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

Mechanistic Studies of Flap Endonuclease-1 (FEN-1)

Supervisor: Prof. Jane. A. Grasby

Sana. I. Algasaier

Dec 2017

Declaration:

Except the information that have been taken from specific references to other sources, this work in this thesis is my original own work, and it has not previously been submitted, wholly or in part, for any other degree.

Sana Algasaier

Publication arising from work covered in this thesis:

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*, 8258-8268.

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.; Sarker, A. H.; Her, M. Z.; Rashid, F.; Hamdan, S. M.; Mirkin, S. M.; Grasby, J. A.; Tainer, J. A. Phosphate steering by Flap Endonuclease 1 promotes 5'-flap specificity and incision to prevent genome instability. *Nat. Commun.* **2017**, *8*, 15855.

Abstract:

Flap endonuclease-1 (FEN-1) is a divalent metal-ion dependent phosphodiesterase that has essential roles during DNA metabolism in all organisms. FEN-1 hydrolyses selective substrate structures that contain duplex-duplex junctions endonucleolytically at single-strand DNA (ssDNA) and double-strand DNA (dsDNA) junctions (fen activity) as well as exonucleolytically on substrates lacking 5'-ssflaps (exo activity). ssDNA flaps may also contain secondary structure and the reaction of these substrates is referred as gap endonuclease activity (gen activity). In all cases, FEN-1 hydrolyses optimal substrates that have a single nucleotide 3'-flap making a single incision one nucleotide into the substrate 5'-duplex. Association of optimal substrates with the enzyme and their correct positioning for reaction requires both global and local conformational changes of DNA substrates. Substrates are bent at the duplex-duplex junction and the terminus of the 5'-duplex must untwist into the active site.

Previously, FEN-1 active site mutations linked with cancers have been shown to affect FEN-1's activity in vitro. However, reports of the effects of these mutations differ from one study to another. It was suggested that E160D and A159V mouse FEN-1 (mFEN-1) mutations cause lung cancer in mice by decreasing exo and gen activities of FEN-1 but not fen activity, but this was contradicted by a later study. To address this discrepancy the properties of mFEN-1 and its mutations were studied. In *chapter 3* it is shown that FEN-1 has the ability to recognise different substrates using same active site and chemistry with a slight preference for the double flap substrates. The ability of the enzyme to bind substrates in a bent conformation was assayed using FRET, with dyes located on each duplex arm. This demonstrated that active site metal ions and presumably the processes that require them are not required to bind substrate in a bent conformation. The divalent metal ion dependent local untwisting of the DNA necessary to place the substrate within the active site was interrogated using exciton coupled CD of tandem 2-amino purines located at the junction terminus of the 5'-duplex. Although the absence of the 5'-flap led to modest reductions in the rate of reaction, the enzyme's ability to bend and twist the substrate were not dramatically affected unless a 5'-phosphate monoester was absent from the exo substrate. The active site mutations E160D and A159V decreased the catalytic efficiency of the enzyme on all substrates to a comparable extent. Furthermore, investigations of individual sterps of the reaction where undertaken and suggested that these rate decreases may be partly explained by hindering the production of ideal local DNA conformational changes and substrate bending.

Studying the structure-function relationships of human FEN-1 (hFEN-1) particularly focused on the ability of the enzyme to bring about global and local substrate structural changes necessary for catalysis *(chapter 4)*. In agreement with the mFEN-1 results, the presence of the 5'-phosphate monoester of the exo substrate was required for perfect conformational changes and substrate bending. In addition, the absence of the 3'-flap destabilised enzyme-substrate complexes. In contrast, mutations of the enzyme generally altered the capacity to bring about local conformational changes required for catalysis rather than altering substrate bending. The active site mutations D181A and the helical arch mutations Y40A, K93A, R100A, helix breaking L130P, and K93AR100A all prevented an ideal local DNA conformational change, leading to a deficit in catalysis. However, mutation of helical arch residue Lys 93 did not alter bending or active site positioning implying a solely catalytic role for this residue. In contrast, repulsive helical arch mutations R104EK132E and Quad E did alter substrate binding abilities, as did the 3'-flap pocket mutation L53A. Taken together, the results imply that while the requirements for bent DNA-protein complex formation are relatively modest (a 3'-flap and its binding site and a 5'-phosphate) most protein mutations are tolerated. However, moving DNA to the active site to allow reaction requires divalent metal ions, the intact structure of helical cap and the presence of a substantive amount of conserved hFEN-1 amino acid residues highlighting this to be a key determinant of hFEN-1 reaction specificity.

Abbreviations:

2-AP	2-Amino Purine (Adenosine analogue).
Ala/A	Alanine.
aa	Amino acid.
APS	Ammonium persulphate.
AI	Auto induction media.
Asp/D	Aspartic acid.
Arg/R	Arginine.
BER	Base excision repair.
BSA	Bovine serine albumin.
bp	Base pair.
Cm	Chloramphenicol.
CD	Circular dichorism.
DF	Double flap.
dHPLC	Denaturing high-pressure liquid chromatography.
DNA	Deoxyribonucleic acid.
DSB	DNA Double strand break.
DNU	Double nucleotide Unpairing.
ds	Double strand.
DTT	Dithiothreitol.
DNTP	Deoxynucleoside triphosphate.
[E]	Enzyme.
(E)	Energy transfer efficiency.
EP	Enzyme-product.
ES	Enzyme-substrate.
ECCD	Exciton coupling circular dichroism.
EDTA	Ethylenediaminetetraacetic acid.
EXO-1	Exonuclease-1.
E_{min}	Minimum energy transfer efficiency.
E_{max}	Maximum energy transfer efficiency.
FEN-1	Flap-dependent 5'-endonuclease.
FB	Folding buffer.
FRET	Fluorescence resonance energy transfer.
GEN-1	Gap-dependent 5'-endonuclease.
Glu/E	Glutamic acid.
h	Human.
HJ	Holliday Junction.
HEPES	Hydroxyethylpiperazineethane sulphonic acid.
IMAC	Immobilised metal ion affinity chromatography.
Kan	Kanamycin.
K_{bend}	Dissociation constant.
K_M	Michaelis constant.

