRB
Total Vol	4800 uL		150	1000000/1000	4	396
10X RB	2400 uL			10000/10	3 set x 4	3 set x 396
H2O	2376 uL			1500/1.5	3 set x 30	3 set x170
1M DTT	24 uL					

1x RRB		Master Mix		[S] + MM Making		
Total Vol	8000 uL	Total Vol	560 uL	MM	160	
5X RRB	1600 uL	5X RRB	126 uL	[S]	20	
H2O	6400 uL	H2O	434 uL	Total Vol	180	
50nM PCNA						
				X[S] Stock In tube Before adding [E]		After adding [E]
					1.111111111	1

	Stock Conc (nM) -->	500 nM @0.5µM			
Subsr Concs for Experiments (nM)	Actual Conc (10X)	500 nM [S] Stock	1x FB	Total Vol	Take out control:
5	50	50	450	500 uL	18µL

Final Substrate (nM)	Substrate MM (uL)	MM (uL)	Volume [E] Injection (uL)	Final [E] (pM)	UREA Quench in each well (uL)	Volume to inject onto the WAVE
5	20	160	18	150	50	60

Dilution [S]: 50µM - 5µM - 0.5µM »		50µM [S] Stock	1xFB	Dilution of 100µM of PCNA
	5µM @ 5000nM	5µL	45µL	100 000nM - 500nM - 50nM
		(heat & anneal)		*50nM of PCNA -
		5µM [S] Stock	1xFB	Use for master mix to get the 2.5, 5, 15, 30, 60, 180nM of PCNA in experiment
	0.5µM @ 500nM	10µL	90µL	

SUBSTRATE Master Mix

[PCNA]; nM (1.125X RRB)	2.5	5	15	30	60	180
5X RRB; µL	220	215	195	165	105	90
H2O; µL	755	735	655	535	295	110
50nM PCNA; µL	25	50	150	300	600	1800
TOTAL; µL	1000	1000	1000	1000	1000	2000

An example threading experiment template for WThFEN1, R103A, R104A, K125A, R129A and R104AK132A with SB5,1 substrate – Bench

	Block		Block	
A	Control	2	Control	4
B	0:00:30	1:00:00	0:00:30	1:00:00
C	0:01:00	2:00:00	0:01:00	2:00:00
D	0:02:00	3:00:00	0:02:00	3:00:00
E	0:04:00	4:00:00	0:04:00	4:00:00
F	0:08:00	6:00:00	0:08:00	6:00:00
G	0:16:00		0:16:00	
H	0:30:00		0:30:00	

	Trap		Trap	
A	Control	6	Control	8
B	0:00:10	0:08:00	0:00:10	0:08:00
C	0:00:20	0:16:00	0:00:20	0:16:00
D	0:00:30	0:30:00	0:00:30	0:30:00
E	0:00:40	1:00:00	0:00:40	1:00:00
F	0:01:00	2:00:00	0:01:00	2:00:00
G	0:02:00	4:00:00	0:02:00	4:00:00
H	0:04:00	6:00:00	0:04:00	6:00:00

	No Streptavidin		No Streptavidin	
A	Control	10	Control	12
B	0:00:10	0:08:00	0:00:10	0:08:00
C	0:00:20	0:16:00	0:00:20	0:16:00
D	0:00:30	0:30:00	0:00:30	0:30:00
E	0:00:40	1:00:00	0:00:40	1:00:00
F	0:01:00	2:00:00	0:01:00	2:00:00
G	0:02:00		0:02:00	
H	0:04:00		0:04:00	

Make 5X CaRRB

Total Vol	2000 uL
10X CaRB	1000 uL
H2O	995 uL
1M DTT	5 uL

1X CaRRB

Total Vol	2000 uL
5X CaRRB	400 uL
H2O	1600 uL

1X MgRRB

Total Vol	2000 uL
10X MgRB	200 uL
H2O	1800 uL

10X Streptavidin

Total Vol	834 uL
1x CaRRB	831 uL
Strp stock	3 uL

10X E @10uM

Total Vol	200 uL
1x CaRRB	180 uL
E	20 uL

Master Mix

Total Vol	2002 uL
5X CaRRB	572 uL
H2O	1430 uL

5µM @ 5000nM

Total Vol	50 uL
50µM	5 uL
1xFB	45 uL

0.5µM @ 500nM

Total Vol	100 uL
5µM	10 uL
1xFB	90 uL

0.1µM @ 100nM

Total Vol	200 uL
0.5µM	40 uL
1xFB	160 uL

Final Concentration
[E]500nM (R104A)
[S]5nM (SB5-1)

Quench
50uL of 8M Urea + 80mM EDTA

Vol. inject (wave)
60uL

Method (wave)
22mer medium 12.5min test2

Block	126uL MM	18uL [S]	18uL Strp	5 min	18uL [E]	2 min	1 min 37°C	20uL cntrl	160uL MgRRB (start)
	0.5 - 360 min (bench)								
Trap	126uL MM	18uL [S]	18uL [E]	2 min	18uL Strp	5 min	1 min 37°C	20uL cntrl	160uL MgRRB (start)
	0.00051 - 360 min (bench) & 0.004017 - 0.23402 min (QF)								
Premix	126uL MM	18uL [S]	18uL [E]	2 min	18uL 1XCaRB	5 min	1 min 37°C	20uL cntrl	160uL MgRRB (start)
	0.166667 - 360 min (bench) & 0.004017 - 0.23402 min (QF)								

An example threading experiment template for WThFEN1, R103A, R104A, K125A, R129A and R104AK132A with SB5,1 substrate – Quenched Flow

Reaction	Loop	10X/1X	Mode	Flow rateSwitch Position	Push 1 Steps	Push 2 Steps	Delay Time (s)	Time (ms)
3	2	1X	C	6	180	95	100	9.1
4	2	1X	C	3	180	95	100	19.4
5	2	1X	C	2	180	95	100	27.6
6	2	1X	C	1	180	95	100	41.8
7	4	1X	C	5	255	95	100	30.6
8	4	1X	C	4	255	95	100	40.8
9	4	1X	C	3	255	95	100	57.5
10	4	1X	C	2	255	95	100	82.1
11	4	1X	C	1	255	95	100	124.2
12	4	1X	I	4	160	95	100	140.8
13	4	1X	I	4	160	95	200	240.8
14	4	1X	I	4	160	95	400	440.8
15	4	1X	I	4	160	95	800	840.8
16	4	1X	I	4	160	95	1600	1640.8
17	4	1X	I	4	160	95	3200	3240.8
18	4	1X	I	4	160	95	6400	6440.8
19	4	10X	I	4	160	95	1280	1281
20	4	10X	I	4	160	95	2560	25641
21	4	10X	I	4	160	95	5120	51241

Left (enzyme/ line A)	
Trap	**µl**
10XCaRB	750
H_20	6635
1M DTT	19
[SB51] 5uM	15
[R104K132A] 100uM	75
Leave 2 minutes	
17U/mg, 2mg/mL Strep	2.7
Leave 5 minutes	
Premix	**µl**
10XCaRB	750
H_20	6635
1M DTT	19
[S] 5uM	15
[E] 100uM	75
Leave 2 minutes	
H_20	2.7
Leave 5 minutes	

Right (line B)	
	µl
10XMgRB	750
H_20	6730
1M DTT	19
Quench	
8M Urea + 300mM EDTA	
Buffer A	
55mM HEPES pH 7.5	
110mM KCL	
1mM $CaCl_2$	
0.02% NaN_3	
15% glycerol	
Buffer B	
55mM HEPES pH 7.5	
110mM KCL	
8mM $MgCl_2$	
0.02% NaN_3	
15% glycerol	

An example ECCD experiments template for WThFEN1, R103A, A107G, K125A, R129A and A98GA107G with FEC1 substrates

1 CD Sample	Volume (µL)	Sample Number +2	Common Volume (µL)	MM
10X RB	50	18	50	1400
1 M $CaCl_2$	0	18	0	0
1M DTT	0.5	18	0.5	14
H_2O	337	18	337	9436
S or FB	50	Total	387.5	10850
E or SB	62.5			
Total	500			

10X RB is: 1.0 M KCl, 100 mM $CaCl_2$, 0.5 M HEPES pH 7.5

SB is: 100 mM KCl, 1 mM $CaCl_2$, 5 mM DTT, 50 mM HEPES pH 7.5, 0.02% NaN_3 (50% glycerol)

for each sample:
run then add 25 µL 250 mM EDTA
mix gently and run again

FEC1 (+1-1 substrate)

Sample	MM (µL)	1×FB (µL)	100 µM ssS (µL)	100 µM S (µL)	SB (µL)	100 µM E (µL)	Total (µL)	Sample Description
1	387.5	50			62.5		500	Blank1 (20°C)
2	387.5		50		62.5		500	ssFEC1 substrate only 10 µM [S] (20°C)
3	387.5			50	62.5		500	dfFEC1 substrate only 10 µM [S] (20°C)
4	387.5			50		62.5	500	12.5 µM hFEN1 with 10 µM [S] (20°C)
5	387.5			50		62.5	500	12.5 µM A107G with 10 µM [S] (20°C)
6	387.5			50		62.5	500	12.5 µM A98GA107G with 10 µM [S] (20°C)
		Total S 100 µM	50	200				
		make at least	75	225				

An example ECCD experiments template for WThFEN1, R103A, A107G, K125A, R129A and A98GA107G with FEC2 substrates

1 CD Sample	Volume (µL)	Sample Number +2	Common Volume (µL)	MM
10X RB	50	18	50	1400
1 M $CaCl_2$	0	18	0	0
1M DTT	0.5	18	0.5	14
H_2O	337	18	337	9436
S or FB	50	Total	387.5	10850
E or SB	62.5			
Total	500			

10X RB is: 1.0 M KCl, 100 mM $CaCl_2$, 0.5 M HEPES pH 7.5

SB is: 100 mM KCl, 1 mM $CaCl_2$, 5 mM DTT, 50 mM HEPES pH 7.5, 0.02% NaN_3 (50% glycerol)

for each sample:
run then add 25 µL 500 mM EDTA
mix gently and run again

FEC2 (-1-2 substrate)

Sample	MM (µL)	1×FB (µL)	100 µM ssS (µL)	100 µM S (µL)	SB (µL)	100 µM E (µL)	Total (µL)	Sample Description
1	387.5	50			62.5		500	Blank1 (20°C)
2	387.5		50		62.5		500	ssFEC2 substrate only 10 µM [S] (20°C)
3	387.5			50	62.5		500	dfFEC2 substrate only 10 µM [S] (20°C)
4	387.5			50		62.5	500	12.5 µM hFEN1 with 10 µM [S] (20°C)
5	387.5			50		62.5	500	12.5 µM A107G with 10 µM [S] (20°C)
6	387.5			50		62.5	500	12.5 µM A98GA107G with 10 µM [S] (20°C)
		Total S 100 µM	50	200				
		make at least	75	225				

K125A sequence references and documentation

Sequence:

```
        10          20          30          40          50          60
MGIQGLAKLI ADVAPSAIRE NDIKSYFGRK VAIDASMSIY QFLIAVRQGG DVLQNEEGET

        70          80          90         100         110         120
TSHLMGMFYR TIRMMENGIK PVYVFDGKPP QLKSGELAKR SERRAEAEKQ LQQAQAAGAE

       130         140         150         160         170         180
QEVEAFTKRL VKVTKQHNDE CKHLLSLMGI PYLDAPSEAE ASCAALVKAG KVYAAATEDM

       190         200         210         220         230         240
DCLTFGSPVL MRHLTASEAK KLPIQEFHLS RILQELGLNQ EQFVDLCILL GSDYCESIRG

       250         260         270         280         290         300
IGPKRAVDLI QKHKSIEEIV RRLDPNKYPV PENWLHKEAH QLFLEPEVLD PESVELKWSE

       310         320         330         340         350         360
PNEEELIKFM CGEKQFSEER IRSGVKRLSK SRQGSTQGRL DDFFKVTGSL SSAKRKEPEP

       370         380
KGSTKKKAKT GAAGKFKRGK
```

References and documentation

Number of amino acids: 380

Molecular weight: 42535.89

Theoretical pI: 8.68

Amino acid composition: CSV format

Ala (A)	32	8.4%
Arg (R)	22	5.8%
Asn (N)	8	2.1%
Asp (D)	16	4.2%
Cys (C)	6	1.6%
Gln (Q)	20	5.3%
Glu (E)	38	10.0%
Gly (G)	27	7.1%
His (H)	8	2.1%
Ile (I)	20	5.3%
Leu (L)	36	9.5%
Lys (K)	37	9.7%
Met (M)	10	2.6%
Phe (F)	14	3.7%
Pro (P)	17	4.5%
Ser (S)	27	7.1%
Thr (T)	12	3.2%
Trp (W)	2	0.5%
Tyr (Y)	8	2.1%
Val (V)	20	5.3%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%
(B)	0	0.0%
(Z)	0	0.0%
(X)	0	0.0%

Total number of negatively charged residues (Asp + Glu): 54
Total number of positively charged residues (Arg + Lys): 59

Atomic composition:

Carbon	C	1876
Hydrogen	H	3033
Nitrogen	N	529
Oxygen	O	564
Sulfur	S	16

Formula: $C_{1876}H_{3033}N_{529}O_{564}S_{16}$
Total number of atoms: 6018

Extinction coefficients:

Extinction coefficients are in units of M^{-1} cm^{-1}, at 280 nm measured in water.

Ext. coefficient	23295
Abs 0.1% (=1 g/l)	0.548, assuming all pairs of Cys residues form cystines

Ext. coefficient	22920
Abs 0.1% (=1 g/l)	0.539, assuming all Cys residues are reduced

Estimated half-life:

The N-terminal of the sequence considered is M (Met).

The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).
>20 hours (yeast, in vivo).
>10 hours (Escherichia coli, in vivo).

Instability index:

The instability index (II) is computed to be 50.09
This classifies the protein as unstable.