k_{STmax}	Single turnover rate constant.
k_{cat}	Catalytic rate.
Lys/K	Lysine.
L/Leu	Leucine.
lpBER	Long patch base excision repair.
LB	Luria borth media.
m	Mouse.
MT	Multiple turnover.
MeCN	Acetonitrile.
MMR	Mismatch repair.
nt (s)	Nucleotide (s).
OD	Optical Density.
P, Q	Product.
Pro/P	Proline.
PCNA	Proliferating cell nuclear antigen.
RNA	Ribonucleic acid.
RB	Reaction buffer.
RRB	Reduced reaction buffer.
RPA	Replication protein A.
r_{min}	Minimum distance between donor and acceptor.
r_{max}	Maximum distance between donor and acceptor.
S	Substrate.
SA	Streptavidin.
spBER	Short-patch base excision repair.
SB	Storage buffer.
ss	Single strand.
ST	Single turnover.
SF	Single flap.
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis.
TAMRA	Tetramethyl rhodamine.
Tyr/Y	Tyrosine.
TBE	Trisborate EDTA.
TCE	Trichloroethanol.
TEMED	Tetramethylethylene diamine.
Tbab	Tetrabutyl ammonium bromide.
Val/V	Valine.
WT	Wild type.
XPG	Xeroderma pigmentosum complementation group G.

Contents:

Chapter 1: Introduction to FEN-1 Enzyme:

The enzyme Flap EndoNuclease-1 (FEN-1), which exists in all kingdoms of life, is essential in mammals. FEN-1 is a structure-specific 5'-nuclease enzyme with a phosphodiesterase activity capable of catalysing a specific incision in a variety of DNA structures. FEN-1 performs critical roles during DNA replication and repair removing single-stranded 5'-flaps from bifurcated nucleic acid structures using its divalent metal ion active site.[1] In presence of water and at physiological pH 7 the enzyme hydrolyses a specific phosphodiester bond between two adjacent nucleotides to produce a double-strand DNA product with a 5'-phosphate end and a single-strand flap product with a 3'-OH end *(figure 1.1)*.[2-5] To understand the enzyme's ability to recognise and cleave a certain single phosphodiester bond in a specific DNA substrate, a detailed review of DNA and FEN-1 structures and properties will be undertaken.
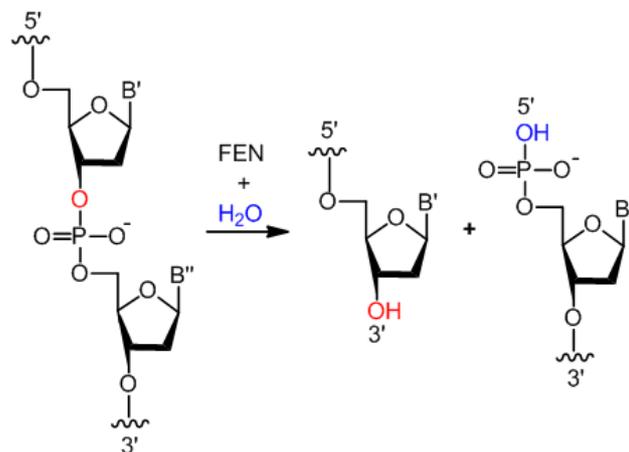
Figure 1.1. Hydrolysis of phosphodiester bond by FEN-1 producing two products with 5'-phosphate monoester end and a 3'-OH end.

1.1. FEN-1 substrate (DNA):

DNA is the genetic information carrier in all of cellular life. Bacterial DNA is carried as a single circular molecule that is supported with different packaging proteins. However, the DNA is packaged differently in eukaryotes, as a set of chromosomes associated with packaging proteins in the cell's nucleus. The chromosomes also contain many other proteins that are required for DNA metabolism such as gene expression, DNA replication and DNA repair. Despite being a tightly folded complex, DNA can interact easily with many enzymes in the cell to be for example: replicated, repaired, or expressed. Amazingly, DNA is the only molecule capable of directing its own synthesis. DNA was first defined as a linear sequence of nucleotides that carries the genetic information. However, it is the three dimensional structure of DNA that allows it to function as

the genetic material. Further studies proposed and then determined the three dimensional atomic structure of DNA. These later models produced further understanding of DNA's biological function.[6]

As shown in *figure 1.2*, DNA is a double helix containing two long polynucleotide strands that are coiled around the same axis, but running in opposite directions (antiparallel, one from 5' to 3' and the other 3' to 5'). Each individual repeating nucleotide unit comprises of a 2-deoxyribose sugar, a phosphate group, and a nitrogen base (purine or pyrimidine). The nucleotides are connected by a 3'-5' phosphodiester linkage between sugar molecules to form one strand of the DNA. The phosphodiester bonds can be considered as the bridges that hold the chain together. Thus the backbone of the DNA strand is formed when a 3'-OH of sugar group of one nucleotide forms a covalent bond with a 5'-OPO_3^{2-} of sugar group of another adjacent nucleotide. In this condensation reaction, one molecule of water is lost.[6-8]
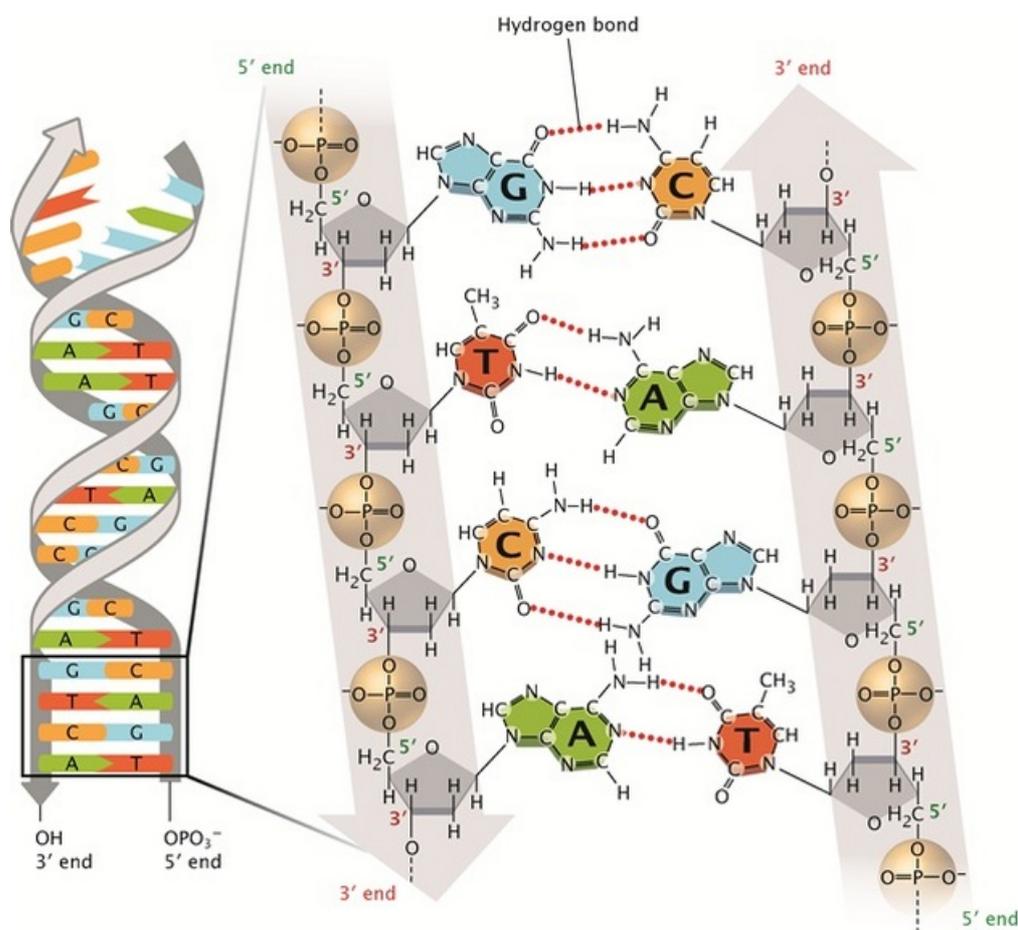
Figure 1.2. *Part of middle of a DNA duplex model to show the general DNA chemical structure contains the pentose sugar (grey rings), bases (identified by the first letters of their names) and the phosphodiester bonds (highlighted by gold balls).[8]*

Alongside the phosphodiester bonds, the DNA structure cannot be completed without the internal hydrogen bonds, which hold the two strands together. These hydrogen bonds occur between purines and pyrimidines. As see in *figure 1.3* the adenine (A) and guanine (G) are purines that have two rings, while the cytosine (C) and thymine (T) are pyrimidines and contain one ring.
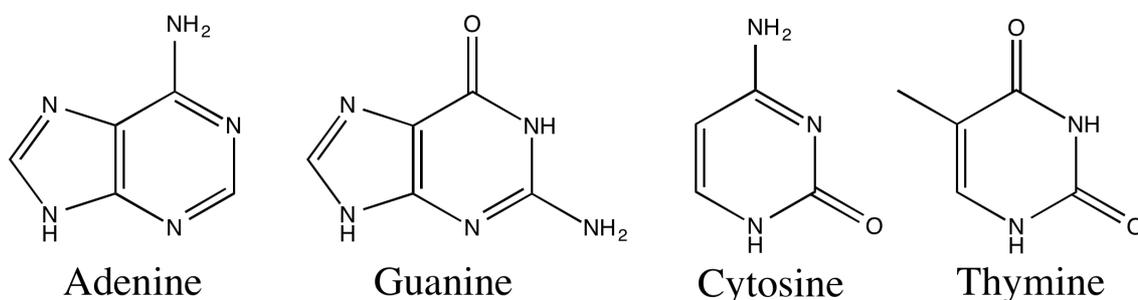
Figure 1.3. *The nitrogenous bases of DNA. Adenine (A) and guanine (G) are double ring purine bases, while the cytosine (C) and thymine (T) are single ring pyrimidine bases.*

Accordingly the base pairing will always follow the rule A pairs only with T, and G pairs only with C. As the hydrogen bond occurs between a partially positively charged hydrogen atom and a partially negatively charged acceptor atom (oxygen or nitrogen), the A-T base pair forms 2 hydrogen bonds, whereas the G-C base pair produces 3 hydrogen bonds. The strength of these bonds depends on the orientations of the participating atoms, however they are still quite weak bonds, which is a very important property during the different DNA functions. In addition to these hydrogen bonds the twisted DNA structure is also supported by another force resulting from stacking of the bases on top of each other.[7] This stacking increases the DNA's stability by enhancing the interaction and reducing the spaces between the base pairs. The hydrophilicity of the phosphate groups places these on the outside of the molecule.

RNA (RiboNucleic Acid) is polynucleotide sequence that is related to the DNA. RNA copies of the DNA are made during DNA transcription process that ultimately leads to translation and protein synthesis. RNA copies and transfers the information from the DNA to direct the synthesis of proteins, transferring the information into another chemical form. These differences in function are facilitated by some variances between DNA and RNA structure *(figure 1.4)*. RNA contains the ribose sugar rather than 2-deoxyribose sugar, which means that the 2nd carbon of the sugar in RNA has an attached OH group, while in DNA it has an attached H. Unlike DNA, RNA is a single-stranded molecule can fold upon itself by base pairing the molecule to form a three-dimminsional RNA structure with lower chemostability than DNA. RNA is also slightly different in base pairing,