Aliphatic index: 81.16

Grand average of hydropathicity (GRAVY): -0.512

```
Alignment of Sequence_1:  [95AB14.seq] with  Sequence_2: [fen1k125a.seq.xdna]

Similarity : 992/1143 (86.79 %)


Seq_1  1079  ------------------------------------------------------------  1080

Seq_2  1     atgggaattcaaggcctggccaaactaattgctgatgtggcccccagtgccatccgggag  60
             M  G  I  Q  G  L  A  K  L  I  A  D  V  A  P  S  A  I  R  E


Seq_1  1079  ------------------------------------------------------------  1080

Seq_2  61    aatgacatcaagagctactttggccgtaaggtggccattgatgcctctatgagcatttat  120
             N  D  I  K  S  Y  F  G  R  K  V  A  I  D  A  S  M  S  I  Y


                                          D  V  L  Q  N  E  E  G  E  X
Seq_1  1079  ----------------------------ggGAtgtgctgCAGAATGaggagGGTGAGAnc  1048
                                         |||||||||||||||||||||||||||||| |
Seq_2  121   cagttcctgattgctgttcgccagggtggggatgtgctgcagaatgaggagggtgaga-c  179
             Q  F  L  I  A  V  R  Q  G  G  D  V  L  Q  N  E  E  G  E  T


             P  Q  X  X  X  G  H  V  L  P  X  P  F  A  *  W  R  T  A  S
Seq_1  1047  CccCAGccnnntgAnggGCATGtTcTACCGncacCAtTcGCATGATGGAGAACGGCATCA  988
             | |||||||:::|||:|||||||||||||||| ||||||||||||||||||||||||||||
Seq_2  180   caccagccacctgatgggcatgttctaccg-caccattcgcatgatggagaacggcatca  238
              T  S  H  L  M  G  M  F  Y  R   T  I  R  M  M  E  N  G  I  K


             S  P  V  Y  V  F  D  G  K  P  P  Q  L  K  S  G  E  L  A  K
Seq_1  987   AGCCCCGTGTATGtCTTTGATGGCAAGCCGcCACAGctCAAGTCAGGCGAGCTGGCCAAA  928
             |||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  239   agccc-gtgtatgtctttgatggcaagccgccacagctcaagtcaggcgagctggccaaa  297
               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  927   CGCAGTGAGCGGCGGGCTGAGGCAGAGAAGCAGCTGCAGCAGGCTCAGGCTGCTgGGGCC  868
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  298   cgcagtgagcggcgggctgaggcagagaagcagctgcagcaggctcaggctgctggggcc  357
             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  A  F  T  K  R  L  V  K  V  T  K  Q  H  N  D
Seq_1  867   GAGCAGGAGGTGGAAGCATTCACTAAGCGGCTGGTGAAGGTCACTAAGCAGCACAATGAT  808
             ||||||||||||||  ||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  358   gagcaggaggtggaaaaattcactaagcggctggtgaaggtcactaagcagcacaatgat  417
             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  807   GAGTGCAAACATCTGCTGAGCCTCATGGGCAtCCCTTATCTTGATGCACCCAGTGAGGCA  748
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  418   gagtgcaaacatctgctgagcctcatgggcatcccttatcttgatgcacccagtgaggca  477
             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  747   GAGGCCAGCTGTGCTGCCCTGGTGAAGGCTGGCAAAGTCTATGCTGCGGCTACCGAGGAC  688
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  478   gaggccagctgtgctgccctggtgaaggctggcaaagtctatgctgcggctaccgaggac  537
             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  687   ATGGACTGCCTCACCTTCGGCAGCCCTGTGCTAATGCGACACCTGACTGCCAGTGAAGCC  628
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  538   atggactgcctcaccttcggcagccctgtgctaatgcgacacctgactgccagtgaagcc  597
             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  627   AAAAAGCTGCCAATCCAGGAATTCCACCTGAGCCGGATTCTGCAGGAGCTGGGCCTGAAC  568
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  598   aaaaagctgccaatccaggaattccacctgagccggattctgcaggagctgggcctgaac  657
```

```
            K  K  L  P  I  Q  E  F  H  L  S  R  I  L  Q  E  L  G  L  N

            Q  E  Q  F  V  D  L  C  I  L  L  G  S  D  Y  C  E  S  I  R
Seq_1  567  CAGGAACAGTTTGTGGATCTGTGCATCCTGCTAGGCAGTGACTACTGTGAGAGTATCCGG  508
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  658  caggaacagtttgtggatctgtgcatcctgctaggcagtgactactgtgagagtatccgg  717
            Q  E  Q  F  V  D  L  C  I  L  L  G  S  D  Y  C  E  S  I  R

            G  I  G  P  K  R  A  V  D  L  I  Q  K  H  K  S  I  E  E  I
Seq_1  507  GGTATTGGGCCCAAGCGGGCTGTGGACCTCATCCAGAAGCACAAGAGCATCGAGGAGATC  448
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  718  ggtattgggcccaagcgggctgtggacctcatccagaagcacaagagcatcgaggagatc  777
            G  I  G  P  K  R  A  V  D  L  I  Q  K  H  K  S  I  E  E  I

            V  R  R  L  D  P  N  K  Y  P  V  P  E  N  W  L  H  K  E  A
Seq_1  447  GTGCGGCGACTTGACCCCAACAAGTACCCTGTGCCAGAAAATTGGCTCCACAAGGAGGCT  388
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  778  gtgcggcgacttgaccccaacaagtaccctgtgccagaaaattggctccacaaggaggct  837
            V  R  R  L  D  P  N  K  Y  P  V  P  E  N  W  L  H  K  E  A

            H  Q  L  F  L  E  P  E  V  L  D  P  E  S  V  E  L  K  W  S
Seq_1  387  CACCAGCTCTTCTTGGAACCTGAGGTGCTGGACCCAGAGTCTGTGGAGCTGAAGTGGAGC  328
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  838  caccagctcttcttggaacctgaggtgctggacccagagtctgtggagctgaagtggagc  897
            H  Q  L  F  L  E  P  E  V  L  D  P  E  S  V  E  L  K  W  S

            E  P  N  E  E  E  L  I  K  F  M  C  G  E  K  Q  F  S  E  E
Seq_1  327  GAGCCAAATGAAGAAGAGCTGATCAAGTTCATGTGTGGTGAAAAGCAGTTCTCTGAGGAG  268
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  898  gagccaaatgaagaagagctgatcaagttcatgtgtggtgaaaagcagttctctgaggag  957
            E  P  N  E  E  E  L  I  K  F  M  C  G  E  K  Q  F  S  E  E

            R  I  R  S  G  V  K  R  L  S  K  S  R  Q  G  S  T  Q  G  R
Seq_1  267  CGAATCCGCAGTGGGGTCAAGAGGCTGAGTAAGAGCCGCCAAGGCAGCACCCAGGGCCGC  208
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  958  cgaatccgcagtggggtcaagaggctgagtaagagccgccaaggcagcacccagggccgc  1017
            R  I  R  S  G  V  K  R  L  S  K  S  R  Q  G  S  T  Q  G  R

            L  D  D  F  F  K  V  T  G  S  L  S  S  A  K  R  K  E  P  E
Seq_1  207  CTGGATGATTTCTTCAAGGTGACCGGCTCACTCTCTTCAGCTAAGCGCAAGGAGCCAGAA  148
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  1018 ctggatgatttcttcaaggtgaccggctcactctcttcagctaagcgcaaggagccagaa  1077
            L  D  D  F  F  K  V  T  G  S  L  S  S  A  K  R  K  E  P  E

            P  K  G  S  T  K  K  K  A  K  T  G  A  A  G  K  F  K  R  G
Seq_1  147  CCCAAGGGATCCACTAAGAAGAAGGCAAAGACTGGGGCAGCAGGGAAGTTTAAAAGGGGA  88
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  1078 cccaagggatccactaagaagaaggcaaagactggggcagcagggaagtttaaaagggga  1137
            P  K  G  S  T  K  K  K  A  K  T  G  A  A  G  K  F  K  R  G

            K  H  H  H  H  H  H  *  A  C  G  R  T  R  A  P  P  P  P  P
Seq_1  87   AAACATCATCATCATCATCACTAAGCTTGCGGCCGCActCGAGCACCACCACCACCACCA  28
            |||                  |||
Seq_2  1138 aaa------------------taa-----------------------------------  1143
            K                    *

            L  R  S  G  C  *  X  S  P
Seq_1  27   CtGAGatCCGGCtgCtaAcnAAgccCg  1

Seq_2  1144 --------------------------  1143
                              X  X
```

```
Alignment of Sequence_1:  [95AB15.seq] with  Sequence_2: [fen1k125a.seq.xdna]

Similarity : 1061/1143 (92.83 %)

                 *  X  F  C  X  T  L  R  R  R  Y  T  M  G  I  Q  G  L  A  K
Seq_1  1     tCtaganaTTTtgTTnAaCTTtAAGAAGGAGATATACCATGGgAaTTCAAGGCCTGGCCA  60
                                                 ||||||||||||||||||||||
Seq_2  1     ---------------------------------------atgggaattcaaggcctggcca  22
                                                    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  Y  F  G
Seq_1  61    AACTAATTGCTGATGTGGCCCCCAGTGCCATCCGGgaGAATGACATCAAGAGCTACTTTG  120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  23    aactaattgctgatgtggcccccagtgccatccgggagaatgacatcaagagctactttg  82
                 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  V  R  Q
Seq_1  121   GCCGTAAGGTGGCCATTGATGCCTCTATGAGCATTTATCAGTTCCTGATTGCTGTTCGCC  180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  83    gccgtaaggtggccattgatgcctctatgagcatttatcagttcctgattgctgttcgcc  142
                 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  G  M  F
Seq_1  181   AGGGTGGGGATGTGCTGCAGAATGAGGAGGGTGAGACCACCAGCCACCTGATGGGCATGT  240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  143   agggtggggatgtgctgcagaatgaggagggtgagaccaccagccacctgatgggcatgt  202
                 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  D  G  K
Seq_1  241   TCTACCGCACCATTCGCATGATGGAGAACGGCATCAAGCCCGTGTATGTCTTTGATGGCA  300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  203   tctaccgcaccattcgcatgatggagaacggcatcaagcccgtgtatgtctttgatggca  262
                 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  E  A  E
Seq_1  301   AGCCGCCACAGCTCAAGTCAGGCGAGCTGGCCAAACGCAGTGAGCGGCGGGCTGAGGCAG  360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  263   agccgccacagctcaagtcaggcgagctggccaaacgcagtgagcggcgggctgaggcag  322
                 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  A  F  T  K
Seq_1  361   AGAAGCAGCTGCAGCAGGCTCAGGCTGCTGGGGCCGAGCAGGAGGTGGAAGCATTCACTA  420
             ||||||||||||||||||||||||||||||||||||||||||||||||  ||||||||
Seq_2  323   agaagcagctgcagcaggctcaggctgctggggccgagcaggaggtggaaaaattcacta  382
                 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  S  L  M
Seq_1  421   AGCGGCTGGTGAAGGTCACTAAGCAGCACAATGATGAGTGCAAACATCTGCTGAGCCTCA  480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  383   agcggctggtgaaggtcactaagcagcacaatgatgagtgcaaacatctgctgagcctca  442
                 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  L  V  K
Seq_1  481   TGGGCATCCCTTATCTTGATGCACCCAGTGAGGCAGAGGCCAGCTGTGCTGCCCTGGTGA  540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  443   tgggcatcccttatcttgatgcacccagtgaggcagaggccagctgtgctgccctggtga  502
                 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  G  S  P
Seq_1  541   AGGCTGGCAAAGTCTATGCTGCGGCTACCGAGGACATGGACTGCCTCACCTTCGGCAGCC  600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  503   aggctggcaaagtctatgctgcggctaccgaggacatggactgcctcaccttcggcagcc  562
                 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  E  F  H
Seq_1  601   CTGTGCTAATGCGACACCTGACTGCCAGTGAAGCCAAAAAGCTGCCAATCCAGGAATTCC  660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  563   ctgtgctaatgcgacacctgactgccagtgaagccaaaaagctgccaatccaggaattcc  622
```

```
                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  E  Q  F  V  D  L  C  I
Seq_1  661    ACCTGAGCCGGATTCTGCAGGAGCTGGGCCTGAACCAGGAACAGTTTGTGGATCTGTGCA  720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  623    acctgagccggattctgcaggagctgggcctgaaccaggaacagtttgtggatctgtgca  682
                L  S  R  I  L  Q  E  L  G  L  N  Q  E  Q  F  V  D  L  C  I

                L  L  G  S  D  Y  C  E  S  I  R  G  I  G  P  K  R  A  V  D
Seq_1  721    TCCTGCTAGGCAGTGACTACTGTGAGAGTATCCGGGGTATTGGgCCCAAGCGGGCTGTGG  780
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  683    tcctgctaggcagtgactactgtgagagtatccggggtattgggcccaagcgggctgtgg  742
                L  L  G  S  D  Y  C  E  S  I  R  G  I  G  P  K  R  A  V  D

                L  I  Q  K  H  K  S  I  E  E  I  V  R  R  L  D  P  N  K  Y
Seq_1  781    ACCTCATCCAGAAGCACAAGAGCATCGAGGAgATCGTGCGGCGACTTGACCCCAACAAGT  840
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  743    acctcatccagaagcacaagagcatcgaggagatcgtgcggcgacttgaccccaacaagt  802
                L  I  Q  K  H  K  S  I  E  E  I  V  R  R  L  D  P  N  K  Y

                P  V  P  E  N  W  L  H  K  E  A  H  Q  L  F  L  E  P  E  V
Seq_1  841    ACCCTGTGCCAGAAAATTGGCTCCACAAGGAGGCTCACCAGCTCTTCTTGGAACCTGAGG  900
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  803    accctgtgccagaaaattggctccacaaggaggctcaccagctcttcttggaacctgagg  862
                P  V  P  E  N  W  L  H  K  E  A  H  Q  L  F  L  E  P  E  V

                L  D  P  E  S  V  E  L  K  W  S  E  P  N  E  E  E  L  I  K
Seq_1  901    TGCTGGACCCAGAGTCTGTGGaGCTGAAGTGGAGCGAGCCAAATGAagAagAGCTGATCA  960
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  863    tgctggacccagagtctgtggagctgaagtggagcgagccaaatgaagaagagctgatca  922
                L  D  P  E  S  V  E  L  K  W  S  E  P  N  E  E  E  L  I  K

                F  M  C  G  E  K  Q  F  S  E  E  R  I  R  S  G  V  X  X  L
Seq_1  961    AGTTCATGTGTGGTGAAAAGCAGTTCTCTGAGGagCGAATCCGCAGTGGGgTCAannGGC  1020
               ||||||||||||||||||||||||||||||||||||||||||||||||||||::|||
Seq_2  923    agttcatgtgtggtgaaaagcagttctctgaggagcgaatccgcagtggggtcaagaggc  982
                F  M  C  G  E  K  Q  F  S  E  E  R  I  R  S  G  V  K  R  L

                N  K  K  P  P  R  Q  H  P  G  P  P  X  *  F  L  Q  X  X  R
Seq_1  1021   TGaatAagaagccgCCAAGGCAGCACCcAGGgccgcctnnaTGaTTtCTTCAAGgnnacc  1080
               ||| |||| ||||||||||||||||||||||||||::||||||||||||||::|||
Seq_2  983    tgagtaag-agccgccaaggcagcacccagggccgcctggatgatttcttcaaggtgacc  1041
                S  K   S  R  Q  G  S  T  Q  G  R  L  D  D  F  F  K  V  T

                X  X  L  F  S  *  X
Seq_1  1081   ggnncnnnctcTTCAgCTAaGng----------------------------------  1103
              ||::|:::|||||||||||||:|
Seq_2  1042   ggctcactctcttcagctaagcgcaaggagccagaacccaagggatccactaagaagaag  1101
              G  S  L  S  S  A  K  R  K  E  P  E  P  K  G  S  T  K  K  K

                                                 X  X
Seq_1  1104   ----------------------------------------  1103

Seq_2  1102   gcaaagactggggcagcagggaagtttaaaaggggaaaataa  1143
              A  K  T  G  A  A  G  K  F  K  R  G  K  *
```

Nucleic Acids Research, 2018 **1**
doi: 10.1093/nar/gky293

Regional conformational flexibility couples substrate specificity and scissile phosphate diester selectivity in human flap endonuclease 1

Ian A. Bennet[1,†], L. David Finger[1,*,†], Nicola J. Baxter[2,3], Benjamin Ambrose[1], Andrea M. Hounslow[2], Mark J. Thompson[1], Jack C. Exell[1], Nur Nazihah B. Md. Shahari[1], Timothy D. Craggs[1,*], Jonathan P. Waltho[2,3] and Jane A. Grasby[1,*]

[1]Centre for Chemical Biology, Department of Chemistry, Krebs Institute for Biomolecular Research, The University of Sheffield, Sheffield S3 7HF, UK, [2]Department of Molecular Biology and Biotechnology, Krebs Institute for Biomolecular Research, The University of Sheffield, Sheffield S10 2TN, UK and [3]Manchester Institute of Biotechnology, School of Chemistry, The University of Manchester, Manchester M1 7DN, UK

Received September 01, 2017; Revised March 26, 2018; Editorial Decision April 04, 2018; Accepted April 09, 2018

ABSTRACT

Human flap endonuclease-1 (hFEN1) catalyzes the divalent metal ion-dependent removal of single-stranded DNA protrusions known as flaps during DNA replication and repair. Substrate selectivity involves passage of the 5′-terminus/flap through the arch and recognition of a single nucleotide 3′-flap by the α2–α3 loop. Using NMR spectroscopy, we show that the solution conformation of free and DNA-bound hFEN1 are consistent with crystal structures; however, parts of the arch region and α2–α3 loop are disordered without substrate. Disorder within the arch explains how 5′-flaps can pass under it. NMR and single-molecule FRET data show a shift in the conformational ensemble in the arch and loop region upon addition of DNA. Furthermore, the addition of divalent metal ions to the active site of the hFEN1–DNA substrate complex demonstrates that active site changes are propagated via DNA-mediated allostery to regions key to substrate differentiation. The hFEN1–DNA complex also shows evidence of millisecond timescale motions in the arch region that may be required for DNA to enter the active site. Thus, hFEN1 regional conformational flexibility spanning a range of dynamic timescales is crucial to reach the catalytically relevant ensemble.

INTRODUCTION

Flap endonuclease 1 (FEN1) is a member of the 5′-nuclease superfamily that is predominantly involved in Okazaki fragment maturation, but also has roles in long-patch base excision repair and ribonucleotide excision repair (1,2). All three of these pathways create bifurcated nucleic acid structures known as 5′-flaps that must be removed precisely to create single-stranded (ss) and nicked double-stranded (ds) DNA products (Figure 1A). In line with a critical role in DNA replication, FEN is present in all organisms, from bacteriophage to mammals (3). In humans, cancer cells overexpress FEN1, and tumor aggressiveness correlates with FEN1 protein levels (4). As such, human FEN1 (hFEN1) has been postulated to be a potential cancer therapeutic target (5,6), and evidence suggests that combinatorial targeting of hFEN1 has therapeutic relevance (7). Moreover, FEN1 is the prototypical member of the 5′-nuclease superfamily whose activities span all major DNA metabolic pathways. As nucleases, specificity is paramount as unwanted hydrolysis of DNA or RNA can be deleterious; thus, how hFEN1 and paralogues achieve substrate and scissile phosphate diester specificity has been an area of considerable debate.

Earlier studies revealed that hFEN1 specificity arises from the recognition of three structural features of its substrate (Figure 1A) (2,8). Firstly, only junction dsDNAs can be bound stably by the protein as two juxtaposed dsDNA binding sites require the substrate to bend roughly 100° (Figure 1B) (9–11). Secondly, despite removing 5′-flaps, hFEN1 recognizes substrates bearing a one nucleotide (1-nt) 3′-flap (12–14). In structures of hFEN1–DNA com-

*To whom correspondence should be addressed. Tel: +44 114 222 9478; Fax: +44 114 222 9346; Email: j.a.grasby@sheffield.ac.uk
Correspondence may also be addressed to L. David Finger. Tel: +44 114 222 9364; Fax: +44 114 222 9346; Email: d.finger@sheffield.ac.uk
Correspondence may also be addressed to Timothy D. Craggs. Tel: +44 114 222 9347; Fax: +44 114 222 9346; Email: t.craggs@sheffield.ac.uk
† The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.
Present address: Jack C. Exell, Department of Chemistry, 1207 French Family Science Centre, Duke University, Durham, NC 27705, USA.

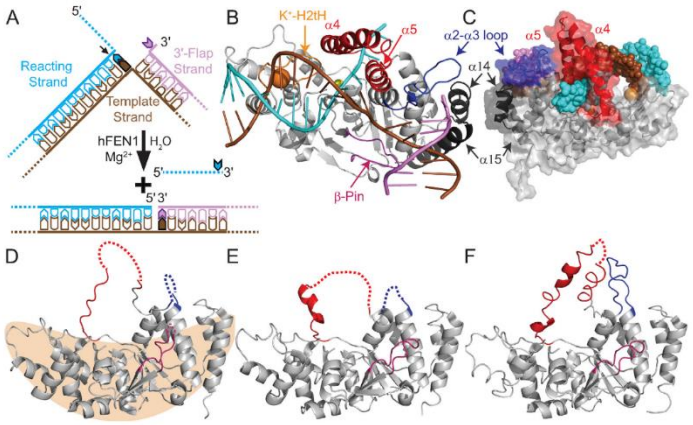

Figure 1. Regions of hFEN1 assume various conformations in crystal structures. (**A**) Representation of hFEN1-catalyzed reaction of double-flap DF#,1 creating ssDNA and nicked dsDNA products (2). # denotes a 5′-flap of any length, 1 denotes a single nucleotide 3′-flap. (**B**) The $hFEN1_{D86N}$-(DF2,1) complex (5UM9) (21) shows that the arch (α4-α5, red), α2–α3 loop (blue) and β-pin (β6-β7 loop) (pink) are ordered in the presence of DNA substrate (colored as in panel A). The α10-α11 loop (H2tH motif) that binds the reacting duplex approximately 20 Å from the scissile phosphate and the α14-loop-α15 that is unique to FEN proteins are shown in orange and black, respectively. The potassium and active site samarium cations are shown as orange and yellow spheres, respectively. (**C**) Rear view of 5UM9 (cartoon with transparent surface representation) with regions colored as in panels A and B and the DNA (DF2,1) shown as spheres. (**D**–**F**) The three crystal structures of apo-hFEN1 (1UL1) show the arch, α2–α3 loop and β-pin regions in various states of order and disorder, dashed lines indicate missing electron density (22). (D) The tan shape illustrates the saddle-region of hFEN1 that binds the dsDNA and active site metals.

plexes, the 3′-flap is cradled in a pocket created from the α14–α15 loop, the first turn of α15 and the α2–α3 loop (11). The α2–α3 loop folds into a non-regular secondary structure known as an Ω-loop (15), often associated with allosterically-regulated regions of proteins (Supplementary Figure S1) (16). Thirdly, the 5′-flap portion of the substrate must pass under the arch region (α4-α5) for biologically-relevant rates of reaction to occur (Figure 1B and C) (17,18). This latter substrate selection feature is also shared by the 5′-nuclease superfamily member human exonuclease-1 (19,20), but not by other family members that act on continuous DNAs like bubble-cutting XPG and the Holliday junction resolvase GEN1 (3). When these three substrate recognition tests are met, the DNA can untwist and enter the active site for catalysis. Of the three hFEN1 substrate feature selection processes, the threading of the 5′-flap through the arch remains the most enigmatic.

Crystal structures of hFEN1–DNA complexes (3Q8K, 3Q8L, 3Q8M, 5UM9) show the arch is composed of two α-helices above the active site (11,21). Together these create a hole in the protein that is large enough to accommodate ss-DNA but not dsDNA (11,21), which could explain hFEN1 specificity for 5′-ss flaps (Figure 1C). However, this is inconsistent with the ability of hFEN1 to process 5′-flaps with some duplex secondary structure (12,18). A further puzzle relates to how threading could occur through a small hole in the absence of an energy source. In contrast, apo-hFEN1 structures show that the regions corresponding to the arch and α2–α3 loop sometimes lack discernible electron density; what has been observed is either random coil or a limited degree of secondary structure (Figure 1D–F) with high B-factors (Supplementary Figure S2) (22). Furthermore, structures of FEN proteins from a wide array of organisms show the arch region in various conformations and positions (Supplementary Figure S3) (2). It is possible that structural heterogeneity of the arch and α2–α3 loop has some functional relevance. However, direct confirmation of the structural status of the arch in free and substrate bound hFEN1 is necessary to define its mechanism.

To extend our understanding, multidimensional NMR spectroscopy was used to assess the properties and dynamics of hFEN1 alone and in complex with DNA in solution. We show that the solution conformation of hFEN1 is consistent with the crystal structure in the dsDNA binding region known as the saddle (Figure 1D), but a large discrepancy is observed in the arch region (traditionally referred to as the helical arch). Using a combination of kinetic, spin relaxation, chemical shift and single molecule FRET data, the contribution of hFEN1 conformationally dynamic regions to substrate specificity is revealed. Chemical shift perturbation mapping provides evidence that divalent cations induce conformational changes in the hFEN1–DNA complex that are not observed in hFEN1 alone. These data demonstrate that the active site communicates with regions involved in dsDNA binding and substrate feature recognition. Combined data raise the possibility that both conformational selection and induced fit mechanisms occur in the arch region in response to structural features of the substrate and reveal that the hFEN1–DNA complex exhibits millisecond timescale motions in the arch. Taken together, our analysis shows that regional conformational flexibility spanning a range of dynamic timescales is channeled into

the catalytically-competent conformational ensemble in response to substrate features, providing the required link between specificity and catalysis in a structure-sensing nuclease. Because the structures of other 5′-nuclease superfamily show disorder or indicate flexibility in the area that corresponds to the hFEN1 arch region (19,20,23–25), a similar conformational ensemble shift mechanism in response to other paralogue-specific triggers is likely important in their mechanisms as well.

MATERIALS AND METHODS

Protein expression

Human hFEN1-336 was expressed from *Escherichia coli* BL21(DE3)-RILP cells transformed with pET29b-hFEN1-336 plasmid, encoding for hFEN1-336 (herein referred to as hFEN1) protein containing a human rhinovirus (HRV) type 14 3C protease-cleavable $(His)_6$-tag located at the C-terminus. Natural abundance protein for kinetic analyses was expressed and purified as described previously (7). ^{15}N-labeled protein was expressed using ^{15}N autoinduction media (26). The final concentrations of ^{15}N autoinduction media components were: 50 mM Na_2HPO_4 and 50 mM KH_2PO_4 pH 7.5, 50 mM $^{15}NH_4Cl$, 5 mM Na_2SO_4, 2 mM $MgSO_4$, 0.5% glycerol, 0.05% D-glucose, 0.2% α-lactose, 12× BME vitamins (USBiological B0110), 1× Trace Metals (Teknova 1000x Trace Metals T1001), 400 μg ml^{-1} kanamycin and 34 μg ml^{-1} chloramphenicol. Cultures for ^{15}N-labeled protein expression were inoculated with a 100-fold dilution of a saturated 2× YT starter culture, and were then allowed to grow at 37°C for 4 h before the temperature was reduced to 18°C to allow for overnight expression by autoinduction.

$^{2}H,^{13}C,^{15}N$-labeled protein was prepared using a high cell-density procedure in combination with isopropyl-β-D-thiogalatopyranoside (IPTG) induction (27). The final concentrations of all media components were: 50 mM Na_2HPO_4 and 25 mM KH_2PO_4 pH 7.5, 18 mM $^{15}NH_4Cl$, 1% $^{13}C_6,^{2}H_7$-D-glucose (U-$^{13}C_6$, 99%; 1,2,3,4,5,6,6-d$_7$ 97-98%), 0.2 mM $CaCl_2$, 5 mM $MgSO_4$, 10 mM NaCl, 0.25 × BME vitamins (USBiological B0110), 0.25 × Trace Metals (Teknova 1000x Trace Metals T1001), 400 μg.ml^{-1} kanamycin and 34 μg ml^{-1} chloramphenicol. Initially, a 100-fold dilution of a saturated culture of BL21(DE3)-RILP cells transformed with pET29b-hFEN1-336 was used to inoculate 2× YT media prepared with 50% D_2O and was grown until saturation. A 100-fold dilution of this saturated 50% D_2O 2× YT culture was then used to inoculate 2× YT media in 100% D_2O to allow the cells to acclimatize for growth. Aliquots (50 ml) of these D_2O acclimatized BL21(DE3)-RILP cells were added to 1 L of 100% D_2O 2× YT media and were grown at 37°C until the A_{600} = 5–6. The culture was pelleted by ultracentrifugation (6000 × *g* at 4°C for 30 min) and the supernatant was discarded. The pellet was then resuspended in 1 L of $^{15}N,^{2}H,^{13}C$ minimal media (as described above) and then allowed to grow at 37°C for a further 2 h to clear unlabeled metabolites. Afterward, IPTG was added to a final concentration of 0.1 mM IPTG, the temperature was reduced to 18°C, and the culture was grown for a further 72 h until A_{600} = 10–13. $^{2}H,^{15}N$-labeled protein was prepared analogously except that the 1% $^{13}C_6,^{2}H_7$-D-glucose (U-$^{13}C_6$, 99%; 1,2,3,4,5,6,6-d$_7$ 97–98%) was replaced with 1% $^{2}H_7$-D-glucose (1,2,3,4,5,6,6-d$_7$ 97–98%).