since RNA uses uracil instead of thymine. Both bases have similar general structure but the uracil lacks a methyl group on its ring.[6,7]
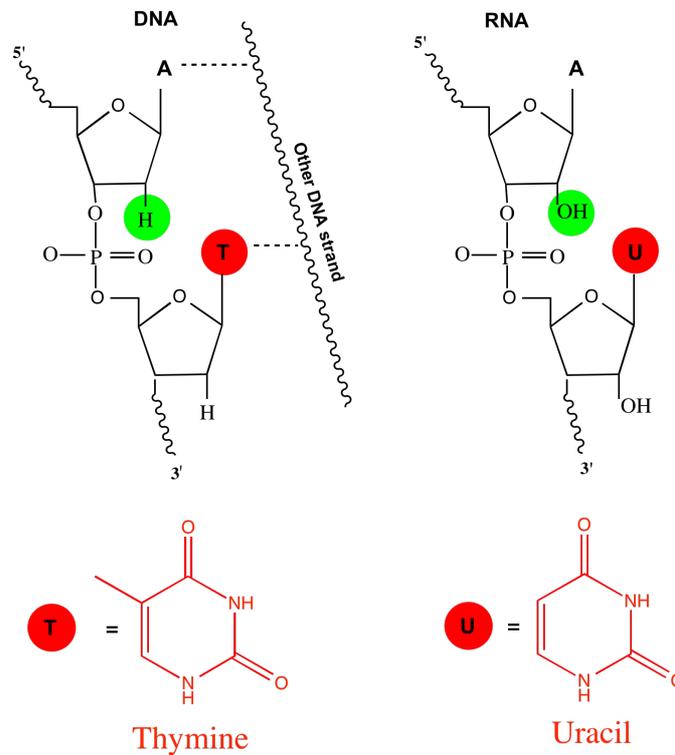
Figure 1.4. *The differences between the DNA and RNA molecules.*

As a result of these small differences, DNA and RNA have different stability reflecting their functions. DNA, used for storage and transmission of genetic information, is required to be more stable than the RNA, which is used to transfer genetic code for protein synthesis. In DNA, the 2-deoxyribose sugar gives more stability to the DNA molecule. This is because in the ribose sugar, the OH group on the 2nd carbon can attack the phosphodiester bond leading to phosphodiester bond cleavage and reduced stability. In addition using thymine in DNA increases its stability by reducing mutations. As known in DNA, cytosine is converted to uracil around 100 times per cell per day, but the cell can repair it by substituting the uracil with cytosine again. However, using uracil instead of thymine in DNA structure can produce a major problem. In this case, the cell cannot recognise if the uracil belongs there or not and will stop repair.[6]

Although the double helix is the most common image of the DNA, this molecule also has the ability to produce a number of other important structures during the different DNA processes. In eukaryotes, DNA is processed in two ways to maintain and transfer the information. In these processes, genetic information is copied into more DNA during DNA replication or ultimately translated to proteins: These processes are based on enzymes and protein complexes that have

major roles in organising and successfully transferring the information. Accordingly with these interactions of DNA, a number of complex structures using one or both strands of DNA will be formed *(figure 1.5)*.[4,9]
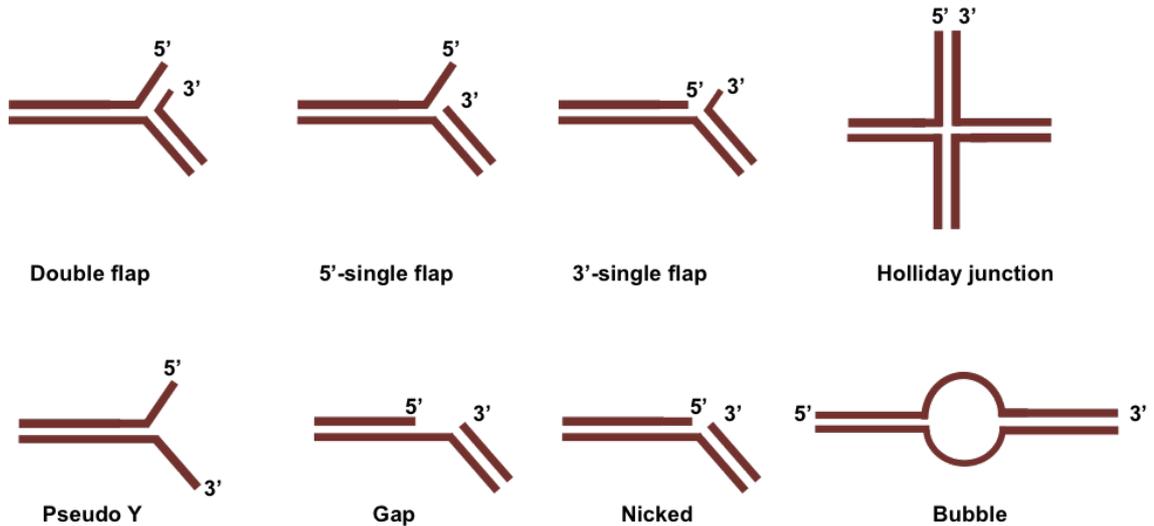
Figure 1.5. Alternative DNA structures that are the result of different DNA processes.

1.2. Synthesis and hydrolysis of the DNA backbone:

Phosphodiester-bonds are key structural features of the DNA backbone. Formation and breaking of these bonds underlie all life process such as DNA replication, repair and recombination. Enzyme-catalysed mechanisms, whether synthesis or hydrolysis, usually make or break phosphodiester bond using metal ions within their active sites.[10,11]