In all cases, cells were harvested by centrifugation (6000 × *g* at 4°C for 15 min) and the supernatant was removed. The cell pellet was resuspended in 50 ml of ice-cold phosphate buffered saline (10 mM Na_2HPO_4 and 2 mM KH_2PO_4 pH 7.4, 137 mM NaCl, 2.7 mM KCl), the cells were pelleted again at (6000 × *g* at 4°C for 15 min) and the supernatant was removed. The cells were then resuspended in 50 ml of Buffer A (20 mM Tris–HCl pH 7.4, 1 M NaCl, 5 mM imidazole) containing 5 mM β-mercaptoethanol (βME), 0.1 mg ml^{-1} lysozyme and 1 EDTA-free protease inhibitor cocktail tablet (Sigma S8830). Cells were incubated on ice for at least 1 h and were frozen at −20°C until required.

Protein purification

To purify hFEN1 protein, cells were thawed and then sonicated on ice for 60 cycles of pulsation for 5 s with 25 s cooling intervals. To each lysate, 5 ml of Buffer A containing 1% Tween-20 was added, and the lysate was clarified by centrifugation (40 000 × *g* at 4°C for 30 min). The supernatant was applied to a column (ID = 1.6 cm, length = 12 cm) containing Chelating Sepharose 6 Fast Flow agarose beads (GE Healthcare Life Sciences) that had been previously charged with Ni^{2+} ions according to the manufacturer's protocol and pre-equilibrated with 5 column volumes of Buffer A. The column was washed with 5 column volumes of Buffer A and 5 column volumes of Buffer B (25 mM Tris–HCl pH 7.4, 0.5 M NaCl, 40 mM imidazole, 0.01% Tween-20, 5 mM βME). The bound hFEN1 protein was eluted with 8 column volumes of Buffer C (250 mM imidazole pH 7.2, 0.5 M NaCl, 5 mM βME). Fractions containing hFEN1 protein were identified by SDS-PAGE, pooled and immediately diluted with an equal volume of Buffer D (20 mM Tris–HCl pH 7.4, 1 mM EDTA, 20 mM βME). The solution was then applied to a 5 mL HiTrap Q HP column (GE Healthcare Life Sciences) pre-equilibrated with the buffer D. hFEN1 protein was found in the flow through, whereas nucleic acid contamination was retained by the column and eluted using 10 column volumes of Buffer D containing a linear gradient of 0–1 M NaCl. The amount of hFEN1 protein in the flow through was estimated by measuring the absorbance at 280 nm (ε_{280} = 22 920 M^{-1} cm^{-1}) (28) using a UV-Vis NanoDrop spectrophotometer (ThermoFisher Scientific). Two units of Turbo3C (HRV3C protease) (BioVision) were added for every mg of protein, and the mixture was allowed to stand overnight at 4°C to catalyze cleavage of the $(His)_6$-tag from the hFEN1 protein. The extent of affinity tag removal was assessed by SDS-PAGE, and the solution was diluted further with an equal volume of buffer E (25 mM Tris–HCl pH 7.5, 1 mM $CaCl_2$, 20 mM βME). The hFEN1 protein solution was then applied to a 20 ml HiPrep Heparin FF 16/10 column (GE Healthcare Life Sciences) pre-equilibrated with 2 column volumes of buffer E and eluted using 15 column volumes of Buffer E containing a linear gradient of 0–1 M NaCl (hFEN1 protein eluted at 350 mM NaCl). Fractions containing hFEN1 protein were pooled and concentrated to less than 10 ml using an Am-

icon Ultrafiltration cell with a 10 kDa MWCO PES membrane (MerckMillipore) pressurised using nitrogen gas. The protein was exchanged into the appropriate buffer (see below) using a HiPrep 26/10 desalting column (GE Healthcare Life Sciences). The hFEN1 protein concentration was determined using the absorbance at 280 nm as described above. Unlabeled samples for kinetic measurements were exchanged into 100 mM HEPES–KOH pH 7.5, 200 mM KCl, 1 mM EDTA, 10 mM DTT and 0.04% NaN_3, and adjusted to a concentration of 200 μM. An equal volume of glycerol was added to each sample to allow optimal storage conditions (final concentration: 100 μM hFEN1, 50 mM HEPES–KOH pH 7.5, 100 mM KCl, 0.5 mM EDTA, 5 mM DTT, 0.02% NaN_3 and 50% glycerol). Samples were stored at −20°C. Isotopically labeled hFEN1 protein samples were exchanged into the appropriate NMR buffer and concentrated to 0.5 mM using a Vivaspin 20 (10 kDa MWCO) concentrator (4000 × *g* at 4°C).

All purification buffers were filtered and degassed prior to use, and an ÄKTA pure chromatography system was used for all column purification steps. Unless otherwise stated, all reagents were purchased from Sigma-Aldrich, Fisher Scientific, or VWR International and used as received.

NMR assignment

All NMR experiments were performed at 298 K (relative to d_4-methanol signal) using standard pulse sequences on either a Bruker 600 MHz Avance DRX spectrometer equipped with a 5-mm TXI cryoprobe with z-axis gradients or a Bruker 800 MHz Avance I spectrometer equipped with a 5-mm TXI probe with z-axis gradients. Both spectrometers were operated with TopSpin 2. Comparison of ^{1}H–^{15}N TROSY spectra of ^{2}H,^{15}N,^{13}C-labeled hFEN1 with ^{15}N-labeled hFEN1 showed that all backbone amide groups had back exchanged from deuterium to protium atoms. For the backbone resonance assignment of hFEN1, 0.5 ml samples contained 0.5 mM ^{2}H,^{15}N,^{13}C-labeled hFEN1 in 10 mM HEPES–KOH pH 7.5, 76 mM KCl, 0.1 mM EDTA, 4 mM NaN_3, 100 mM βME. D_2O (10% v/v) and 0.05 mM trimethylsilyl propanoic acid (TSP) were added for the deuterium lock and chemical shift reference standard, respectively. After transferring the sample into a 5 mm NMR tube, the sample was sealed with a Precision Seal® rubber septa cap (Sigma-Aldrich Z554014) and argon was passed over the sample to help minimize oxidation.

hFEN1 backbone $^{1}H^{N}$, ^{15}N, $^{13}C^{\alpha}$, $^{13}C^{\beta}$ and $^{13}C'$ were assigned using the standard suite of ^{1}H–^{15}N TROSY and 3D TROSY-based constant time experiments (HNCA, HN(CO)CA, HNCACB, HN(CO)CACB, HN(CA)CO and HNCO) (29). ^{1}H chemical shifts were referenced relative to the internal TSP signal, whereas ^{15}N and ^{13}C chemical shifts were indirectly referenced using nuclei-specific gyromagnetic ratios. Peak picking and frequency matching was performed within CCPNMR Analysis version 2.4 (30), and the backbone assignments were confirmed independently using a simulated annealing algorithm employed by the *asstools* assignment program (31). The secondary structure content and Random Coil Index-S^2 (RCI-S^2) prediction of hFEN1 was conducted by uploading the backbone $^{1}H^{N}$, ^{15}N, $^{13}C^{\alpha}$, $^{13}C^{\beta}$ and $^{13}C'$ chemical shifts to the TALOS-N webserver (32).

NMR relaxation measurements

hFEN1 samples for ^{15}N backbone fast timescale relaxation measurements were performed using ^{2}H,^{15}N-labeled hFEN1 in a 5 mm Shigemi NMR microtube assembly matched with D_2O. Sample conditions were 10 mM HEPES–KOH pH 7.5, 76 mM KCl, 4 mM NaN_3 100 mM βME, 10% D_2O, 0.05 mM TSP and 0.1 mM EDTA. Spin-lattice ^{15}N relaxation rates (R_1), rotating frame ^{15}N relaxation rates ($R_{1\rho}$) and heteronuclear steady-state ^{15}N-$\{^{1}H\}$ NOE (hNOE) values were obtained using interleaved TROSY-readout pulse sequences (33). Temperature compensation was applied in the R_1 experiment by incorporating a spin-lock pulse placed off-resonance in the inter-scan delay, equal to the longest spin-lock time and RF power of the $R_{1\rho}$ experiment. Relaxation delays of 0, 80, 240, 400, 400, 640, 800, 1200, 1760 and 2400 ms were used to calculate R_1, and delays of 1, 20, 20, 30, 40, 60, 90, 110, 150 and 200 ms were used to calculate $R_{1\rho}$. The inter-scan delay was 3.5 s and the strength of the RF spin-lock field during $R_{1\rho}$ measurement was 1400 Hz at 600 MHz and 1866.7 Hz at 800 MHz. For the hNOE measurement, two interleaved experiments were acquired with relaxation delays of 10 s. For the determination of R_1 and $R_{1\rho}$ rates, the decay of backbone amide peak intensities were fitted using an exponential function in CCPNMR (30). Relaxation parameters were obtained for 192 residues because 59 residues were omitted from further analysis due to peak overlap or poor signal to noise ratios. For hNOE experiments, the ratio of the saturated to unsaturated peak height was measured. R_2 values were calculated from R_1 and $R_{1\rho}$ rates according to Equation ((1)).

$$R_2 = R_{1\rho}/\sin^2\theta - R_1/\tan^2\theta \quad (1)$$

where $\tan\theta = \omega_1 / \Omega$, ω_1 is the spin lock RF field (1400 Hz at 600 MHz and 1866.7 Hz at 800 MHz) and Ω is the offset of the ^{15}N resonance of interest with respect to the ^{15}N carrier frequency (33).

Model-free analysis

Model-free analysis was performed using *relax* (34–38) on the Sheffield-WRGRID ICEBERG high performance computing cluster. Using R_1, R_2 and hNOE values at 600 MHz and 800 MHz and the coordinate geometry of backbone amide N-H bonds as provided by the crystal structure 3Q8K (11), model-free analysis was executed. However, residues from the arch region, α2–α3 loop and the C-terminus were excluded due to lack of coordinates in the crystal structures (1UL1) (22) or low R_2/R_1 values. Instead, these 16 residues were modeled with a spherical diffusion tensor. Both analyses were conducted for the 192 residues for which relaxation data at both field strengths was available; of these, only 179 were processed fully, as 13 were removed due to large errors or computational eliminations.

Titration of Ca^{2+} and Mg^{2+} cations into $hFEN1_{K93A}$

A ^{15}N-labeled hFEN1 K93A mutant ($hFEN1_{K93A}$), which hindered catalysis, was prepared in the pET29b-hFEN1-336 vector as described previously (11). The $hFEN1_{K93A}$ protein was prepared as described for the wild type above using the ^{15}N autoinduction method. ^{1}H–^{15}N TROSY spectra were separately recorded at 10 mM HEPES–KOH pH 7.5, 76 mM KCl, 0.1 mM EDTA, 4 mM NaN_3, 100 mM βME, 10% D_2O and 0.05 mM TSP with either 0 mM, 8 mM $MgCl_2$ and 8 mM $CaCl_2$ added. Absolute changes in weighted chemical shifts (ω) were determined using Equation ((2)) where the correction factor for ^{15}N was $\alpha = 0.14$.

$$\omega = \left(\Delta\delta_H{}^2 + (\alpha\Delta\delta_N)^2\right)^{1/2} \quad (2)$$

DNA preparation

DNA oligonucleotides were purchased with purification from LGC Biosearch Technologies. The sequence of the oligonucleotides are shown in Table 1. Oligonucleotide concentrations were quantified using absorbance at 260 nm and molar extinction coefficients calculated according to the 'nearest-neighbor' method (39,40).

$hFEN1_{K93A}$-DNA complex formation

DHPS1 (DNA substrate) was heated and annealed in 10 mM HEPES–KOH pH 7.5, 6 mM KCl, 4 mM NaN_3 and 0.1 mM EDTA. Initially, 0.5 mM ^{15}N-labeled $hFEN1_{K93A}$, ^{2}H,^{13}C,^{15}N-labeled $hFEN1_{K93A}$ or ^{2}H,^{15}N-labeled $hFEN1_{K93A}$, prepared as described above, in 10 mM HEPES–KOH pH 7.5, 76 mM KCl, 4 mM NaN_3 and 0.1 mM EDTA was titrated with sub-stoichiometric aliquots of lyophilized DHPS1 until the DNA was in excess of the protein (>500 μM). At 50–100 μM of DNA, the sample precipitated and the quality of the spectra decreased rapidly. To overcome protein precipitating in complex with DNA, an $hFEN1_{K93A}$-DNA complex was prepared in 10 mM HEPES–KOH pH 7.5, 6 mM KCl, 4 mM NaN_3 and 0.1 mM EDTA in the presence of excess DNA substrate (1:1.1) under dilute conditions of protein and DNA (roughly 5 μM). Low monovalent salt concentrations were used to slow dissociation of the complex. The sample was concentrated to ∼500 μM using Vivaspin 20 (10 kDa MWCO) spin columns and 100 mM βME, 10% D_2O (10% v/v) and 0.05 mM TSP were added. The same resonance assignment strategy using 3D NMR spectra was conducted with a ^{2}H,^{15}N,^{13}C-labeled $hFEN1_{K93A}$–DNA complex prepared in this manner. ^{1}H-^{15}N TROSY spectra were acquired at 0, 0.5, 1, 2, 4, 6 and 8 mM $CaCl_2$, and chemical shift differences were recorded as described in Equation (2).

Preparation of hFEN1 cysteine mutants labeled with fluorophores

Surface residues C235 and C311 were successively mutated to alanine in the pET28b-hFEN1 plasmid using Agilent site-directed mutagenesis protocols (41) and the appropriate oligonucleotides (Supplementary Table S1). The resulting double mutant hFEN1 protein (C235A/C311A) was expressed and purified as described above. This mutant did not react with maleimide dyes on a timescale of 1 h. Subsequent mutagenesis as above created $hFEN1_Q$, E120C/C235A/S293C/C311A. $hFEN1_Q$ protein was expressed and purified as described above. ESI-MS calculated: 43210.6 Da experimental: 43210.9 Da.

To stochastically label $hFEN1_Q$ with the appropriate fluorophores, 250 μl of 100 μM (25 nmol) purified $hFEN1_Q$ in 50% glycerol storage buffer was diluted 1:1 with buffer F (50 mM Tris–HCl $pH_{25°C}$ 7, 150 mM NaCl, 1 mM EDTA and 0.1 mM TCEP) to dilute the glycerol. DTT was added to the diluted $hFEN1_Q$ to a final concentration of 20 mM, and the solution allowed to incubate on ice for 15 min. The $hFEN1_Q$ solution was then subjected to size exclusion chromatography (SEC) using a Superdex 75 GL 10/300 column (GE Healthcare Life Sciences) and buffer F. Protein containing fractions were pooled and concentrated using a 10 kDa MWCO Vivaspin-20. The $hFEN1_Q$ protein concentration was quantified using the absorbance at A_{280} nm and calculated extinction coefficient (28). Stocks of 10 mM Atto 647N maleimide (Sigma-Aldrich) and 10 mM Cy3b maleimide (GE Healthcare Life Sciences) were prepared in DMSO. To label $hFEN1_Q$, the protein was added to a tube containing the appropriate volume of each stock of fluorophore maleimide conjugate to give a molar ratio of 1:2:2 of hFEN1:Atto 647N:Cy3b. Aliquots (usually 50 μl) of the reaction were removed at 2, 4, 8, 16 and 26 min and quenched by addition of excess DTT. To remove excess free fluorophore, the quenched aliquots were exchanged into buffer F using micro Bio-Spin 6 columns (BioRad) according to the manufacturer's instructions. The labeled protein was again subjected to SEC as described above. The protein containing fractions were pooled and concentrated as before. The concentrations of the protein and the covalently attached Cy3b and Atto 647N were quantified using a UV/Vis Nanodrop spectrohphotometer (ThermoFisher) at A_{280}, A_{559} and A_{646}, respectively, and the associated extinction coefficients (hFEN1 protein $\varepsilon_{280} = 22\,920$ M^{-1} cm^{-1}, Cy3b $\varepsilon_{559} = 130\,000$ M^{-1} cm^{-1} and Atto 647N $\varepsilon_{646} = 150\,000$ M^{-1} cm^{-1}). A_{280} values were corrected for contributions from both dyes using A_{280}/A_{556} and A_{280}/A_{646} ratios.

Kinetic measurements of hFEN1, mutants and fluorescently labeled proteins

The activities of hFEN1, $hFEN1_Q$ and $hFEN1_Q$ with attached fluorophores ($hFEN1_{QF}$) were assessed using a 5′-fluorescein (FAM) labeled bimolecular DNA oligo construct DF5,1*, which was prepared by heating and annealing F1* and T1 in a 1:1.1 ratio (Table 1), respectively, in 50 mM HEPES–KOH pH 7.5, 100 mM KCl and 0.02% NaN_3. Reactions mixtures (180 μl) containing 50 nM DF5,1* in reaction buffer (55 mM HEPES–KOH pH 7.5, 110 mM KCl, 8 mM $MgCl_2$, 0.1 mg ml^{-1} BSA. 1 mM DTT) were initiated by adding appropriate amounts of hFEN1 proteins (3–6 pM final concentration) that afforded less than ∼10% product formation in 10 min at 37°C. Aliquots (20 μl) of reaction were quenched in excess EDTA (50 μl of 500 mM EDTA) at 2, 4, 6, 8, 10 and 20 min. The amount of product formation was assessed by denaturing high performance liquid chromatography (dHPLC) as described previously (18,42). Plots

Table 1. Oligonucleotide sequences used herein

Name	Sequence
DHPS1	5′dTGAAAGGCAGAGCGCTAGCTCTGCCTTTCGAGCGAAGCTCC3′
*F1**	5′FAM*-dTTTTTACAAGGACTGCTCGACAC3′
F1	5′dTTTTTACAAGGACTGCTCGACAC3′
T1	5′dGTGTCGAGCAGTCCTTGTGACGACGAAGTCGTCC3′

*5′FAM is a 5′-terminal fluorescein modification produced using the single isomer 5-carboxyfluorescein-aminohexyl amidite.

of concentration of product versus time yielded the initial rate of reaction (v_0 nM min^{-1}), which could be normalized for enzyme concentration ($v_0/[E]_0$ min^{-1}).

For second order rate constant determination as a function of viscosity, reaction mixtures (180 μl) were assayed at 2.5, 5, 7.5 and 10 nM DF5,1* substrate in reaction buffer at 37°C and the indicated relative viscosity. Buffer viscosity was adjusted using either glycerol (0–36% v/v) or sucrose (0–35%). Glycerol relative viscosities and their corresponding %v/v were calculated using temperature corrected density calculations at 37°C (43). Relative viscosity was adjusted with sucrose as described previously (42). Product formation was monitored by dHPLC and normalized initial rates of reaction were generated as described above. Estimates of second order rate constants were derived from the slope of normalized initial rate (v/[E]) versus the concentration of substrate [S]. A calculated second order rate constant $(k_{cat}/K_M)_0$ value at relative viscosity of 1 was derived from the Y-intercept from a plot of k_{cat}/K_M versus relative viscosity (44).

Single molecule FRET measurements and analysis

Labeled hFEN1 ($hFEN1_{QF}$) was diluted to ∼50 pM in binding buffer (55 mM HEPES–KOH pH 7.5, 110 mM KCl, 8 mM $CaCl_2$, 0.1 mg ml^{-1} BSA, 1 mM DTT)) and incubated at room temperature for 5 min with or without 20 nM substrate DNA (i.e. DF5,1, prepared by heating and annealing F1 and T1 in a 1:1.1 ratio (Table 1)). Glass slides were passivated with 1 mg/ml BSA for 5 min prior to each measurement. Three triplicate 10-min data sets were acquired for each sample giving a total of 90 min of data for each condition, yielding ∼4000 bursts.

smFRET data were acquired using a custom built confocal microscope and alternating laser excitation (45). Two dioide lasers (515 nm and 635 nm – LuxX plus) were directly modulated (100 us, duty cycle 45%) and combined into an optical fibre. The output beam was collimated and then cropped to 2.5 mm diameter by an iris. The beam was directed into the back of the objective (Olympus UPLSAPO 60× NA = 1.35 oil immersion) using a dichroic mirror (Chroma ZT532/640 rpc 3 mm) with the fluorescence emission collected by the same objective, focussed onto a 20 um pinhole and then split (dichroic mirror: Chroma NC395323 – T640lpxr) for detection by two avalanche photodiodes (SPCM-AQRH-14 and SPCM-NIR-14, Excilitas). Photon arrival times were recorded by a national instruments card (PCIe-6353), with the acquisition controlled using custom software (LabView 7.1).

smFRET data analysis was performed using custom Jupyter notebooks, based on the open source python software, FRETBusts, described previously (46). Briefly, background rates in each channel were obtain for each 60 seconds of the acquisition by fitting a poisson distribution to inter-photon delays. This background rate was recalculated every 60 s to take account of any change in background throughout the measurement time. Bursts were identified using a dual channel burst search as previously described (47).

Apparent FRET efficiencies (E^*) and apparent stoichiometries (S^*) were calculated for each burst according to Equations (3) and (4), respectively:

$$E^* = \frac{I_{A|D}}{I_{D|D} + I_{A|D}} \quad (3)$$

$$S^* = \frac{I_{D|D} + I_{A|D}}{I_{D|D} + I_{A|D} + I_{A|A}} \quad (4)$$

where I represents the background corrected intensity in the (i) acceptor emission channel after donor excitation ($I_{A|D}$), (ii) donor emission channel after donor excitation ($I_{D|D}$) and (iii) acceptor emission channel after acceptor excitation ($I_{A|A}$). Apparent FRET efficiencies were corrected for donor leakage, acceptor direct excitation, and the detection efficiencies/quantum yields (gamma correction) according to published protocols (48) (arXiv:1710.03807), using labeled DNA standards for the gamma correction.

After correction factors were applied, bursts were selected with at least 40 photons under green excitation, and 10 photons under red excitation, with a maximum of 300 photons to filter out an acceptor heavy aggregate observed in the data set. Bursts with an observed stoichiometry between 0.5 and 0.85 were plotted to generate a histogram of relative frequency versus FRET efficiency and fitted with an unrestrained double Gaussian function.

RESULTS

NMR assignments are consistent with known secondary structure and peak intensities suggest regions with unusual dynamic properties

Human FEN1 (hFEN1) is a 380 amino acid protein that has a nuclease core domain (amino acid residues 1–336) that is sufficient for catalysis *in vitro*. As we were most interested in the dynamics associated with catalysis, the hFEN1–336 construct from X-ray crystallographic studies was used for NMR studies (11,21). Thus, the final 38 kDa hFEN1-336 construct (herein referred to as hFEN1) contains 336 native amino acids with an additional six residues present at the C-terminus from the rhinovirus 3C protease recognition site (49). Optimization of both buffer and temperature yielded stable hFEN1 samples that afforded good ^{1}H–^{15}N TROSY spectra (Figure 2A and Supplementary Figure

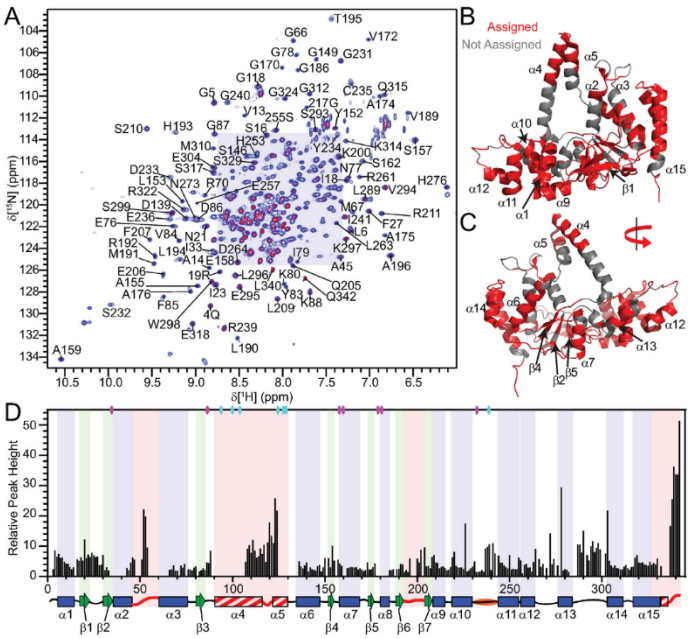

Figure 2.)) Relative peak heights suggest differences in dynamics in regions of hFEN1. (**A**) ^{1}H–^{15}N-TROSY spectrum of hFEN1 under experimental conditions. Peaks from amide side chains and the two tryptophan side-chain indole groups (H^{ε}–N^{ε}) are also observed. Expanded views of the shaded region can be found in Supplementary Figure S4. (**B**) Front and (**C**) rear views of the hFEN1 structure (3Q8K) (11) with labeled secondary structure elements and colored backbone to denote assigned (red) or unassigned (gray) residues. (**D**) Relative peak height (based on the lowest intensity peak E257) obtained from the ^{1}H–^{15}N-TROSY spectrum of hFEN1 plotted versus residue number show that loops are generally more intense and are flanked by decreasing peak intensities and sometimes missing residues. A secondary structure schematic of hFEN1 (3Q8K) is included (11). Blue rectangles, green arrows and black lines indicate α-helices, β-strands and loops, respectively. Loops known to have structural heterogeneity (1UL1 versus 3Q8K) are indicated by red lines (11,22). Red and white stripped rectangles highlight regions of structural heterogeneity. The H2tH α10–α11 loop is highlighted in orange. The location of the active site carboxylate (D34, D86, E158, E160, D179, D181 and D233) and basic residues (K93, R100, R104, K125, K128, R129 and R238) with respect to secondary structure elements are indicated by magenta and cyan ovals, respectively.

S4). Using standard ^{1}H–^{15}N TROSY and TROSY-based 3D experiments (29), 251 amino acid residues out of the 324 theoretically-detectable backbone amides (excluding prolines and initiator methionine) were assigned in the absence of divalent metal ions. Many of the unassigned residues were localized to the base of the arch region, the active site amino acids and some of the α2–α3 loop (Figure 2B and C). Furthermore, the peaks in the ^{1}H–^{15}N TROSY spectrum displayed variable intensities, suggesting that not all residues relaxed uniformly. Some of the interfaces between the assignable and unassignable regions had peaks with decreased intensity, which suggested that the missing regions were undergoing exchange on the millisecond timescale (Figure 2D).

Chemical shift analysis of protein backbone nuclei (^{15}N, $^{1}H^{N}$, $^{13}C^{\alpha}$, $^{13}C^{\beta}$ and $^{13}C'$) was conducted using TALOS-N (32). The predicted secondary structure elements agreed with the structure of the saddle region of the protein in 3Q8K and 1UL1x, thereby supporting the accuracy of our assignments (Supplementary Figure S5). Prediction of β-strand ϕ and ψ dihedral angles for residues E295–S299 were confirmed with further analysis of this region in 3Q8K and 1UL1z (11,22), which showed that the dihedral angles were consistent with the β-strand prediction despite being a loop. Furthermore, a single α-helical turn in the β1–β2 loop (residues D22–Y26) is present consistent with the TALOS-N assignment (11,22). In the arch region, the analysis for A106–A111 suggested that these residues were in an α-helical conformation. For Q112–Q115, α-helix was also predicted, but with decreasing probability. Finally, an extended loop conformation was predicted for residues A116–E124. Therefore, only the N-terminal half of the arch region was consistent with the secondary structure present in 3Q8K, whereas the C-terminal portion was not.

The observable residues of the arch region and α2–α3 loop are disordered

To characterize the dynamics of the free protein, ^{15}N spin-lattice (R_1) and spin-spin (R_2) relaxation rates and ^{15}N-{^{1}H} heteroNOE (hNOE) values were measured using

$^{2}H,^{15}N$-labeled hFEN1 (33). Of the 251 residues that were assigned, only 203 and 197 were analyzed at 600 and 800 MHz (Supplementary Figure S6A-F), respectively, due to peaks either being too weak to analyze or overlapped. The data generally showed that residues in secondary structure have similar relaxation rates, indicative of the overall rotational correlation time of the molecule (τ_c). Using R_2/R_1 ratios (Supplementary Figure S6G and H), an overall τ_c was estimated to be 25 ns (50), consistent for a protein of this size (38 kDa) (51). Loops including β6–β7 (i.e. β-pin), α10–α11, α12–α13 and α13–α14 as well as the start of α15 displayed intermediate R_2/R_1 ratios. Regions that showed particularly low R_2/R_1 ratios and low hNOE values were the α2–α3 loop, the arch region and the non-native C-terminal residues. These regions were associated with higher than average *B*-factors or lacked observable electron density in crystal structures (Figure 1D–F and Supplementary Figure S2) (11,22). Furthermore, the higher peak intensities of the α2–α3 loop, the arch region and the non-native C-terminal residues (Figure 2D) are consistent with the lower R_2/R_1 ratios (Supplementary Figure S6G and H).

We used *relax* in combination with the 3Q8K protein structure to calculate model-free order parameters (52). Residues with low R_2/R_1 ratios and a lack of density in the free protein structures were treated separately as an isotropic spherical diffusion tensor (16 residues), while the other backbone amide residues were derived using an oblate rotational diffusion tensor (163 residues) (Supplementary Figure S6I and J). Although a large range of S^2 values were generated with various models (Supplementary Table S2), most residues in the protein in secondary structure elements afforded an average S^2 of ~0.8, showing that relaxation of most residues in the saddle region of hFEN1 was influenced predominantly by the tumbling of the molecule in solution (Figure 3). Therefore, these residues were relatively rigid. The α10–α11, α12–α13 and α13–α14 loops, the C-terminal end of α10 and α14 as well as most of α15 displayed intermediate S^2 ($0.5 < S^2 < 0.8$) values indicating increased flexibility in these regions. The lowest S^2 values (<0.5) were generally observed in the C-terminal tail, arch and α2–α3 loop residues, thereby indicating that these residues were extremely flexible in solution. R_{ex} terms were reported for approximately 70% of the residues in hFEN1 that were amenable to model-free analysis (Supplementary Figure S7), indicating that millisecond timescale motions contributed to spin-spin relaxation (R_2). Most of the R_{ex} terms were associated with residues in the saddle region.

S^2 values were compared with Random Coil Index-S^2 (RCI-S^2) values predicted by TALOS-N (32) and were congruent apart from an underestimation of the degree of regional flexibility (Figure 3A). Considering the flexibility of the arch, α2–α3 loop and C-terminal tail, we decided to confirm the nature of these regions using the neighbor-corrected random coil chemical shift library for intrinsically disordered proteins (ncIDP) (53). Deviations greater than 1 ppm from ncIDP ^{15}N and ^{13}C chemical shift values were shown to be a powerful indicator of secondary structure. The observable regions of the arch, α2–α3 loop and C-terminal residues showed only small differences (≤ 1) (Figure 4A and B), suggesting that these residues were disordered in the free protein. The small deviations from the random coil value suggested that residues A107–Q113 were transiently sampling α-helical ϕ,ψ-space. The postulated preference for occupying α-helical-ϕ,ψ space showed a gradual decrease occurring from Q110–A114. For residues A114–A120, a progressive switch to extended ϕ,ψ-space was implied (53). These data agreed with AGADIR predictions (54) and TALOS-N analysis (32) (Figure 4C and Supplementary Figure S5A). Overall, these data suggested that the middle of α4 was poised to form α-helical structure, whereas the top of α5 was not.

The analysis of the sequence of the arch residues (P90–L130) using CIDER (55,56) (Classification of Intrinsically Disordered Ensemble Regions) categorized this IDR as a polyampholytic coil or hairpin. Whether a polyampholytic IDR prefers to be a hairpin collapsed on itself or an extended coil (Supplementary Figure S3C,D,F) was shown to depend on the linear sequence distribution of charged residues, which can be assessed by the value κ (57). A κ value of 0 indicates no self-assembly, whereas a κ value of 1 indicates full collapse. The arch region has a very low κ value of 0.12, which suggested that it prefers to adopt an extended coil rather than collapsed conformation in solution.

Divalent cations induce chemical shift changes in the vicinity of the hFEN1 active site

Because hFEN1 binds two Mg^{2+} metal ions in the active site to catalyze the hydrolysis of the scissile phosphate diester bond of the optimal substrate (1,2), we wanted to assess the effects of Mg^{2+} and Ca^{2+} metal ion coordination by hFEN1 using chemical shift perturbation mapping. Ca^{2+} was included in the analysis for studies of an hFEN1–DNA complex, because it facilitates the hFEN1-induced DNA conformational change required for scissile phosphate diester placement on the active site metal ions without substrate hydrolysis (9,58). Furthermore, the analysis was performed using $hFEN1_{K93A}$ because addition of Ca^{2+} alone was unable to completely inhibit phosphate diester hydrolysis, presumably due to small amounts of Mg^{2+} contamination (59). Removal of K93 by mutation to alanine reduces scissile phosphate diester hydrolysis by 2000-fold, because it acts as an electrophilic catalyst during scissile phosphate diester hydrolysis (42). Nonetheless, $hFEN1_{K93A}$ still facilitates substrate conformational change similarly to hFEN1 (9,58). The K93A mutation is in a region of hFEN1 that was unable to be assigned. Comparison of the $^{1}H–^{15}N$ TROSY spectra of hFEN1 and $hFEN1_{K93A}$ showed only small and localized changes in chemical shifts, thereby indicating that it had no widespread effect on the protein (Figure 5A).

Titration of either $MgCl_2$ or $CaCl_2$ (0–10 mM) into samples of hFEN1 was monitored using $^{1}H–^{15}N$ TROSY spectra and showed saturation between 8–10 mM for both divalent salts, consistent with the optimal concentrations of $MgCl_2$ used in routine kinetic assays. Furthermore, $^{1}H–^{15}N$ TROSY spectra of hFEN1 with 8 mM Ca^{2+} or 8 mM Mg^{2+} are nearly identical, but show small differences in residues near to the active site including the N-terminal residues (Supplementary Figure S8A) This indicated that Ca^{2+} binds in the active site, consistent with Ca^{2+} being competitive with respect to Mg^{2+} (60). The small deviations in chemical

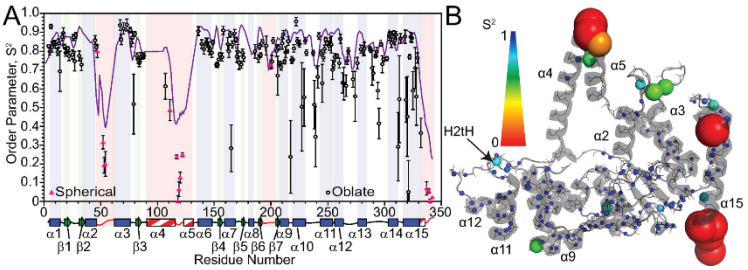

Figure 3. Model-free analysis of hFEN1 relaxation data identifies regions with increased flexibility. (**A**) Generalized order parameters (S^2) were derived from *relax* (52) using backbone ^{15}N relaxation data acquired at two field strengths and plotted versus residue number. Black circles represent data fitted to the oblate spheroid diffusion tensor, whereas pink triangles were fitted to a spherical diffusion tensor. The purple line in (**A**) represents the TALOS-N (32) random coil index S^2 prediction. Secondary structure map as described in Figure 2D. (**B**) S^2 values plotted on a cartoon depiction of the hFEN1 protein structure (3Q8K) (11). The spheres represent the nitrogen nuclei for which data were derived. The S^2 and R_{ex} spectrum bars illustrate the magnitude of S^2 and R_{ex} values with respect to sphere color and size.

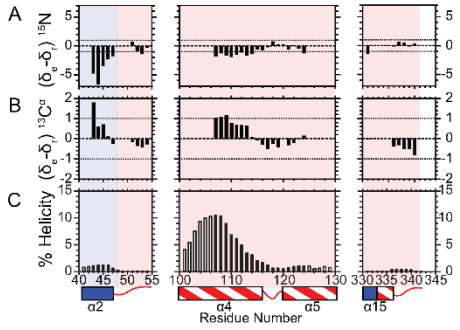

Figure 4. $^{15}N^H$ and $^{13}C^\alpha$ chemical shifts for the disordered arch region indicate that only some of the residues are sampling helical conformations. Bar graphs indicate the magnitude of the difference of experimental (δ_e) and sequence-corrected random coil chemical shifts (δ_r) for (**A**) $^{15}N^H$ and (**B**) $^{13}C^\alpha$ for the indicated regions. Differences less than one (dotted line) suggest disorder. The anti-correlated nature of $^{15}N^H$ and $^{13}C^\alpha$ is used to suggest which segments of the disordered region populate either helical or extended peptide backbone angles. Positive and negative $^{15}N^H$ and $^{13}C^\alpha$ chemical shift differences, respectively, imply extended backbone angles, whereas negative and positive $^{15}N^H$ and $^{13}C^\alpha$ chemical shift differences, respectively, suggest sampling α-helical backbone angles (53). (**C**) Agadir helical percentage predictions for the indicated residues assessed at 114 mM ionic strength, pH 7.5 and 25°C (NMR conditions) (54). Black bars indicate the residues for which NMR data is available to validate the prediction, whereas open bars indicate an absence of NMR assignments. A secondary structure map is illustrated below as described in Figure 2D.

shifts of residues closest to the active site could be due to the larger ionic radius and looser coordination geometry of Ca^{2+} compared to Mg^{2+}. Comparison of ^1H-^{15}N TROSY spectra of hFEN1$_{K93A}$ in the absence and presence of 8 mM Ca^{2+} showed similar chemical shift perturbations as observed in spectra of hFEN1 when Mg^{2+} or Ca^{2+} were added. The largest changes between apo-hFEN1$_{K93A}$ and Ca^{2+}-hFEN1$_{K93A}$ occurred in regions close to the metal chelating carboxylates (D34, D86, E158, E160, D179, E181, D233) and the N-terminal residues (Figure 5B–D). Other regions affected included β5, α3, α7, β4–α7 loop, the H2tH motif, the β3–α4 loop and the β6-β7 β-pin, which are all regions that are near active site residues. Some of the β-strands also showed smaller perturbations further from the active site that suggested that conformational rearrangement in the active site is propagated throughout the β-sheet. These data indicated that divalent ions predominantly affected the structure immediately surrounding the active site in agreement with their locations in X-ray crystal structures (11). Notably, the addition of Ca^{2+} did not change the higher relative intensities of the peaks associated with the arch region and α2–α3 loop, suggesting that Ca^{2+} did not significantly alter the dynamics of these regions.

Substrate binding changes the exchange regime for the arch region and α2–α3 loop

To prepare an hFEN1$_{K93A}$–DNA complex and to analyze NMR spectral changes upon binding, consecutive sub-stoichiometric amounts of DNA substrate (DHPS1; Table 1 and Supplementary Figure S9A) were added to an hFEN1$_{K93A}$ sample containing with 76 mM KCl until the DNA was in excess. However, addition of substrate induced protein precipitation in the sample suggesting non-specific DNA binding at concentrations used for NMR spectroscopy. Despite this, chemical shift changes upon addition of DNA were observed for a number of residues, consistent with locations known to interact with duplex DNA. Residues that were affected by DNA binding, but remained in the fast exchange regime (e.g. Q4, A175, L190, R192, R239 and G324) could be followed with the addition of DNA to saturation. Interestingly, the assignable α2–α3 loop and arch residues disappeared from the spectrum at sub-stoichiometric amounts of DNA, suggesting that they entered an exchange regime that prevented their detection (29). The magnitude of the exchange rate at which this happens at these field strengths should be faster than 10 s^{-1} but slower than 1000 s^{-1} as the peaks do not re-appear upon further addition of DNA to saturation. The switch from sub-nanosecond dynamics to millisecond exchange

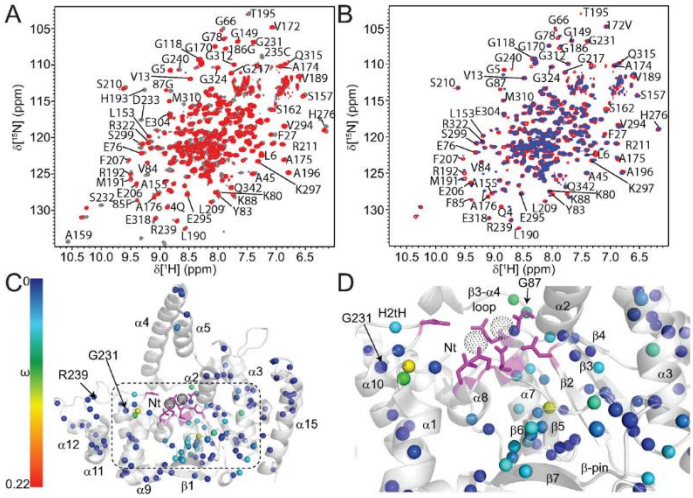

Figure 5. Addition of Ca^{2+} ions to $hFEN1_{K93A}$ perturbs chemical shifts nearest the active site. (**A**) Superposed $^{1}H–^{15}N$ TROSY spectra of hFEN1 (gray) and $hFEN1_{K93A}$ (red) shows only minimal changes in localized regions close to the mutation site indicating no widespread effects on global structure. (**B**) Superposed ^{1}H-^{15}N TROSY spectra of $hFEN1_{K93A}$ in the absence (red) and presence (blue) of Ca^{2+}. Well-dispersed residues are labeled accordingly. (**C**) Chemical shift changes observed on the addition of Ca^{2+} ions to $hFEN1_{K93A}$ are mapped onto a cartoon representation of the hFEN1 structure (3Q8K) (11). The black-dotted spheres indicate the locations of active site metal ions that are coordinated by the side chains of carboxylate residues shown as magenta sticks. Secondary structure elements, the H2tH motif and the N-terminus (Nt) are labeled. The locations of the G231 and R239 spheres are highlighted. The magnitudes of the nitrogen nuclei chemical shift perturbations (ω) are represented by sphere color according to the associated spectrum bar. The absence of a sphere either indicates the lack of assignment in the protein or the inability to follow chemical shift changes due to peak overlap. (**D**) Magnified view of the area indicated by the dashed box in panel C to highlight the location of residues most affected by the addition of Ca^{2+} ions to $hFEN1_{K93A}$. Labels are as described in panel C except for the omission of R239 and addition of G87.

suggested a conformational change of the α2–α3 loop and the arch upon DNA binding.