On one hand, formation of phosphodiester bonds are required to synthesise DNA strands by adding nucleotides. The bond results from a nucleotide transfer reaction from a dNTP molecule to the growing DNA chain and is catalysed by a polymerase. This polymerisation reaction mechanism requires the two metal ions of the polymerase active site and was deduced to be S_N2P forming a pentacovalent phosphate transition state *(figure 1.6)*. This substrate alignment allows deprotonation of the 3'-OH of the DNA strand by the adjacent water molecule. The new bond forms when the deprotonated 3'-OH of the DNA strand attacks the α-phosphate of the dNTP molecule breaking the phosphoanhydride bond between the α and β-phosphates of the dNTP molecule. As a result, a pyrophosphate (PPi) leaves and a new nucleotide is added to the DNA strand.[10]
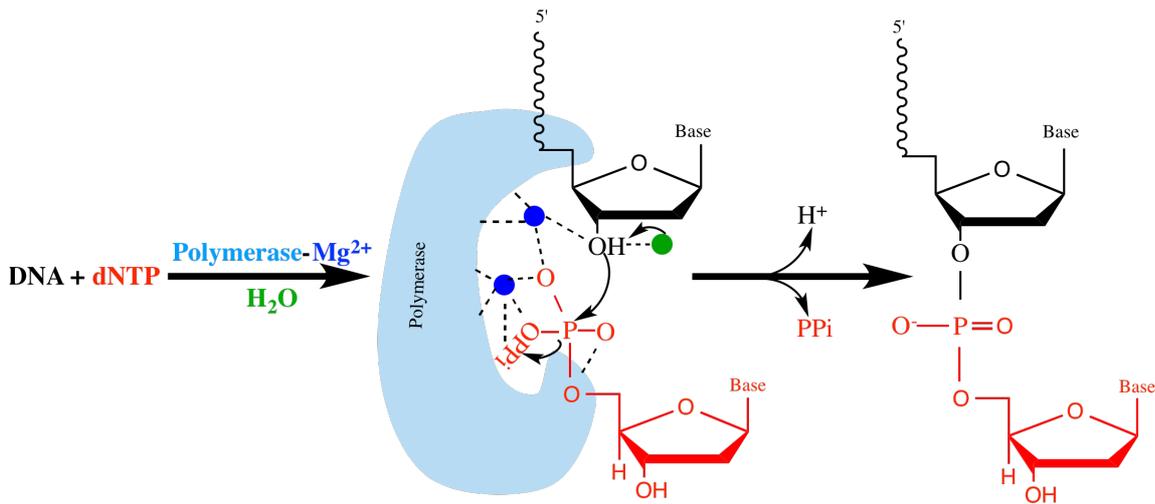
Figure 1.6. *Phosphodiester-bond synthesis to extend DNA structure with one nucleotide from a dNTP molecule using the two metal ions active site of the polymerase in presence of water.*

Per contra, phosphodiester bond hydrolysis facilitates most of the biological processes of the DNA such as synthesis, manipulation and repair. The great stability of these bonds requires phosphodiesterases to catalyse their hydrolysis reactions. Nucleases are phosphodiesterases that typically work with a high degree of specificity in function and reaction sites of hydrolysis. As these bonds form the sugar-phosphate-sugar backbone of the DNA, each bond between two nucleotides has two bridges of P-O *(figure 1.7)*. Accordingly, nucleases can cleave one of the two bridges, 5' or 3', in the DNA strand to produce a 3'-OH and 5'-phosphate products *(Figure 1.7.a)* or a 3'-phosphate and 5'-OH products *(figure 1.7.b)* respectively. The attacking nucleophile is always opposite of the leaving group in a penta-covalent intermediate or transition state to form the 3' or 5' OH-product. Most of the nucleases prefer to attack the 5' side (figure 1.7.a) likely because the 3'-OH product is a nucleophile that can be passed onto other reactions such as DNA synthesis, whereas the 5'-phosphate product is an ideal complementary substrate for DNA metabolism such as replication, recombination, or repair.[12,13]
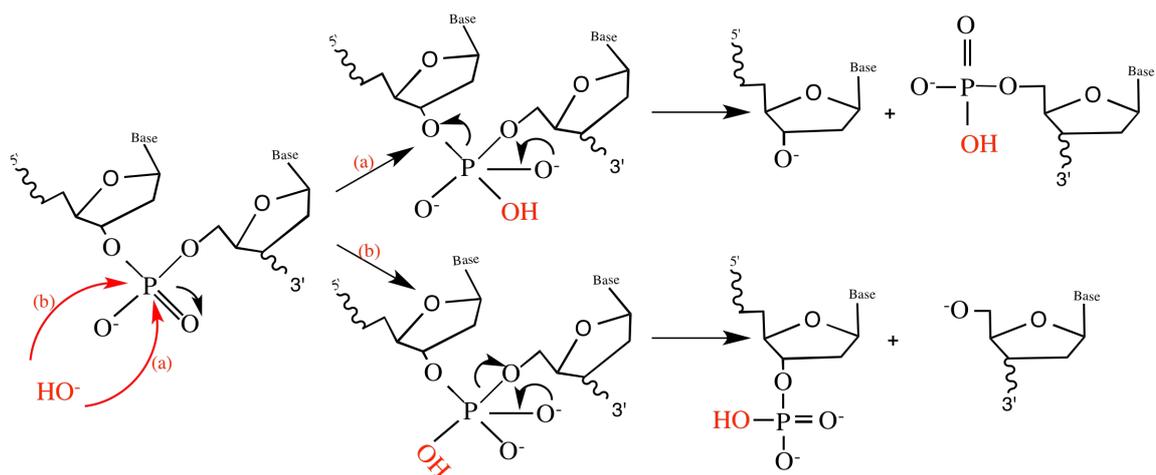
Figure 1.7. *The two different ways the phosphpodiester bond can be attacked by the HO⁻ nucleophile. In route (a) the nucleophile attacks the 5' side producing a 3'-OH and a 5'-phosphate products, while in (b) it attacks the 3' side producing a 3'-phosphate and a 5'-OH products. (The five co-ordinate transition sate or intermediate is shown).*

In general, both the synthesis and the hydrolysis of DNA backbone can be considered as a nucleophilic attack, with RO⁻ replacing PPi during DNA synthesis and HO⁻ replacing RO⁻ in DNA hydrolysis. There are three possible mechanisms that have been suggested for this nucleophilic substitution. In the first mechanistic model, the leaving group departs after formation of a highly negatively charged pentacovalent phosphate intermediate, which results from the nucleophilic addition. This is usually referred to as an addition-elimination mechanism or associative mechanism with intermediate (*figure 1.8.a*).[14,15] In the second possibility of the mechanism, the reaction follows a dissociative mechanism. The leaving group is eliminated first; forming a metaphosphate intermediate, then the nucleophile rapidly attacks *(figure 1.8.b).*[16,17] Between these two extreme mechanisms, a third associative mechanism with transition state lies between both. In this model, the leaving group departure and nucleophilic attack happen simultaneously during the transition state *(figure 1.8.c)*. Accordingly, the concerted mechanism is extremely complicated since the transition state can be considered as a deletion, an addition or a combination of both. In effect, the transition state as a removing step is similar to the dissociative mechanism, while it is close to the simple addition-elimination mechanism if it is deemed as an addition step leading to a concurrent transition state between leaving group departure and addition. To date, there is no strong evidence to suggest exactly what happens between starting substrate and final product in enzymatic reactions. This is because this kind of reactions depends on many factors such as nature of the nucleophile, the electrophile and the leaving group. Furthermore, the active site architecture of the enzyme that catalyses the reaction has significant effect on the mechanism because of the contribution of its metal ions.[14,15]

a) Associative mechanism with intermediate (or addition-elimination mechanism)
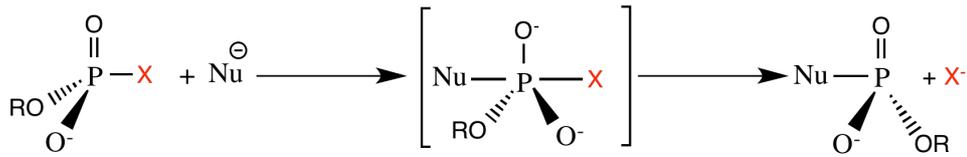
b) Dissociative mechanism
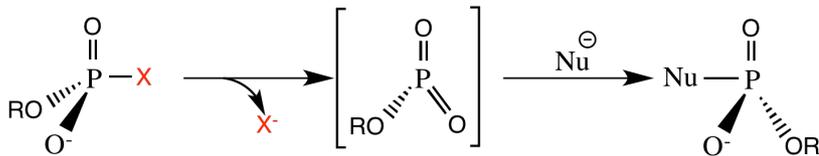
c) Associative mechanism with transition state
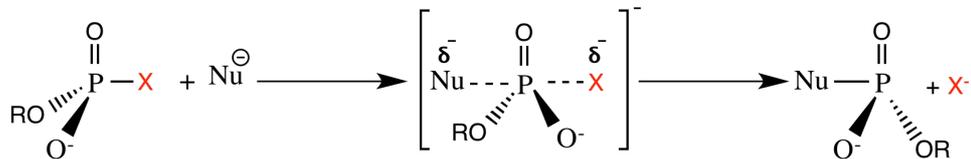
Figure 1.8. The suggested mechanism of the phoshpodiester-bond synthesis and hydrolysis when nucleophile (Nu) attacks and leaving group eliminates (X). Nu is RO⁻ and X is OPPi⁻ in synthesis, but Nu is HO⁻ and X is RO⁻ in hydrolysis. a) Associative mechanism with intermediate (or addition-elimination mechanism) via a pentacoordinate intermediate. b) Dissociative mechanism via tri-coordinate metaphosphate intermediate. c) Associative mechanism with transition state.

Phosphodiesterases are typically characterised by their binding of two metal ions such as magnesium ions (Mg^{2+}) to catalyse hydrolysis of the target phosphate diester bond in combined actions known as the two-metal ion mechanism. In such way, one metal ion binds a non-bridging phosphate oxygen of the scissile phosphate diester bond acting as an electrophilic catalyst and the nucleophilic hydroxide ion, and the other metal ion binds the same non-bridging phosphate oxygen and oxygen of the leaving group *(figure 1.9)*. One very important feature of these biological enzymatic reactions is that the electrophilic power of the phosphorus atom is increased by the Lewis acid (electron-accepting) effect of one or more Mg^{2+}.
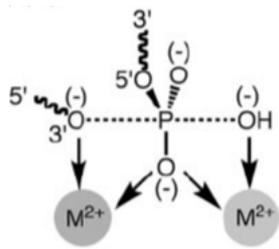
Figure 1.9. The two-metal ion mechanism. Interaction of enzymatic ions with a non-bridging phosphate oxygen and nucleophilic hydroxide ion with one ion and leaving group with the other ion to increase phosphorus electrophilicity.

1.3. FEN-1 in metabolism:

Flap endonuclease-1 (FEN-1) is an important factor to maintain genetic stability in higher organisms. Previous studies have indicated an increase in cancer progression related to defects of FEN-1 function.[18,19] FEN-1 has essential roles during DNA replication and long-patch base excision repair (lpBER). In both processes, FEN-1 contributes to removal of 5'-single stranded DNAs (5'-ssDNA), known as 5'-flaps, independently of the sequence using its phosphodiesterase activity.[5] The proceeding sections describe the specific role of FEN-1 in each of DNA replication and lpBER:

1.3.1. FEN-1 in DNA Replication:

Before a cell divides, DNA has to replicate so that each daughter cell will capture a copy of the genome from the original DNA.[20] Because cells pack the DNA into a condensed structure named chromatin under the normal situations, they need to unfold, unwind and separate the two strands from each other to start replication at specific locations in the genome called replication origins. In this case, each of the two strands works as a template for the creation of the new strand. This structure is ordinarily termed a replication fork that is organised by an enormous number of proteins. Each enzyme performs a core function for a specific purpose such as to unwind the DNA strands, organise the ssDNA, prime, copy the DNA, load co-proteins (such as PCNA), or process the Okazaki fragments.[21,22]

DNA replication starts at a specific point in the DNA duplex known as the origin of replication. This specific point is a destination for many enzymes that will bind specific binding sites on the DNA to proceed with the replication. Initially, the DNA helicase unzips the DNA-double helix by breaking the hydrogen bonds between the two polynucleotide chains to generate a replication fork *(figure 1.10)*. Each strand will act as a parental strand to template the synthesis of a new strand. Because of the antiparallel nature of the DNA structure, one daughter strand will be synthesised continuously in the same direction of polymerase function (leading strand), on the other hand, the other daughter strand will be synthesised discontinuously (lagging strand) as segments named Okazaki fragments. Therefore, this manner of discontinuous synthesis will produce further problems.[23,24]

In both cases, the polymerisation is initiated by a short complementary RNA oligonucleotide (RNA primer), which is synthesised by primase enzyme. The RNA primer is then extended with DNA primer synthesised by DNA polymerase α (Pol α), in order to form an RNA-DNA primer (approximately 20-25 bp long). From the mature RNA-DNA primer, proofreading-proficient pol ε

and pol δ will carry out synthesis the daughter DNA strands on the leading and lagging strands, respectively. Pol ε moves continuously along the leading strand to synthesize a new daughter strand in the 5' to 3' direction. On the other arm Pol δ produces the new daughter strand as long series of fragments known as Okazaki fragments.[1]