Divalent metal ions induce large chemical shift changes far from the active site

To overcome issues associated with the precipitation of $hFEN1_{K93A}$ when preparing an $hFEN1_{K93A}$–DNA complex (Supplementary Figure S9A), a different strategy was adopted where the complex was prepared under dilute conditions and at lower KCl concentrations in the presence of EDTA, followed by concentration. Preparing $hFEN1_{K93A}$-DNA complexes in this manner resulted in stable samples that afforded acceptable $^{1}H–^{15}N$ TROSY spectra (Supplementary Figure S8B). Comparison of $^{1}H–^{15}N$ TROSY spectra of $hFEN1_{K93A}$ and the $hFEN1_{K93A}$-DNA complex showed chemical shift changes for residues in or near areas known to interact with the substrate, such as the H2tH motif (e.g. G231 and R239), active site (e.g. Q4, G5, L6, G87, K88, A175 and R192), the hydrophobic wedge (e.g. A45 and G66,) and the 3′-flap binding pocket (e.g. M310, E318 and G324). This indicated that the solution complex was consistent with the hFEN1–DNA complex crystal structures (3Q8K and 5UM9) (11,21). Before assigning the $hFEN1_{K93A}$–DNA complex, Ca^{2+} was titrated into the sample (0–8 mM) and chemical shift perturbations upon each successive addition were monitored by $^{1}H–^{15}N$ TROSY spectra. Assignment of the $hFEN1_{K93A}$-DNA complex in 8 mM Ca^{2+} was conducted using the same methodology as for hFEN1; however, only 57% of the visible peaks (132 peaks out 230 peaks in the spectrum) could be assigned presumably due to signal attenuation in the 3D spectra. Some assignments were corroborated by following the DNA titration data (i.e. R19, G149, A175, S255, R261).

Addition of Ca^{2+} to $hFEN1_{K93A}$ and the $hFEN1_{K93A}$-DNA complex showed that residues immediately surrounding the active site were perturbed similarly (Figure 5B–D *cf.* Figure 6A and B), indicating that Ca^{2+} is coordinated in the active site in both free and bound forms. In the $hFEN1_{K93A}$–DNA complex, however, the presence of Ca^{2+} additionally resulted in the disappearance of the N-terminal residues Q4, G5 and L6 (Supplementary Figure S8C). For example, the first addition of Ca^{2+} to the complex resulted in an upfield chemical shift with peak broadening for Q4 (Figure 6C). These N-terminal residues remained observable in spectra when Ca^{2+} was added to $hFEN1_{K93A}$ (Figure 5B); thus, this unique effect was due to the presence of the DNA. Together this suggested that the addition of Ca^{2+} to the complex resulted in the N-terminal residues experiencing changes in their chemical environment due to movement of the residues themselves or movement of the DNA into the active site, thereby resulting in intermediate exchange broadening.

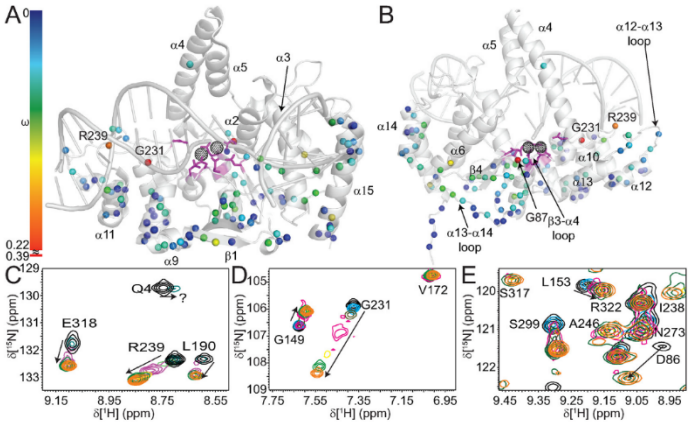

Figure 6. Addition of Ca^{2+} ions to the $hFEN1_{K93A}$–DNA complex show chemical shift perturbations (ω) throughout the protein. (**A**) Front and (**B**) rear views of a cartoon depiction of the hFEN1–DNA structure (3Q8K) (11) with the magnitudes of nitrogen nuclei chemical shift perturbations (ω) on addition of Ca^{2+} represented by sphere color according to the associated spectrum bar. Note, because the chemical shift change for G231 was larger than for most other residues, the spectrum bar is discontinuous. (C–E) Highlighted regions from Supplementary Figure S8C showing several residues with significant chemical shift perturbations that occur upon titration with Ca^{2+} at 0 mM (black), 0.5 mM (cyan), 1 mM (olive), 2 mM (magenta), 4 mM (yellow), 6 mM (green) and 8 mM (orange). The panels show a large ω for (**C**) L190, R239 and E318, (**D**) G149 and G231 and (**E**) the appearance of S317 at higher Ca^{2+} concentrations. Intermediate exchange behavior is highlighted for (C) Q4 and (E) D86.

Unlike $hFEN1_{K93A}$, the addition of Ca^{2+} to the $hFEN1_{K93A}$–DNA complex resulted in larger perturbations in regions significantly removed from the active site (Figure 5C and D *cf.* Figure 6A and B). The most pronounced weighted ^{1}H–^{15}N chemical shift change (ω) was observed for G231, which lies at the beginning of the H2tH motif. In the absence of DNA, this residue showed a slight change in chemical shift upon addition of Ca^{2+} ions (ω = 0.022 ppm, Figure 5C), whereas the change increased greatly (ω = 0.39 ppm, Figure 6D) when DNA was present. Another H2tH motif residue whose Ca^{2+} induced chemical shift change was significantly larger for the $hFEN1_{K93A}$–DNA complex was R239 (ω = 0.013 ppm, Figure 5B *cf.* ω = 0.17 ppm, Figure 6C), which lines the minor groove of the dsDNA in the reacting duplex. This showed that the H2tH DNA binding motif and residues in that region, which are approximately 20 Å from the active site metal ions, responded to the metal occupancy status of the active site in the presence of DNA. Residue G87, which is next to the divalent metal binding residue D86, also showed a larger chemical shift perturbation in response to Ca^{2+} in the $hFEN1_{K93A}$-DNA complex (ω = 0.22 ppm, Figure 6B and Supplementary Figure S8C) compared to $hFEN1_{K93A}$ (ω = 0.07 ppm, Figure 5B and Supplementary Figure S8B). In the threaded hFEN1–DNA substrate crystal structure (5UM9), residue G87, which is in the β3–α4 loop, is near the 5′-flap exit, and this larger perturbation was likely due to its proximity to the exit of the 5′-flap (Figure 1C) (21). Moreover, chemical shift changes in this region are consistent with the reported changes in the 5′-flap threading equilibrium when Ca^{2+} was added to hFEN1-DNA complexes (18). Furthermore, addition of Ca^{2+} to the $hFEN1_{K93A}$–DNA complex also gave rise to larger chemical shift perturbations than in $hFEN1_{K93A}$ in the β3–α4, α12–α13 and α13–α14 loops and helices α6, α14 and α15 (Figure 6A and B). These regions were proximal to the 5′-flap exit site and close to the 3′-flap binding pocket.

In contrast to the N-terminal residues, S317 of the $hFEN1_{K93A}$–DNA complex reappeared near the end of the Ca^{2+} titration (Figure 6E). In the hFEN1–DNA product and substrate complexes (3Q8K and 5UM9) (11,21), the S317 backbone amide group hydrogen bonds with one of the non-bridging phosphate oxygen atoms of the 3′-flap phosphate diester moiety (Supplementary Figure S1A). In ^{1}H–^{15}N TROSY spectra of $hFEN1_{K93A}$, S317 was observable in the presence and absence of Ca^{2+}; however, the peak was missing from spectra of the $hFEN1_{K93A}$-DNA complex in the absence of Ca^{2+} (Supplementary Figure S8B). The weighted chemical shift perturbation (ω) for S317 in $hFEN1_{K93A}$ and the $hFEN1_{K93A}$–DNA complex in the presence of Ca^{2+} was 0.68 ppm downfield, consistent with it forming a hydrogen bond with a non-bridging phosphate oxygen atom in the DNA. Regardless of the presence or absence of Ca^{2+}, residues associated with the arch region and α2–α3 loop were neither observable nor assignable in spectra of the $hFEN1_{K93A}$–DNA complex

Single molecule FRET measurements confirm conformational changes in the arch region

To investigate further the range of conformations adopted by FEN1 in the presence and absence of substrate DNA (Supplementary Figure S9B without the 5'-FAM label), we used single-molecule Forster Resonance Energy transfer (smFRET). Modeling of dye attachment sites for donor

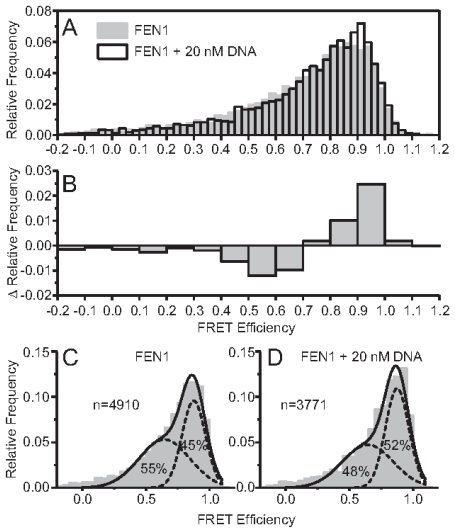

Figure 7. smFRET shows a change in conformational ensemble upon addition of DNA and reveals millisecond conformational dynamics occurring in the arch region. (**A**) Relative frequencies of FRET efficiencies for $hFEN1_{QF}$ alone (gray) and in complex with 20 nM DNA (white). (**B**) The difference in relative frequencies of FRET efficiencies upon addition of 20nM DNA. (**C**) $hFEN1_{QF}$ alone and (**D**) $hFEN1_{QF}$ with 20 nM DNA showing the unrestrained fit to the sum of two Gaussian functions (see Supplementary Table S4 for fitting parameters) to show that upon addition of DNA the lower FRET population decreases and the higher FRET population increases.

(Cy3B) and acceptor (Atto647N) dyes identified sites in the arch (E120) and the saddle region (S293) opposite the DNA binding site that would produce a FRET pair sensitive to the conformation and position of the arch (Supplementary Figure S10A). However, hFEN1 has two partially solvent accessible cysteines (C235 and C311) that easily label with maleimide dyes (data not shown). Thus, a quadruple mutant was generated by successive site-directed mutagenesis (E120C/C235A/S293C/C311A; $hFEN1_Q$) (41), and stochastic incorporation of the maleimide dyes was achieved at the desired postions to generate $hFEN1_{QF}$ (Supplementary Table S3) (61). Measurement of $hFEN1_Q$ and $hFEN1_{QF}$ activities showed that the mutations and attachment of the dyes had only a small effect on enzymatic activity (Supplementary Figure S10B). In the absence of DNA but with Ca^{2+}, a broad range of FRET efficiencies were observed (Figure 7A), indicative of multiple conformations and positions of the arch. In these experiments, single donor–acceptor labeled FEN1 molecules are observed as they diffuse through the confocal volume, a process which takes ~ 1 ms. Conformations that interconvert much faster than the observation time yield a single, tight FRET efficiency signal, whereas a broad profile is indicative of multiple conformations converting on slower (> 1 ms) time scales, as was observed here (Figure 7A).

On addition of 20 nM substrate DNA, subtle changes in the FRET efficiency distribution were observed (Figure 7A); a decrease in the relative frequency of FRET efficiencies below 0.7 with a concomitant increase in FRET efficiencies above 0.7 (Figure 7B). This is indicative of a change in the conformational ensemble for FEN1 upon binding the DNA substrate. This change in the conformational ensemble could explain the unusual viscogen dependence of hFEN1 reaction (Supplementary Figure S11). The FRET efficiency distributions were well described by the sum of two gaussians, with free fits yielding the same means for the high and low FRET populations in both the presence and absence of the DNA substrate (Supplementary Table S4). On addition of DNA, a shift in the relative amplitude from the low-FRET to the high-FRET ensemble was observed (Figure 7C and D, Supplementary Table S4), showing a change in the conformational ensemble of the arch with respect to the saddle upon binding DNA. These data indicated that a change in the conformational ensemble in the arch occurred when substrate was added.

DISCUSSION

Earlier studies have suggested that hFEN1-specificity for DNA structure rather than sequence arises from the recognition of three substrate structural features: (i) the bending of the two-way dsDNA junction, (ii) 5′-flap threading and (iii) 3′-flap binding (8–12,21). The latter two substrate-feature selection steps involve regions of the protein that we show are conformationally dynamic (i.e., the arch region and the α2–α3 loop). NMR spin relaxation data for the free protein shows that residues in the top of the arch (A107, L111, A117–E120 and E122) and the middle of the α2–α3 loop (V52–Q54) are very flexible, exhibiting motions on the timescale of $\sim 10^9$ s^{-1}; hence, they are disordered. However, residues either side of these flexible regions cannot be detected in NMR experiments presumably due to intermediate exchange broadening caused by millisecond timescale motions. Consistent with this, smFRET data for the free protein suggests that the arch region is moving relative to the saddle region of hFEN1 at a rate slower than a 1000 s^{-1} (Figure 8). The highly mobile residues in the arch and α2–α3 loop that are observable by NMR may experience this millisecond motion as well, but because these regions likely experience negligible chemical shift changes, the subnanosecond timescale motions dominate the relaxation.

The observable residues in arch and α2–α3 loop undergo a change of motional timescale from $\sim 10^9$ s^{-1} in the free protein to an intermediate exchange regime (10–1000 s^{-1}) in the hFEN1–DNA complex, suggesting their chemical environment is perturbed more significantly when bound to DNA. Such perturbation can arise from changes in rates of exchange, secondary structure formation, and changes in chemical environment due to proximity to the DNA. In line with this, the smFRET data indicate a change in the conformational ensemble upon addition of DNA. Interestingly, both the smFRET and NMR data suggest that there are still millisecond time motions involving the arch and α2–α3 loop. Thus, the earlier disorder to order model upon correct substrate binding that was based on crystallographic observations (2,11,21,62,63) is an oversimplification as it suggests

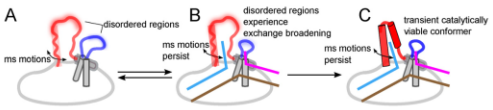

Figure 8. The flexible regions of hFEN1 respond to the appropriate structural features of the substrate, resulting in a shift to the catalytically viable ensemble. (**A**) In the absence of substrate, hFEN1 contains flexible regions in the 'arch' (red) and the α2–α3 loop (blue). The tops of these regions are disordered and display very fast motions ($\sim 10^9$ s^{-1}), whereas the flanking regions are absent presumably due to intermediate exchange broadening. The observable regions of the arch and loop are likely in fast exchange because they experience little to no change in chemical environment from the millisecond timescale motion. (**B**) Upon binding DNA, the disordered regions experience exchange broadening, likely due to a change in chemical environment and/or change in rate of motions. However, the slow millisecond timescale movements persist in both the arch and the α2–α3 loop despite the change in the conformational ensemble. (**C**) Eventually, the catalytically viable ensemble forms possibly coupling motions within the arch, the α2–α3 loop and the DNA.

there are no motions in the hFEN1–DNA complex. Rather, the complex is likely better described as an ensemble of conformations. The smFRET data shows that addition of DNA shifts the equilibrium bias of the arch to higher FRET efficiency states, implying the net position of the arch moves closer to the saddle. Interestingly, the arch has been observed in different positions relative to the saddle in hFEN1 (11,21) and hEXO1 X-ray structures (20,64). Notably, in structures where DNA has untwisted and moved into the active site (21), the arch has moved forward toward the saddle. It is possible that arch motions play an important role in the catalytic cycle, by facilitating untwisting of the DNA to transfer the scissile phosphate to the active site. Indeed, the single turnover rates of the hFEN1-catalyzed reaction are 15–30 s^{-1} (12); thus, it is possible that the slower conformational changes in the arch and α2–α3 loop regions could limit the rate of reaction.

In the hFEN1–DNA complex, residues near the active site are perturbed upon addition of Ca^{2+} as is also seen with substrate-free hFEN1. Interestingly, the N-terminal residues disappear from the spectra upon addition of Ca^{2+} to hFEN1–DNA, but not with hFEN1 alone. The recent crystal structures of the hFEN1–DNA complex (5UM9) suggests that the N-terminal amine moiety of G2 in conjunction with a metal ion and D233 activates the attacking water to generate the hydroxide ion nucleophile (21). Hence, the dynamics resulting in the disappearance of the N-terminal residues upon the addition of Ca^{2+} may be associated with this role. We also observed large chemical shift changes in response to addition of Ca^{2+} to the hFEN1–DNA complex in regions of the protein distal to the active site that were not observed at the same magnitude upon addition of Ca^{2+} to hFEN1. Several studies have shown that divalent metal ions and arch region ordering are necessary to move the scissile phosphate into the active site by untwisting (65) the reacting duplex (9,18,58,59). Thus, these large Ca^{2+} induced chemical shift changes in the H2tH motif, which directly contacts the reacting duplex 20 Å from the active site metal ions, are likely due the untwisting of the reacting duplex to move the scissile phosphate of into the active site. Remarkably, chemical shift perturbations are also propagated to the region around the 3′-flap binding pocket. The reappearance of S317 and other chemical shift changes in the 3′-flap binding region strongly suggests that the positioning of the 3′-flap duplex is altered upon the addition of Ca^{2+}. Therefore, our study provides direct evidence of DNA-mediated allostery (66) between the active site and both reacting duplex and 3′-flap recognition, thereby linking regions involved in DNA structural recognition and catalysis.

When the 5′-flap is threaded through the helical arch, the top of α4 and α5 form a mini-hydrophobic core (21). The importance of this ordered state has previously been demonstrated; it is required for untwisting of the reacting duplex DNA to place the scissile phosphate on the active site while delivering two basic residues to the active site that are essential for biologically-relevant rates of reaction (11,18,59). Chemical shift analysis (53) of the NMR-observable region corresponding to the top of α4 (A107−A116) suggests that it is transiently sampling an α-helical conformation in its disordered state, whereas the corresponding region of α5 is not. This suggests that the disorder-to-order transition for α4 arises due to conformational selection (i.e., population shift), whereas the transition in α5 results from induced fit (67–69). Having a completely induced fit model for the arch would allow for enhanced specificity but would involve large entropic penalties and slow ordering. Using a conformational ensemble shift for the larger portion of the arch (α4) reduces these penalties. A potential trigger for the ensemble population shift for the α4 region could be the catalytically-crucial contact to the +1 phosphate of the 5′-flap as well as the formation of the mini-hydrophobic core (21). On the other side of the arch, several factors are likely to induce the disorder to α-helical transition of α5. These include interaction with the α2–α3 loop when it forms an Ω-loop (15) in response to the 3′-flap and electrostatic interaction of K125, K128 and R129 with template DNA phosphate diesters (11,21). Interestingly, the distance these lysine residues are from the template phosphate diesters is shorter when the arch moves forward towards the saddle (21).

Because hFEN1 requires a disordered region that can fold in response to the appropriate ligand, the evolutionary conservation of disorder and order promoting residues to maintain the delicate balance is necessary (55,70). CIDER (56) analyses of the arch suggest that substrate-free hFEN1 has evolved to populate an extended rather than a collapsed conformation. The extended conformation is necessary to allow DNA to thread. Seemingly innocuous mutations or even interaction with small molecules may alter the conformation of the disordered state (extended coil vs. collapsed) or change the balance between disorder and order, thereby inhibiting catalysis (55). Thus, the arch region and α2–α3 loop, which are not conserved with other superfamily members, could be potential drug targets for hFEN1 (71).

Why hFEN1 has evolved to have an intrinsically disordered region and loop to achieve catalysis may seem counterintuitive initially, because enzymology has traditionally focused on structure-function paradigms (70,71). However, there is a rationale for hFEN1 evolution of intrinsically disordered regions. Firstly, because hFEN1 needs to thread

a 5′-flap through the helical arch, doing so when the arch is disordered would be easier due a larger capture radius and would provide the ability to accommodate 5′-flaps with some secondary structure. Secondly, the combination of induced fit and conformational selection with subsequent ensemble shifts for the arch allows fast ordering with low energy barriers in response to the appropriate substrate features necessary for specificity. However, there is also fast unfolding and release of incorrect binding partners (72). Thirdly, it allows for low affinity for the catalytically-relevant ensemble, which arises when the enthalpic gains in ordering are balanced by the entropic cost (70,71). Hence, specificity can be achieved while allowing the enzyme to disassociate once catalysis has occurred (i.e. turnover). Together, this work illustrates how conformational dynamic regions respond to appropriate substrate features, resulting in a shift in the ensemble towards the catalytically viable state (Figure 8).

Although all 5′-nuclease superfamily members recognize DNA junctions of some kind, each protein has its own specific DNA substrate (3). However, the use of flexible regions spanning a range of dynamic timescales for substrate recognition and control of active site assembly is likely to be a common theme in these structure-sensing nucleases. As hEXO1 also uses the threading mechanism to enforce substrate free 5′-ends (19,20) and requires a helical arch for active site assembly, it too is likely to use changes in protein dynamics for substrate accommodation and catalysis. In contrast, the Holiday junction resolvase GEN1 lacks the arch feature. A much shorter peptide linker that is the equivalent in GEN1 is not observable in current structures (23,24), despite containing superfamily conserved active site residues. This is the predicted dimerization interface, as two GEN1 monomers come together to sense the four-way-junction. Similarly, the very large equivalent of the arch in XPG, which also contains active site residues, presumably assembles as a result of the binding of other nucleotide excision repair protein partners and DNA (25). Thus, somewhat surprisingly, protein reversible plasticity appears to be the answer to coupling recognition with reaction for a range of nucleic acid hydrolases.