To avoid turning the DNA to RNA during continuation of the replication process, the previous segments of the RNA primer should be displaced during polymerisation of the lagging strand. This displacement will cause a flap in the nucleic acid polymer approximately 50 million times in an average human cell replication. These flaps should be eliminated to allow DNA ligase to connect the newly synthesised DNA parts together to form a continuous polymer as shown in the figure below.
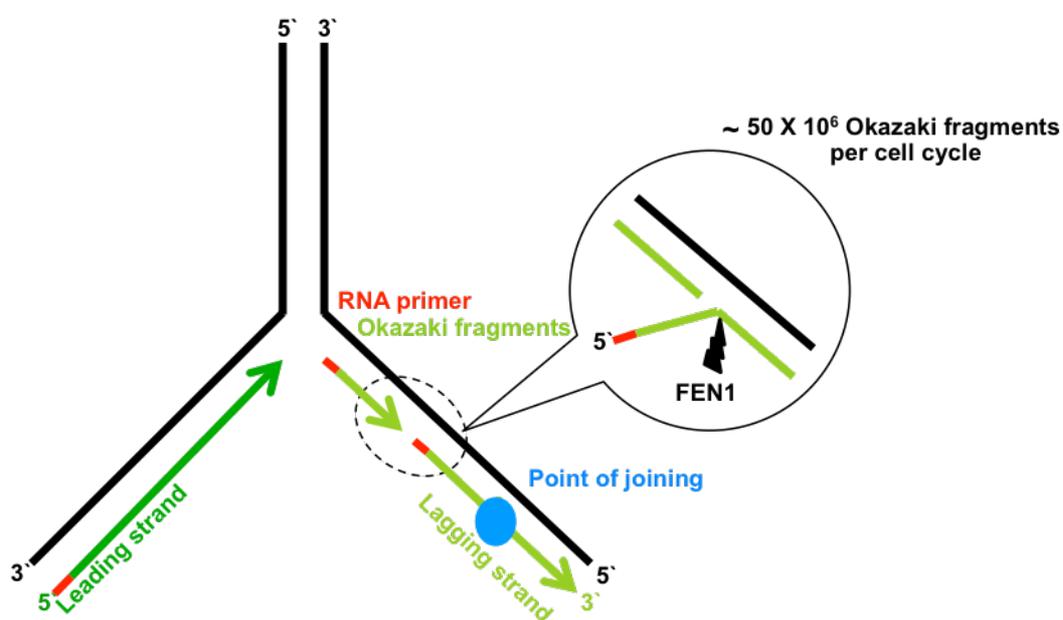
Figure 1.10. *Role of FEN-1 in lagging-strand DNA replication. As the lagging strand is synthesised discontinuously, the* ***RNA primer*** *and a part of the* ***Okazaki fragment*** *should be removed by FEN-1 before* ***joining****.*

In essence, completion of this process requires joint enzymatic work with a complicated mechanism. Although indicating most of the proteins that constitute the central players in this process solved the mystery, details are still a point of contention between many studies. For example, a previous study suggested that the RNA primer is removed gradually by the combined action of Pol δ and PCNA as a 5'-flap with 1 or 2 nucleotides length, which is removed directly by FEN1. After degradation of the RNA primer, the resulting nicked DNA will be connected by ligase to generate the completed DNA duplex.[22] Another explanation was proposed to explain this mechanism in a slightly different way. When Pol δ arrives to the 5' end of the previous Okazaki frag-

ment, it will continue the synthesis by replace the mature primer and a part of the new DNA as a 30-35 nucleotides flap length. The combined action of RPA and Dna2 protein cleaves the flap to leave only 5-7 nucleotides. Then FEN1 cleaves the shortened flap to form a nicked DNA duplex that can be sealed by ligase to generate the completed new DNA daughter.[25] Another proposed model offers a role for the ribonuclease (RNase) H2 protein that cuts the RNA primer leaving the last RNA nucleotide, which is removed as a 5'-flap by FEN1. At the same time the gap is filled by Pol δ and then ligated by ligase to complete the new resulted DNA duplex.[1]

In some cases, the replication mechanism happens imprecisely forming some accidental genomic lesions. Most of these changes will activate DNA repair mechanisms to maintain the genetic stability of the DNA. As a result, the high efficiency of DNA repair will eliminate mutations accumulation in the DNA sequence.

1.3.2. FEN-1 in Base Excision Repair (BER):

DNA is the genetic information store in living cells; consequently, any DNA damage or instability could be extremely deleterious. For example, the link between the DNA damage and disease is that of mutations, which can lead to cancer. Because of this, DNA repair processes are necessary for the survival of all organisms. Due to different types of environmental insult, the damage response occurs by a number of diverse pathways with participation of many enzymes. The major DNA repairing mechanism are base excision repair (BER), nucleotide excision repair (NER), mismatch repair (MMR). Each proceeds by the action of subsequent enzymes to generate various intermediates by removing the damage and incision of the phosphodiester bond to fix it. FEN-1 is one of the enzymes that has critical role in lpBER *(figure 1.11)*.