DATA AVAILABILITY

The backbone ^{1}H, ^{15}N and ^{13}C chemical shifts of the free protein in EDTA and the hFEN1$_{K93A}$–DNA complex in Ca^{2+} have been deposited in the BioMagResBank (http://www.bmrb.wisc.edu/) under the BMRB accession codes 27160 and 27404, respectively.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The author wish to acknowledge Dr Susan Tsutakawa for assistance with the graphical abstract.
Author Contributions: I.A.B. and L.D.F. prepared and purified isotopically-enriched protein samples and their DNA complexes. I.A.B., L.D.F., A.M.H. and N.J.B. conducted NMR spectroscopic measurements for assignment purposes, relaxation measurements and divalent metal titrations. I.A.B., L.D.F. and A.M.H. assigned the protein. I.A.B., J.C.E. and N.N.B.Md.S. performed enzyme kinetic measurements. I.A.B., L.D.F. and M.J.T. performed site-directed mutagenesis, protein purification and fluorophore maleimide attachment to hFEN1 for single molecule FRET measurements. I.A.B., B.A. and T.D.C. performed single molecule measurements. I.A.B., L.D.F., N.J.B., B.A., T.D.C., J.P.W. and J.A.G. contributed to the interpretation of the results and preparation of the manuscript. L.D.F., J.P.W. and J.A.G. conceived and coordinated the project. I.A.B., L.D.F., B.A., T.D.C. and J.A.G. prepared the figures and wrote the manuscript.

FUNDING

Biotechnology and Biological Sciences Research Council [BB/K009079/1 and BB/M00404X/1 to J.A.G.]; Biotechnology and Biological Sciences Research Council White Rose Doctoral Training Program in Mechanistic Biology Studentship [BB/J014443/1 to I.A.B.]. B.A. thanks the Engineering and Physical Sciences Research Council for his PhD stipend. N.N.B.Md.S. thanks the Public Service Department of Malaysia, Human Capitol Development Division for her PhD stipend. Funding for open access charge: The University Publication Fund of The University of Sheffield from RCUK.
Conflict of interest statement. None declared.

REFERENCES

1. Balakrishnan,L. and Bambara,R.A. (2013) Flap endonuclease 1. *Annu. Rev. Biochem.*, **82**, 119–138.
2. Finger,L.D., Atack,J.M., Tsutakawa,S., Classen,S., Tainer,J., Grasby,J. and Shen,B. (2012) In: MacNeill,S (ed). *The Eukaryotic Replisome: A Guide to Protein Structure and Function*. Springer, Dordrecht, pp. 301–326.
3. Grasby,J.A., Finger,L.D., Tsutakawa,S.E., Atack,J.M. and Tainer,J.A. (2012) Unpairing and gating: sequence-independent substrate recognition by FEN superfamily nucleases. *Trends Biochem. Sci.*, **37**, 74–84.
4. Zheng,L., Jia,J., Finger,L.D., Guo,Z., Zer,C. and Shen,B. (2011) Functional regulation of FEN1 nuclease and its link to cancer. *Nucleic Acids Res.*, **39**, 781–794.
5. van Pel,D.M., Barrett,I.J., Shimizu,Y., Sajesh,B.V., Guppy,B.J., Pfeifer,T., McManus,K.J. and Hieter,P. (2013) An evolutionarily conserved synthetic lethal interaction network identifies FEN1 as a broad-spectrum target for anticancer therapeutic development. *PLoS Genet.*, **9**, e1003254.
6. Ward,T.A., McHugh,P.J. and Durant,S.T. (2017) Small molecule inhibitors uncover synthetic genetic interactions of human flap endonuclease 1 (FEN1) with DNA damage response genes. *PLoS ONE*, **12**, e0179278.
7. Exell,J.C., Thompson,M.J., Finger,L.D., Shaw,S.J., Debreczeni,J., Ward,T.A., McWhirter,C., Sioberg,C.L., Martinez Molina,D., Abbott,W.M. *et al.* (2016) Cellularly active N-hydroxyurea FEN1 inhibitors block substrate entry to the active site. *Nat. Chem. Biol.*, **12**, 815–821.
8. Rashid,F., Harris,P.D., Zaher,M.S., Sobhy,M.A., Joudeh,L.I., Yan,C., Piwonski,H., Tsutakawa,S.E., Ivanov,I., Tainer,J.A. *et al.* (2017) Single-molecule FRET unveils induced-fit mechanism for substrate selectivity in flap endonuclease 1. *eLife*, **6**, e21884.
9. Algasaier,S.I., Exell,J.C., Bennet,I.A., Thompson,M.J., Gotham,V.J.B., Shaw,S.J., Craggs,T.D., Finger,L.D. and Grasby,J.A. (2016) DNA and protein requirements for substrate conformational changes necessary for human flap endonuclease-1-catalyzed reaction. *J. Biol. Chem.*, **291**, 8258–8268.

10. Chapados,B.R., Hosfield,D.J., Han,S., Qiu,J., Yelent,B., Shen,B. and Tainer,J.A. (2004) Structural basis for FEN-1 substrate specificity and PCNA-mediated activation in DNA replication and repair. *Cell*, **116**, 39–50.
11. Tsutakawa,S.E., Classen,S., Chapados,B.R., Arvai,A.S., Finger,L.D., Guenther,G., Tomlinson,C.G., Thompson,P., Sarker,A.H., Shen,B.I. *et al.* (2011) Human flap endonuclease structures, DNA double-base flipping, and a unified understanding of the FEN1 superfamily. *Cell*, **145**, 198–211.
12. Finger,L.D., Blanchard,M.S., Theimer,C.A., Sengerová,B., Singh,P., Chavez,V., Liu,F., Grasby,J.A. and Shen,B. (2009) The 3′-flap pocket of human flap endonuclease 1 is critical for substrate binding and catalysis. *J. Biol. Chem.*, **284**, 22184–22194.
13. Kao,H.I., Henricksen,L.A., Liu,Y. and Bambara,R.A. (2002) Cleavage specificity of Saccharomyces cerevisiae flap endonuclease 1 suggests a double-flap structure as the cellular substrate. *J. Biol. Chem.*, **277**, 14379–14389.
14. Lyamichev,V., Brow,M.A., Varvel,V.E. and Dahlberg,J.E. (1999) Comparison of the 5′ nuclease activities of taq DNA polymerase and its isolated nuclease domain. *Proc. Natl. Acad. Sci. U.S.A.*, **96**, 6143–6148.
15. Fetrow,J.S. (1995) Omega loops: nonregular secondary structures significant in protein function and stability. *FASEB J.*, **9**, 708–717.
16. Papaleo,E., Saladino,G., Lambrughi,M., Lindorff-Larsen,K., Gervasio,F.L. and Nussinov,R. (2016) The Role of Protein Loops and Linkers in Conformational Dynamics and Allostery. *Chem. Rev.*, **116**, 6391–6423.
17. AlMalki,F.A., Flemming,C.S., Zhang,J., Feng,M., Sedelnikova,S.E., Ceska,T., Rafferty,J.B., Sayers,J.R. and Artymiuk,P.J. (2016) Direct observation of DNA threading in flap endonuclease complexes. *Nat. Struct. Mol. Biol.*, **23**, 640–646.
18. Patel,N., Atack,J.M., Finger,L.D., Exell,J.C., Thompson,P., Tsutakawa,S., Tainer,J.A., Williams,D.M. and Grasby,J.A. (2012) Flap endonucleases pass 5′-flaps through a flexible arch using a disorder-thread-order mechanism to confer specificity for free 5′-ends. *Nucleic Acids Res.*, **40**, 4507–4519.
19. Shaw,S.J., Finger,L.D. and Grasby,J.A. (2017) Human Exonuclease 1 Threads 5′-Flap Substrates through Its Helical Arch. *Biochemistry*, **56**, 3704–3707.
20. Shi,Y., Hellinga,H.W. and Beese,L.S. (2017) Interplay of catalysis, fidelity, threading, and processivity in the exo- and endonucleolytic reactions of human exonuclease I. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, 6010–6015.
21. Tsutakawa,S.E., Thompson,M.J., Arvai,A.S., Neil,A.J., Shaw,S.J., Algasaier,S.I., Kim,J.C., Finger,L.D., Jardine,E., Gotham,V.J.B. *et al.* (2017) Phosphate steering by Flap Endonuclease 1 promotes 5′-flap specificity and incision to prevent genome instability. *Nat. Commun.*, **8**, 15855.
22. Sakurai,S., Kitano,K., Yamaguchi,H., Hamada,K., Okada,K., Fukuda,K., Uchida,M., Ohtsuka,E., Morioka,H. and Hakoshima,T. (2005) Structural basis for recruitment of human flap endonuclease 1 to PCNA. *EMBO J.*, **24**, 683–693.
23. Lee,S.H., Princz,L.N., Klugel,M.F., Habermann,B., Pfander,B. and Biertumpfel,C. (2015) Human Holliday junction resolvase GEN1 uses a chromodomain for efficient DNA recognition and cleavage. *Elife*, **4**, e12256.
24. Liu,Y., Freeman,A.D., Declais,A.C., Wilson,T.J., Gartner,A. and Lilley,D.M. (2015) Crystal Structure of a Eukaryotic GEN1 Resolving Enzyme Bound to DNA. *Cell Rep.*, **13**, 2565–2575.
25. Miętus,M., Nowak,E., Jaciuk,M., Kustosz,P., Studnicka,J. and Nowotny,M. (2014) Crystal structure of the catalytic core of Rad2: insights into the mechanism of substrate binding. *Nucleic Acids Res.*, **42**, 10762–10775.
26. Studier,F.W. (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.*, **41**, 207–234.
27. Sivashanmugam,A., Murray,V., Cui,C., Zhang,Y., Wang,J. and Li,Q. (2009) Practical protocols for production of very high yields of recombinant proteins using Escherichia coli. *Protein Sci.*, **18**, 936–948.
28. Gasteiger,E., Gattiker,A., Hoogland,C., Ivanyi,I., Appel,R.D. and Bairoch,A. (2003) ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.*, **31**, 3784–3788.
29. Cavanagh,J., Fairbrother,W.J., Palmer,A.G. III, Rance,M. and Skelton,N.J. (2007) *Protein NMR Spectroscopy Principles and Practice*. 2nd edn. Elsevier Academic Press, London.
30. Vranken,W.F., Boucher,W., Stevens,T.J., Fogh,R.H., Pajon,A., Llinas,M., Ulrich,E.L., Markley,J.L., Ionides,J. and Laue,E.D. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins*, **59**, 687–696.
31. Reed,M.A.C., Hounslow,A.M., Sze,K.H., Barsukov,I.G., Hosszu,L.L.P., Clarke,A.R., Craven,C.J. and Waltho,J.P. (2003) Effects of domain dissection on the folding and stability of the 43 kDa protein PGK probed by NMR. *J. Mol. Biol.*, **330**, 1189–1201.
32. Shen,Y. and Bax,A. (2013) Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J. Biomol. NMR*, **56**, 227–241.
33. Lakomek,N.A., Ying,J. and Bax,A. (2012) Measurement of ^{15}N relaxation rates in perdeuterated proteins by TROSY-based methods. *J. Biomol. NMR*, **53**, 209–221.
34. d'Auvergne,E.J. and Gooley,P.R. (2003) The use of model selection in the model-free analysis of protein dynamics. *J. Biomol. NMR*, **25**, 25–39.
35. d'Auvergne,E.J. and Gooley,P.R. (2006) Model-free model elimination: a new step in the model-free dynamic analysis of NMR relaxation data. *J. Biomol. NMR*, **35**, 117–135.
36. d'Auvergne,E.J. and Gooley,P.R. (2007) Set theory formulation of the model-free problem and the diffusion seeded model-free paradigm. *Mol. Biosyst.*, **3**, 483–494.
37. d'Auvergne,E.J. and Gooley,P.R. (2008) Optimisation of NMR dynamic models II. A new methodology for the dual optimisation of the model-free parameters and the Brownian rotational diffusion tensor. *J. Biomol. NMR*, **40**, 121–133.
38. d'Auvergne,E.J. and Gooley,P.R. (2008) Optimisation of NMR dynamic models I. Minimisation algorithms and their performance within the model-free and Brownian rotational diffusion spaces. *J. Biomol. NMR*, **40**, 107–119.
39. Cantor,C.R., Warshaw,M.M. and Shapiro,H. (1970) Oligonucleotide interactions. 3. Circular dichroism studies of the conformation of deoxyoligonucleotides. *Biopolymers*, **9**, 1059–1077.
40. Cavaluzzi,M.J. and Borer,P.N. (2004) Revised UV extinction coefficients for nucleoside-5′-monophosphates and unpaired DNA and RNA. *Nucleic Acids Res.*, **32**, e13.
41. Braman,J., Papworth,C. and Greener,A. (1996) Site-directed mutagenesis using double-stranded plasmid DNA templates. *Methods Mol. Biol.*, **57**, 31–44.
42. Sengerova,B., Tomlinson,C., Atack,J.M., Williams,R., Sayers,J.R., Williams,N.H. and Grasby,J.A. (2010) Bronsted analysis and rate-limiting steps for the T5 flap endonuclease catalyzed hydrolysis of exonucleolytic substrates. *Biochemistry*, **49**, 8085–8093.
43. Cheng,N.-S. (2008) Formula for the viscosity of a Glycerol−Water mixture. *Ind. Eng. Chem. Res.*, **47**, 3285–3288.
44. Brouwer,A.C. and Kirsch,J.F. (1982) Investigation of diffusion-limited rates of chymotrypsin reactions by viscosity variation. *Biochemistry*, **21**, 1302–1307.
45. Kapanidis,A.N., Lee,N.K., Laurence,T.A., Doose,S., Margeat,E. and Weiss,S. (2004) Fluorescence-aided molecule sorting: analysis of structure and interactions by alternating-laser excitation of single molecules. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 8936–8941.
46. Ingargiola,A., Lerner,E., Chung,S., Weiss,S. and Michalet,X. (2016) FRETBursts: an open source toolkit for analysis of freely-diffusing single-molecule FRET. *PLoS One*, **11**, e0160716.
47. Nir,E., Michalet,X., Hamadani,K.M., Laurence,T.A., Neuhauser,D., Kovchegov,Y. and Weiss,S. (2006) Shot-noise limited single-molecule FRET histograms: comparison between theory and experiments. *J. Phys. Chem. B*, **110**, 22103–22124.
48. Hohlbein,J., Craggs,T.D. and Cordes,T. (2014) Alternating-laser excitation: single-molecule FRET and beyond. *Chem. Soc. Rev.*, **43**, 1156–1171.
49. Cordingley,M.G., Callahan,P.L., Sardana,V.V., Garsky,V.M. and Colonno,R.J. (1990) Substrate requirements of human rhinovirus 3C protease for peptide cleavage in vitro. *J. Biol. Chem.*, **265**, 9062–9065.
50. Kay,L.E., Torchia,D.A. and Bax,A. (1989) Backbone dynamics of proteins as studied by ^{15}N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry*, **28**, 8972–8979.

51. Rossi,P., Swapna,G.V., Huang,Y.J., Aramini,J.M., Anklin,C., Conover,K., Hamilton,K., Xiao,R., Acton,T.B., Ertekin,A. *et al.* (2010) A microscale protein NMR sample screening pipeline. *J. Biomol. NMR*, **46**, 11–22.
52. Bieri,M., d'Auvergne,E.J. and Gooley,P.R. (2011) relaxGUI: a new software for fast and simple NMR relaxation data analysis and calculation of ps-ns and mus motion of proteins. *J. Biomol. NMR*, **50**, 147–155.
53. Tamiola,K., Acar,B. and Mulder,F.A. (2010) Sequence-specific random coil chemical shifts of intrinsically disordered proteins. *J. Am. Chem. Soc.*, **132**, 18000–18003.
54. Lacroix,E., Viguera,A.R. and Serrano,L. (1998) Elucidating the folding problem of alpha-helices: local motifs, long-range electrostatics, ionic-strength dependence and prediction of NMR parameters. *J. Mol. Biol.*, **284**, 173–191.
55. Das,R.K., Ruff,K.M. and Pappu,R.V. (2015) Relating sequence encoded information to form and function of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.*, **32**, 102–112.
56. Holehouse,A.S., Das,R.K., Ahad,J.N., Richardson,M.O. and Pappu,R.V. (2017) CIDER: Resources to analyze Sequence-Ensemble relationships of intrinsically disordered proteins. *Biophys. J.*, **112**, 16–21.
57. Das,R.K. and Pappu,R.V. (2013) Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 13392–13397.
58. Finger,L.D., Patel,N., Beddows,A., Ma,L., Exell,J.C., Jardine,E., Jones,A.C. and Grasby,J.A. (2013) Observation of unpaired substrate DNA in the flap endonuclease-1 active site. *Nucleic Acids Res.*, **41**, 9839–9847.
59. Patel,N., Exell,J.C., Jardine,E., Ombler,B., Finger,L.D., Ciani,B. and Grasby,J.A. (2013) Proline scanning mutagenesis reveals a role for the flap endonuclease-1 helical cap in substrate unpairing. *J. Biol. Chem.*, **288**, 34239–34248.
60. Syson,K., Tomlinson,C., Chapados,B.R., Sayers,J.R., Tainer,J.A., Williams,N.H. and Grasby,J.A. (2008) Three metal ions participate in the reaction catalyzed by T5 flap endonuclease. *J. Biol. Chem.*, **283**, 28741–28746.
61. Kim,Y., Ho,S.O., Gassman,N.R., Korlann,Y., Landorf,E.V., Collart,F.R. and Weiss,S. (2008) Efficient site-specific labeling of proteins via cysteines. *Bioconjug. Chem.*, **19**, 786–791.
62. Tsutakawa,S.E., Lafrance-Vanasse,J. and Tainer,J.A. (2014) The cutting edges in DNA repair, licensing, and fidelity: DNA and RNA repair nucleases sculpt DNA to measure twice, cut once. *DNA Repair (Amst.)*, **19**, 95–107.
63. Tsutakawa,S.E. and Tainer,J.A. (2012) Double strand binding-single strand incision mechanism for human flap endonuclease: implications for the superfamily. *Mech. Ageing. Dev.*, **133**, 195–202.
64. Orans,J., McSweeney,Elizabeth A., Iyer,Ravi R., Hast,Michael A., Hellinga,Homme W., Modrich,P. and Beese,Lorena S. (2011) Structures of human exonuclease 1 DNA complexes suggest a unified mechanism for nuclease family. *Cell*, **145**, 212–223.
65. Kannan,S., Kohlhoff,K. and Zacharias,M. (2006) B-DNA under stress: over- and untwisting of DNA during molecular dynamics simulations. *Biophys. J.*, **91**, 2956–2965.
66. Chaires,J.B. (2008) Allostery: DNA does it, too. *ACS Chem. Biol.*, **3**, 207–209.
67. Boehr,D.D., Nussinov,R. and Wright,P.E. (2009) The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.*, **5**, 789–796.
68. Csermely,P., Palotai,R. and Nussinov,R. (2010) Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends Biochem. Sci.*, **35**, 539–546.
69. Ma,B. and Nussinov,R. (2010) Enzyme dynamics point to stepwise conformational selection in catalysis. *Curr. Opin. Chem. Biol.*, **14**, 652–659.
70. Oldfield,C.J. and Dunker,A.K. (2014) Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu. Rev. Biochem.*, **83**, 553–584.
71. Habchi,J., Tompa,P., Longhi,S. and Uversky,V.N. (2014) Introducing protein intrinsic disorder. *Chem. Rev.*, **114**, 6561–6588.
72. Nussinov,R., Ma,B. and Tsai,C.J. (2014) Multiple conformational selection and induced fit events take place in allosteric propagation. *Biophys. Chem.*, **186**, 22–30.

SUPPLEMENTARY DATA

Regional Intrinsic Disorder Couples Substrate Specificity and Scissile Phosphate Diester Selectivity in Human Flap Endonuclease 1

Ian A. Bennet, L. David Finger, Nicola J. Baxter, Benjamin Ambrose, Andrea M. Hounslow, Mark J. Thompson, Jack C. Exell, Nur Nazihah B. Md. Shahari, Timothy D. Craggs, Jonathan P. Waltho, Jane A. Grasby

Table S1
Oligonucleotide* sequences used for site directed mutagenesis of pET28b-hFEN1

NAME	SEQUENCE
C235A_F	5'gctaggcagtgactacgctgagagtatccggggt3'
C235A_R	5'accccggatactctcagcgtagtcactgcctagc3'
C311A_F	5'aagagctgatcaagttcatggctggtgaaaagcagttctctg3'
C311A_R	5'cagagaactgcttttcaccagccatgaacttgatcagctctt3'
E120C_F	5'aatttttccacctcctggcaggccccagcagcctgag3'
E120C_R	5'ctcaggctgctggggcctgccaggaggtggaaaaatt3'
S293C_F	5'agctccacacactctgggtccagcacc3'
S293C_R	5'ggtgctggacccagagtgtgtggagct3'

*Oligonucleotides were purchased from ThermoFisher without purification.

Table S2 (related to Figure 3)
Model selection for all 179 analyzed residues

Models	Number of Residues selected for each model
$m1 = \{S^2\}$	21
$m2 = \{S^2, \tau_e\}$	11
$m3 = \{S^2, R_{ex}\}$	62
$m4 = \{S^2, \tau_e, R_{ex}\}$	24
$m5 = \{S^2, S^2_f, \tau_s\}$	18
$m6 = \{S^2, \tau_f, S^2_f, \tau_s\}$	2
$m7 = \{S^2, S^2_f, \tau_s, R_{ex}\}$	28
$m8 = \{S^2, \tau_f, S^2_f, \tau_s, R_{ex}\}$	13
TOTAL	**179**

Table S3
Efficiency of fluorophore-maleimide labelling of hFEN1$_D$ determined using UV/Vis spectrophotometry

Wavelength λ	$A_{\lambda max}$	$A_{\lambda max\ Corrected}$	Concentration#	[Chromophore]/[Protein] Ratio
280 nm	2.16	1.52*	66.19	1.00
559 nm	6.23	6.23‡	47.96	0.72
646 nm	7.05	7.05	47.00	0.71

*Corrected for Cy3b and ATTO 647N absorbtion at 280 nm

‡Corrected for ATTO 647N absorbtion at 559 nm

#Extinction coefficients used (ε_{280} = 2.292×10^4 M^{-1}cm^{-1}, ε_{559} = 1.3×10^5 M^{-1}cm^{-1}, ε_{646} = 1.5×10^5 M^{-1}cm^{-1})

Table S4
Parameters from fitting the smFRET data to a two Gaussian fit model (related to Figure 7)

Gauss Curve	Mean E	Sigma	Amplitude
FEN1 Alone low FRET	0.6455	0.1989	0.5304
FEN1 Alone high FRET	0.8762	0.0924	0.4428
+20nM DNA low FRET	0.6450	0.1990	0.4732
+20nM DNA high FRET	0.8760	0.0920	0.5044

Figure S1 (Related to Figure 1)

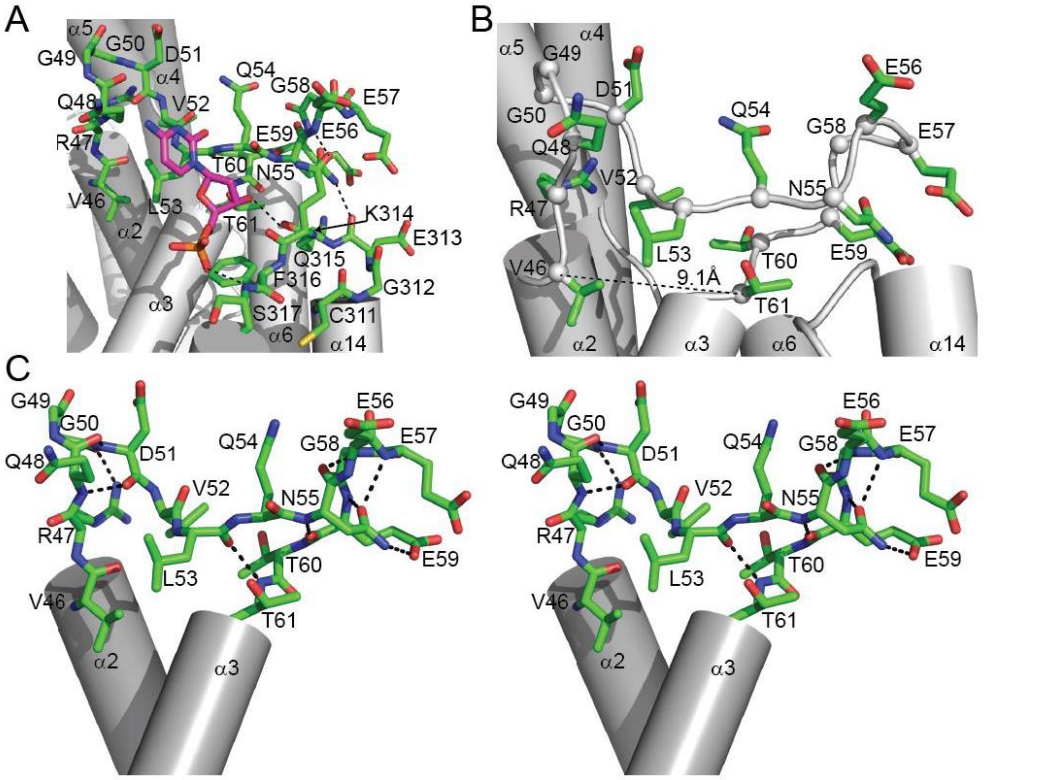

Figure S1. The 3'-flap is bound in a pocket made by the α14-α15 loop, the first turn of α15 and the α2-α3 loop, which conforms to the definition of an omega (Ω)-loop. (**A**) View of 3'-flap nucleotide of the DNA substrate (magenta) (3Q8K) cradled in its pocket (1). Residue identities and α-helices are indicated. Dashed lines indicate the hydrogen bonds formed between the α14–α15 and α2–α3 loops and the α14–α15 loop and the 3'-flap nucleotide. The N55 side chain amide forms a hydrogen bond with the E313 backbone carbonyl. The backbone amide of E56 forms a hydrogen bond with the side chain carbonyl of Q315. The backbone carbonyl of K314 forms a hydrogen bond with the 3'-hydroxyl moiety of the 3'-flap nucleotide. The backbone amide of S317 forms a hydrogen bond with a non-bridging phosphoryl oxygen atom of the phosphate diester moiety of the 3'-flap nucleotide. (**B**) Similar view as in panel A with the 3'-flap nucleotide and α14–α15 loop omitted for clarity. The traditional definition of an Ω-loop is a loop of 6 to 16 amino acid residues with the C^{α}–C^{α} distance of the loop termini both less than 10 Å (dashed line V46–T61 = 9.1 Å) and less than two-thirds the longest pairwise C^{α}–C^{α} distance in the loop (V46–E57 and G50–E57 = 18.4 Å). (**C**) Wall-eyed stereo-view of the α2-α3 loop to show the intra-loop network of hydrogen bonds among backbone and sidechain atoms when structured as an Ω-loop (2).

Figure S2 (Related to Figure 1)

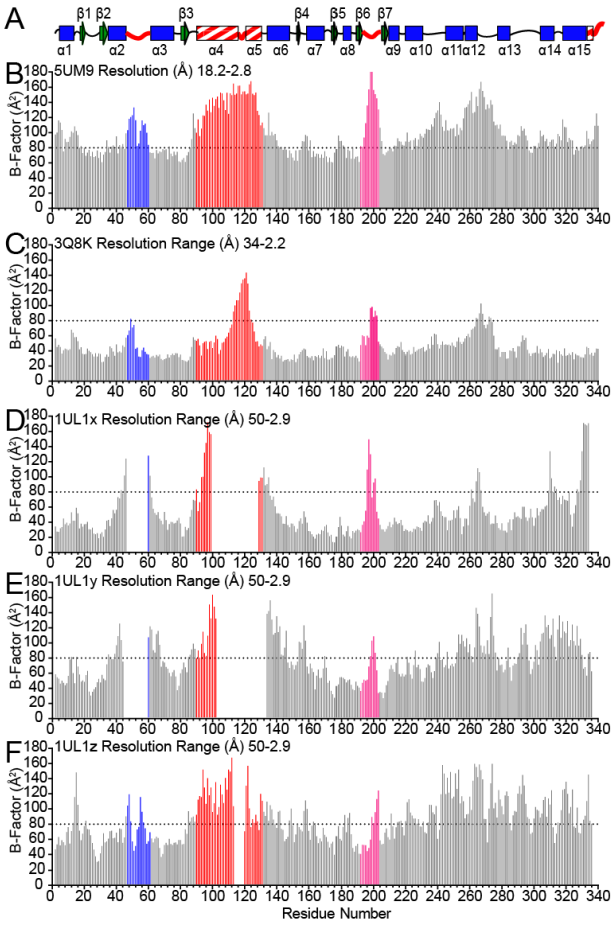

Figure S2. The α2-α3 loop, arch region, β-pin and the C-terminal residues consistently show higher than average temperature-factors (or B-factors). (**A**) Secondary structure schematic of hFEN1 from 3Q8K. Blue rectangles, green arrows and black lines indicate α-helices, β-strands, and loops, respectively. Loops known to have structural heterogeneity (1UL1 vs. 3Q8K) are indicated by red lines. Red and white stripped rectangles highlight regions where either α-helix or disorder has been observed. Plots of backbone amide nitrogen B-factors versus residue number for (**B**) 5UM9 (3), (**C**) 3Q8K (1), (**D**) 1UL1x, (**E**) 1UL1y and (**F**) 1UL1z (4). The α2-α3 loop, arch region, and β-pin are highlighted in blue, red and magenta, respectively. The dashed line indicates a B-factor of 80 Å^2, which is equivalent to a root mean square displacement of 1 Å.

Figure S3 (Related to Figure 1)

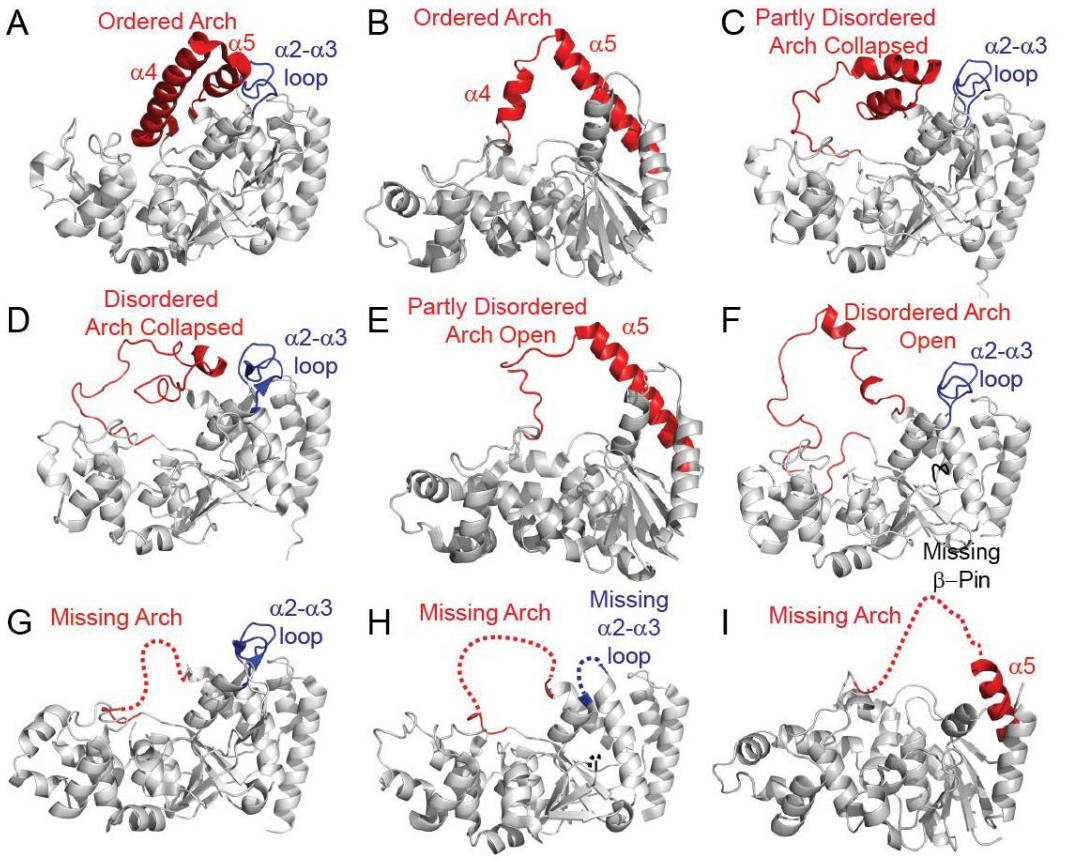

Figure S3. The arch regions and α2-α3 loops of FEN1 proteins from other organisms also show structural heterogeneity. Crystal structures of (**A**) *Desulfurococcus solfataricus* (3ORY) (7) and (**B**) several T5 FEN (1EXN, 1UT5 and 1UT8) (5,6) show an ordered arch region and α2-α3 loop. Two crystal structures of FEN1 from (**C**) *Pyrococcus furiosus* (1B43) (8) and (**D**) *Pyrococcus horikoshii* (1MC8) (9) show an ordered α2–α3 loop and an arch region with various degrees of secondary structure and random coil present, but it is collapsed over the active site. Crystal structures from (**E**) T5 FEN (5HML and 5HMM) (10) and (**F**) *Methanococcus jannaschii* (1A76 and 1A77) (11) show an ordered α2–α3 loop and an arch region with partly disordered areas, some visible secondary structure and an extended, open conformation. (**G**) The crystal structure of *Methanopyrus kandleri* (4WA8) (12) shows an ordered α2–α3 loop, but the arch region is completely missing suggesting disorder in these segments. The crystal structures of hFEN1 in complex with (**H**) an inhibitor (5FV7) (13) and (**i**) T5 FEN (1XO1) (14) lack density for both the arch region and the α2–α3 loop.

Figure S4 (Related to Figure 2)

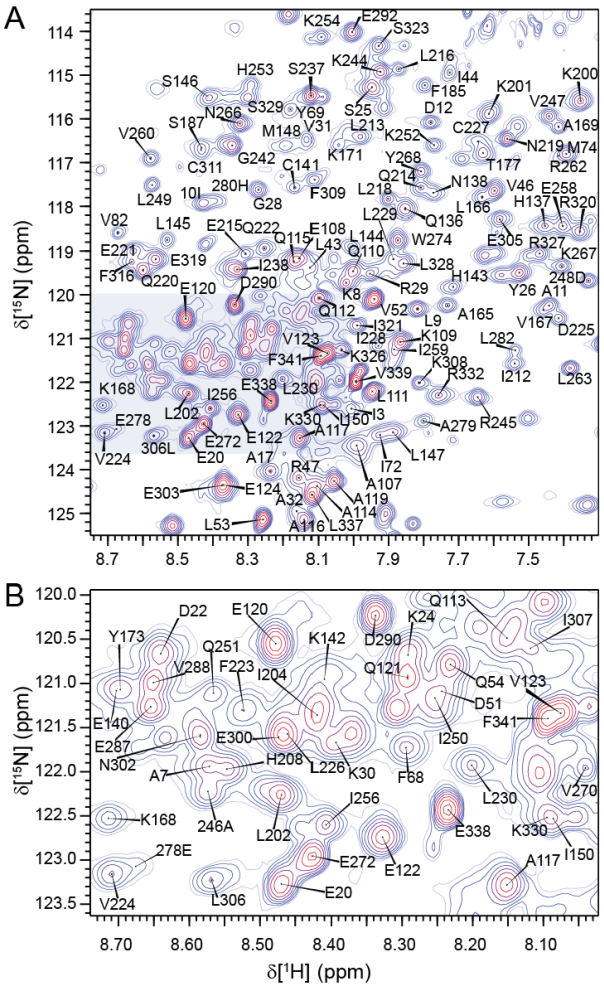

Figure S4. Expanded views of the crowded regions of the hFEN1 ^{1}H-^{15}N TROSY spectrum. (**A**) Expanded view of the shaded area in Figure 2A with assignments shown. (**B**) An expanded view of the shaded area in panel A with resonance assignments shown.

Figure S5 (Related to Figures1 and 2)

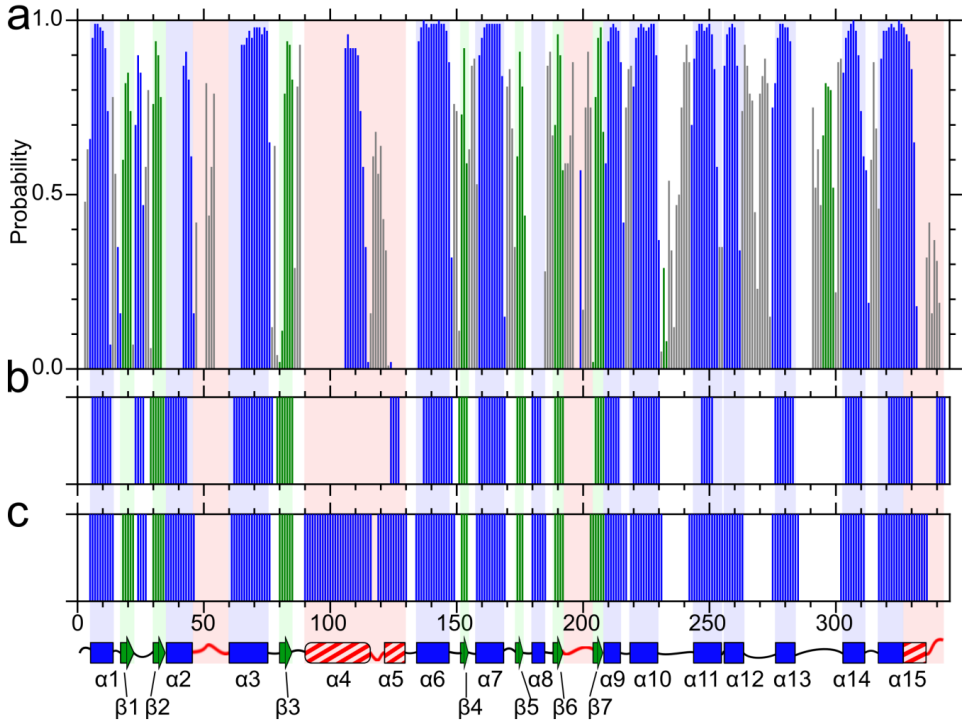

Figure S5. Secondary structure prediction for hFEN1 using backbone chemical shifts corresponds with the X-ray crystal structure of hFEN1 for the saddle region and the α2-α3 loop, but reveals differences in the arch region. (**A**) TALOS-N (15) secondary structure prediction based on assigned backbone chemical shifts ($^{1}H^{N}$, ^{15}N, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, $^{13}C'$) for hFEN1. Predicted α-helices, β-strands and loops are shown as blue, green and gray bars, respectively. Absence of a bar in panel (**a**) indicates that the residue was not assigned. Secondary structure assignments for (**B**) 1UL1z (4) and (**c**) 3Q8K (1). Well-characterized α-helices and β-strands are shown, but loops are indicated by the absence of a bar in the panels (**B,C**). Secondary structure schematic derived from 3Q8K as described Figure S2A is included at the bottom of the figure.

Figure S6 (Related to Figure 3)

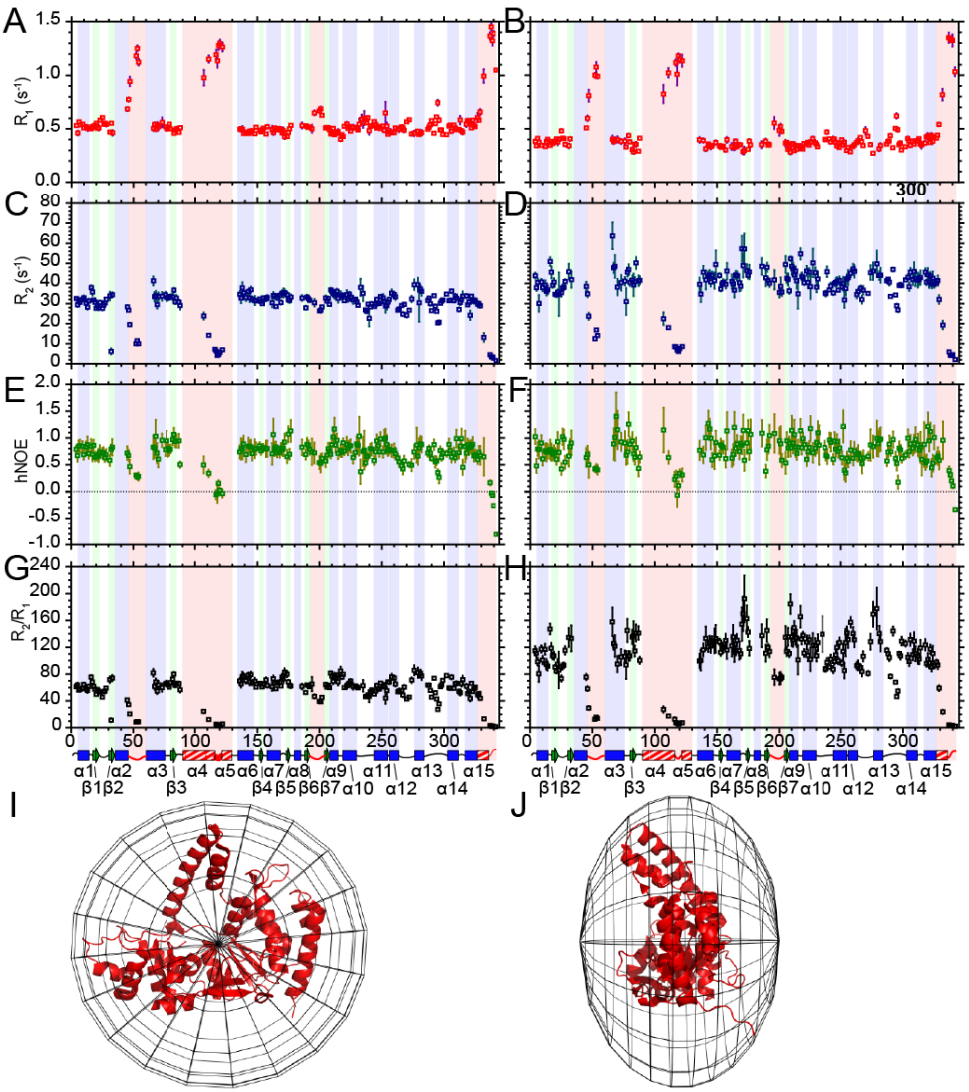

Figure S6. Experimentally-determined backbone ^{15}N relaxation parameters for hFEN1. (**A,B**) ^{15}N spin-lattice (R_1) rates, (**C,D**) spin-spin (R_2) rates and (**E,F**) ^{15}N-{^{1}H} NOE (hNOE) values at (**A,C,E**) 600 MHz and (**B,D,F**) 800 MHz for ^{2}H,^{15}N-labelled hFEN1 measured using interleaved TROSY-readout pulse sequences (16). R_2/R_1 plots generated from the data above for (**G**) 600 MHz and (**H**) 800 MHz, showing the per-residue average molecular correlation time in model-free methodology. Secondary structure maps are provided as described in Figure S2A. The oblate spheroid diffusion tensor best describes the tumbling of hFEN1 in solution. View of the diffusion tensor with the hFEN1 protein structure (3Q8K) (1) along the (**I**) z-axis and (**J**) x,y plane to illustrate agreement between the structure and the selected diffusion tensor.

Figure S7 (Related to Figure 3)

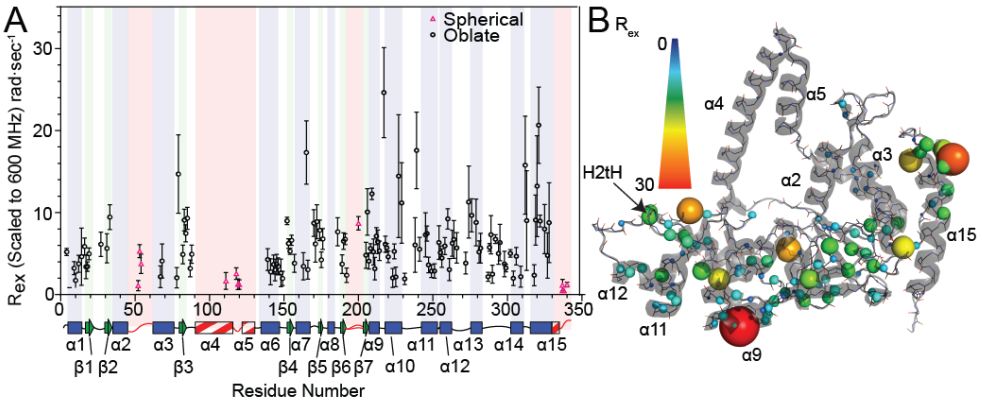

Figure S7. Model-free analysis of hFEN1 relaxation data identifies regions with millisecond timescale motions. (**A**) Chemical exchange terms (R_{ex}) were derived from *relax* (17) using backbone ^{15}N relaxation data acquired at 600 and 800 MHz (Figure S6A-F) and plotted versus residue number. Black circles represent data fitted to the oblate spheroid diffusion tensor (Figure 6I and J), whereas pink triangles were fitted to a spherical diffusion tensor. Locations of the residues with respect to protein secondary structure are illustrated below the panel as described in Figure S2A. (**B**) R_{ex} values plotted on a cartoon depiction of the hFEN1 protein structure (3Q8K) (1). The spheres represent the nitrogen nuclei for which data were derived. The R_{ex} spectrum bars illustrate the magnitude of R_{ex} values with respect to sphere color and size.

Figure S8 (Related to Figures 5 and 6)

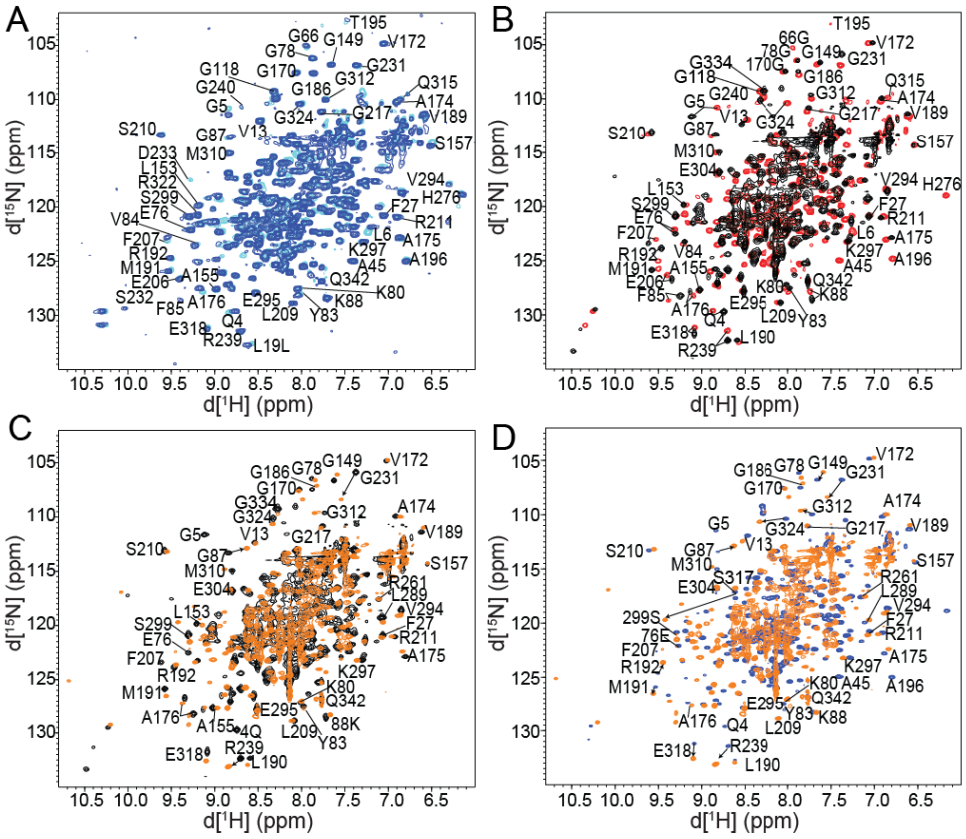

Figure S8. Superposed ^{1}H-^{15}N TROSY spectra illustrating chemical shift changes that arise from the addition of divalent cations to hFEN1, hFEN1$_{K93A}$ and the hFEN1$_{K93A}$-DNA complex. (**A**) ^{1}H-^{15}N TROSY spectra of hFEN1 in the presence of 8 mM Mg^{2+} (cyan) and 8 mM Ca^{2+} (blue) are nearly identical. Minor differences are seen for residues close to the active site (e.g., Q4, G5, L6, S157 and A176) and may arise due to the larger size and looser coordination geometry of Ca^{2+}. These small changes indicate that the Mg^{2+} and Ca^{2+} divalent metal ions are coordinated in the same active site locations. (**B**) ^{1}H-^{15}N TROSY spectra of hFEN1$_{K93A}$ (red) and the hFEN1$_{K93A}$-DNA complex (black) in the presence of 0.1 mM EDTA. (**C**) ^{1}H-^{15}N TROSY spectra of the hFEN1$_{K93A}$-DNA complex in the presence of 0.1 mM EDTA (black) or 8 mM Ca^{2+} (orange). (**D**) ^{1}H-^{15}N TROSY spectra of hFEN1$_{K93A}$ (blue) and the hFEN1$_{K93A}$-DNA complex (orange) in the presence of 8 mM Ca^{2+}. Peaks with good chemical shift dispersion are labelled.

Figure S9 (Related to Figures 6 and 7)

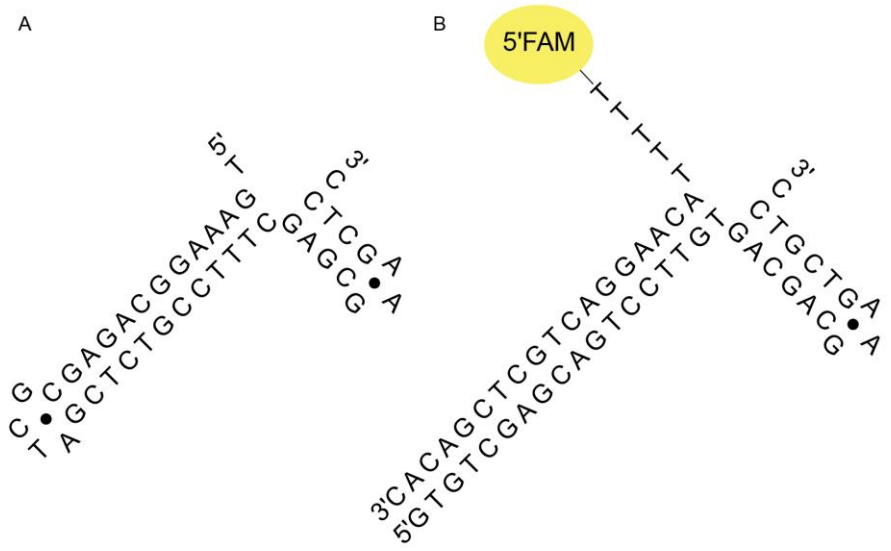

Figure S9. Sequence and secondary structure of the DNA substrate used herein. (**A**) The hFEN1$_{K93A}$-DNA complex uses a unimolecular substrate design bearing stable tri- (dGAA) (18) and tetra- (dGCTA) (19) loops to prevent hFEN1 inadvertently binding blunt-ended duplex. (**B**) The bimolecular substrate was used to assess the initial rates of reaction used to determine second order rate constants. In addition, this same substrate was used in smFRET experiments except it lacked the 5'-FAM label.

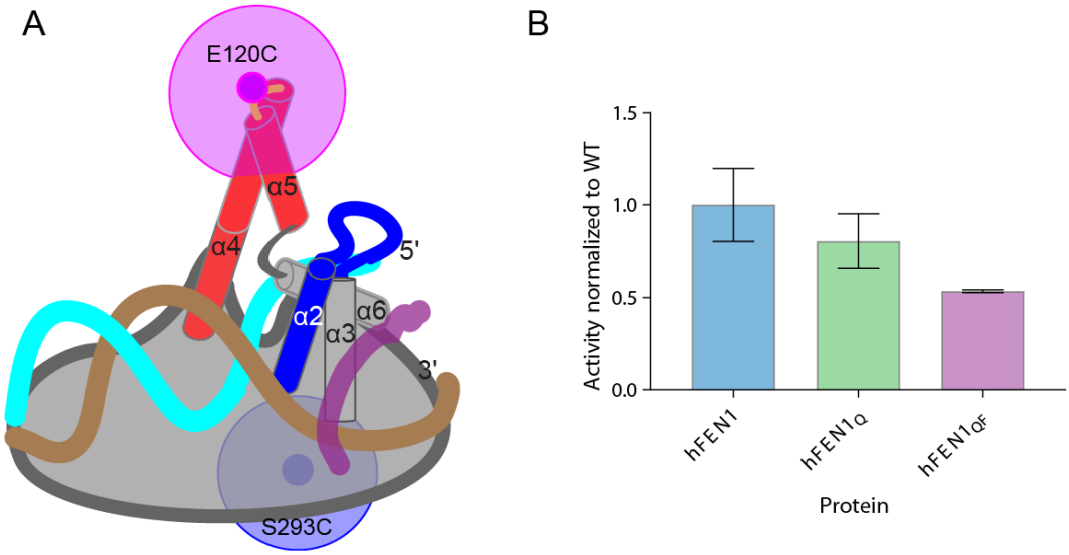

Figure S10. Relative labelling positions, activity and mass spectrometry results for hFEN1$_{Q}$ and hFEN1$_{QF}$ proteins. (**A**) Relative labelling positions of the fluorophores on hFEN1$_{QF}$ with substrate. (**B**) Activity of hFEN1 proteins assessed by normalized initial rate measurements at 50 nM substrate relative to WT protein. Values represent the average of three measurements (n=3), and error bars represent the standard error of the mean (SEM).

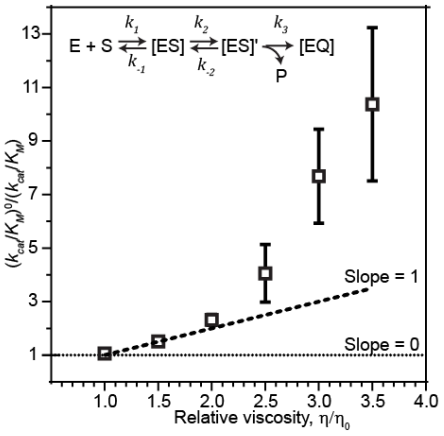

Figure S11. The effect of viscogen on relative second order rate constants is larger than expected and non-linear. Graph of mean relative second order rate constants $(k_{cat}/K_M)_0/(k_{cat}/K_M)$ for hFEN1-catalyzed hydrolysis of DF5,1 (Figure S9B) versus relative viscosity (SEM, n=3) shows a non-linear dependence. Inset: simplified scheme of the hFEN1-catalyzed reaction illustrating bimolecular complex [ES] formation, a conformational change to the catalytically-competent state [ES]' and substrate hydrolysis into ssDNA (P) and enzyme-dsDNA [EQ] products. Microrate constants (k) up to and including the first irreversible step (k_3) contribute to k_{cat}/K_M. hFEN1 second-order reaction rates (k_{cat}/K_M) on double-flap DNA substrates approached diffusion-control (10^7–10^9 $M^{-1}\cdot s^{-1}$); therefore, hFEN1 has likely achieved catalytic perfection with its optimal substrate (Figure S9B) (20). Observation of a linear dependence for the reciprocal of relative k_{cat}/K_M with respect the relative viscosity of the solution has been the standard manner by which to confirm catalytic perfection or diffusion-control for enzymes (21,22). Assuming that only the rates of enzyme and substrate association and complex disassociation are affected by viscogen, a linear relationship is expected for the reciprocal of normalized k_{cat}/K_m with respect to relative viscosity. Moreover, a diffusion-controlled reaction is expected to afford a slope of one (dashed line), whereas an enzyme that is not diffusion controlled is expected to afford a line with a slope of zero (dotted line) (21,22). Although increasing the relative viscosity of the reaction buffer with either glycerol or sucrose did increase the reciprocal of normalized k_{cat}/K_m with the optimal double flap substrate, the effects were non-linear especially at relative viscosities greater than two. Similar non-linear viscogen dependence has been reported for an intrinsically disordered protein that folds upon binding its protein partner (23). In addition to slowing diffusional encounter, microviscogens have also been shown to slow protein conformational rearrangements (24) and protein-catalyzed reactions when conformational changes in the catalytic cycle were involved (25). Indeed, the impact of viscogen on rates of hFEN1-catalyzed reaction under single-turnover conditions measured by single molecule techniques is consistent with viscogen slowing conformational change steps after enzyme-substrate complex formation (26). Therefore, the non-linear viscogen dependence on relative second order rate constants could be due to the combined effect of viscogen on the rates of bimolecular association and subsequent conformational changes. However, we cannot rule out the possibility that sucrose and glycerol are acting either as inhibitors of the reaction by some unknown mechanism or that these polyols are affecting the energy landscape of the various conformational ensembles by acting as osmolytes.

SUPPLEMENTARY REFERENCES

...1. Tsutakawa, S.E., Classen, S., Chapados, B.R., Arvai, A.S., Finger, L.D., Guenther, G., Tomlinson, C.G., Thompson, P., Sarker, A.H., Shen, B.i. *et al.* (2011) Human flap endonuclease structures, DNA double-base flipping, and a unified understanding of the FEN1 superfamily. *Cell*, **145**, 198-211.

...2. Fetrow, J.S. (1995) Omega loops: nonregular secondary structures significant in protein function and stability. *FASEB J.*, **9**, 708-717.

...3. Tsutakawa, S.E., Thompson, M.J., Arvai, A.S., Neil, A.J., Shaw, S.J., Algasaier, S.I., Kim, J.C., Finger, L.D., Jardine, E., Gotham, V.J.B. *et al.* (2017) Phosphate steering by Flap Endonuclease 1 promotes 5′-flap specificity and incision to prevent genome instability. *Nat. Commun.*, **8**, 15855.

...4. Sakurai, S., Kitano, K., Yamaguchi, H., Hamada, K., Okada, K., Fukuda, K., Uchida, M., Ohtsuka, E., Morioka, H. and Hakoshima, T. (2005) Structural basis for recruitment of human flap endonuclease 1 to PCNA. *EMBO J.*, **24**, 683-693.

...5. Ceska, T.A., Sayers, J.R., Stier, G. and Suck, D. (1996) A helical arch allowing single-stranded DNA to thread through T5 5'-exonuclease. *Nature*, **382**, 90-93.

...6. Feng, M., Patel, D., Dervan, J.J., Ceska, T., Suck, D., Haq, I. and Sayers, J.R. (2004) Roles of divalent metal ions in flap endonuclease-substrate interactions. *Nat. Struct. Mol. Biol.*, **11**, 450-456.

...7. Mase, T., Kubota, K., Miyazono, K., Kawarabayasi, Y. and Tanokura, M. (2009) Crystallization and preliminary X-ray analysis of flap endonuclease 1 (FEN1) from Desulfurococcus amylolyticus. *Acta Crystallogr Sect F Struct Biol Cryst Commun*, **65**, 923-925.

...8. Hosfield, D.J., Mol, C.D., Shen, B. and Tainer, J.A. (1998) Structure of the DNA repair and replication endonuclease and exonuclease FEN-1: coupling DNA and PCNA binding to FEN-1 activity. *Cell*, **95**, 135-146.

...9. Matsui, E., Musti, K.V., Abe, J., Yamasaki, K., Matsui, I. and Harata, K. (2002) Molecular structure and novel DNA binding sites located in loops of flap endonuclease-1 from Pyrococcus horikoshii. *J Biol Chem*, **277**, 37840-37847.

...10. AlMalki, F.A., Flemming, C.S., Zhang, J., Feng, M., Sedelnikova, S.E., Ceska, T., Rafferty, J.B., Sayers, J.R. and Artymiuk, P.J. (2016) Direct observation of DNA threading in flap endonuclease complexes. *Nat. Struct. Mol. Biol.*, **23**, 640-646.

...11. Hwang, K.Y., Baek, K., Kim, H.-Y. and Cho, Y. (1998) The crystal structure of flap endonuclease-1 from Methanococcus jannaschii. *Nat. Struct. Mol. Biol.*, **5**, 707-713.

...12. Shah, S., Dunten, P., Stiteler, A., Park, C.K. and Horton, N.C. (2015) Structure and specificity of FEN-1 from Methanopyrus kandleri. *Proteins*, **83**, 188-194.

...13. Exell, J.C., Thompson, M.J., Finger, L.D., Shaw, S.J., Debreczeni, J., Ward, T.A., McWhirter, C., Sioberg, C.L., Martinez Molina, D., Abbott, W.M. *et al.* (2016) Cellularly active N-hydroxyurea FEN1 inhibitors block substrate entry to the active site. *Nat. Chem. Biol.*, **12**, 815-821.

...14. Garforth, S.J., Ceska, T.A., Suck, D. and Sayers, J.R. (1999) Mutagenesis of conserved lysine residues in bacteriophage T5 5′-3′ exonuclease suggests separate mechanisms of endoand exonucleolytic cleavage *Proc. Natl. Acad. Sci. USA*, **96**, 38-43.

...15. Shen, Y. and Bax, A. (2013) Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J. Biomol. NMR*, **56**, 227-241.

...16. Lakomek, N.A., Ying, J. and Bax, A. (2012) Measurement of ^{15}N relaxation rates in perdeuterated proteins by TROSY-based methods. *J. Biomol. NMR*, **53**, 209-221.

...17. Bieri, M., d'Auvergne, E.J. and Gooley, P.R. (2011) relaxGUI: a new software for fast and simple NMR relaxation data analysis and calculation of ps-ns and mus motion of proteins. *J. Biomol. NMR*, **50**, 147-155.

...18. Yoshizawa, S., Kawai, G., Watanabe, K., Miura, K.-i. and Hirao, I. (1997) GNA Trinucleotide Loop Sequences Producing Extraordinarily Stable DNA Minihairpins. *Biochemistry*, **36**, 4761-4767.

...19. Nakano, M., Moody, E.M., Liang, J. and Bevilacqua, P.C. (2002) Selection for thermodynamically stable DNA tetraloops using temperature gradient gel electrophoresis reveals four motifs: d(cGNNAg), d(cGNABg),d(cCNNGg), and d(gCNNGc). *Biochemistry*, **41**, 14281-14292.

...20. Finger, L.D., Blanchard, M.S., Theimer, C.A., Sengerová, B., Singh, P., Chavez, V., Liu, F., Grasby, J.A. and Shen, B. (2009) The 3′-flap pocket of human flap endonuclease 1 is critical for substrate binding and catalysis. *J. Biol. Chem.*, **284**, 22184-22194.

...21. Brouwer, A.C. and Kirsch, J.F. (1982) Investigation of diffusion-limited rates of chymotrypsin reactions by viscosity variation. *Biochemistry*, **21**, 1302-1307.

...22. Sengerova, B., Tomlinson, C., Atack, J.M., Williams, R., Sayers, J.R., Williams, N.H. and Grasby, J.A. (2010) Bronsted analysis and rate-limiting steps for the T5 flap endonuclease catalyzed hydrolysis of exonucleolytic substrates. *Biochemistry*, **49**, 8085-8093.
...23. Rogers, J.M., Steward, A. and Clarke, J. (2013) Folding and binding of an intrinsically disordered protein: fast, but not 'diffusion-limited'. *J. Am. Chem. Soc.*, **135**, 1415-1422.
...24. Sekhar, A., Latham, M.P., Vallurupalli, P. and Kay, L.E. (2014) Viscosity-dependent kinetics of protein conformational exchange: microviscosity effects and the need for a small viscogen. *J Phys Chem B*, **118**, 4546-4551.
...25. Rauscher, A., Derenyi, I., Graf, L. and Malnasi-Csizmadia, A. (2013) Internal friction in enzyme reactions. *IUBMB Life*, **65**, 35-42.
...26. Rashid, F., Harris, P.D., Zaher, M.S., Sobhy, M.A., Joudeh, L.I., Yan, C., Piwonski, H., Tsutakawa, S.E., Ivanov, I., Tainer, J.A. *et al.* (2017) Single-molecule FRET unveils induced-fit mechanism for substrate selectivity in flap endonuclease 1. *eLife*, **6**, e21884.