The first step of the BER pathway is initiated by recognition of then removing the damaged nucleobase via a damage-specific DNA glycosylase. DNA glycosylases remove the damaged base by flipping it into the active site and with hydrolysis of the glycosidic bond to generate an abasic site (AP-site) as a product. Some of these enzymes are multiple damaged bases removers, however, others are specific for only one or two damaged bases. Interestingly, some glycosylases are bifunctional enzymes. In addition to the glycosidic bond cleavage ability, they have less efficient AP lyase activity. Consequently, the AP site can be recognised by DNA glycosylase again or other enzyme.[22,25]

The BER pathway is resumed by incision the deoxyribose phosphate bond of the resulted AP-site. Cleavage of this AP-site by glycosylase with its AP-lyase activity or endonuclease-1 (APE-1) leaves a nick with 3'-unsaturated aldehyde to the AP-site and 5'-phosphate end or 3'-OH and 5'-

deoxyribose phosphate (dRP) end to the AP site respectively.[25] The 3' AP-site intermediate always follows the short patch BER (spBER) pathway to insert a single nucleotide, however the 5' AP-site intermediate can follow either spBER or lpBER to incorporate multiple nucleotides.
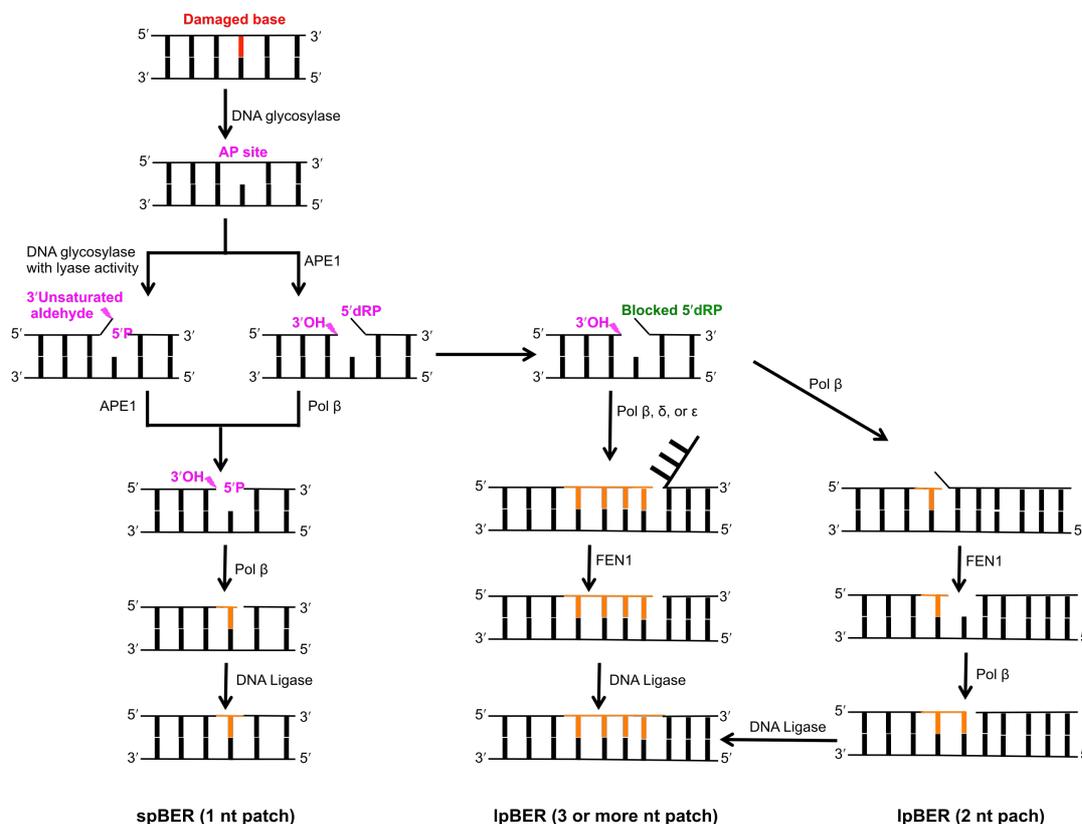
Figure 1.11. *The two ways of the DNA base excision repair (BER) the long-patch (lp) and the short-patch (sp) to remove **damaged base** in a part of DNA.[26]*

The spBER is mediated by APE1 or polymerase β (pol β) to process the resulted 3' AP-site intermediate or 5' AP-site intermediate respectively. In both cases this leads to creation of a gap with 3'-OH and 5'-phosphate intermediate. Then pol β fills the gap by incorporation of a new single nucleotide to the 3'-OH termini. Finally DNA ligase seals the nick in the backbone by creation a new phosphodiester bond between the 3'-OH of the new nucleotide and the 5'-phosphate of the next nucleotide.

By contrast, lpBER is an alternative pathway of spBER when pol β cannot hydrolyse the 5'dRP group. In this case lpBER will be organised by polymerase β, δ, or ε to incorporate multiple nucleotides to fill the gap. In the meantime, incorporation of the new nucleotides displaces a single strand of DNA to generate a flap. Then the flap is removed by FEN-1 forming nicked DNA. The process ends when the fresh DNA strand is joined by DNA ligase.[27,28]

1.4. FEN-1 Superfamily (5'-Nucleases):

FEN-1 superfamily members are structure specific nucleic acid hydrolysing enzymes that have similar structural features and an absolute need for two central metal ions to play critical roles during DNA replication, repair and recombination. In eukaryotic organisms the superfamily has been divided to flap endonuclease-1 (FEN-1); exonuclease-1 (EXO-1); gap endonuclease-1 (GEN-1); and xeroderma pigmentosum complementation group G (XPG).[2] These enzymes are characterised with a high specificity for particular substrate structures and tight regulation of function *(Table 1.1) (figure 1.12)*. This specificity and regulation keep the information and function of nucleic acid avoiding the danger of inappropriate phosphodiester-bond hydrolysis. In vitro, the superfamily enzymes hydrolyse various structures of DNA specifically to catalyse reactions of 5'-bifurcated or blunt ended DNA substrates endonucleotlytically and exonucleolytically respectively. The endonuclease activity can cleave oligonucleotides by attacking the middle of the substrate while the exonuclease activity has ability to remove one nucleotide at a time by attacking the end of the substrate.[2-5,12] Each superfamily member has a characteristic structure to organise substrate specifically and function. Although structural comparison between the superfamily members shows a considerable sequence and structural similarity, also important structural differences are identified, which cause variances in many of the superfamily properties that differentiate the superfamily members from each other. Comparison of these similarities and differences is very important to explain the substrate recognition and processing for the superfamily members.[2]

Superfamily fold	Conserved motifs	Superfamily member	Nuclease activity	Biological pathway	Preferred substrate
SAM/PIN Domain-like	Rossman (α/β) structure	hFEN-1	5'-endo/exo-nuclease	DNA replication lb-BER	df-DNA[2,5]
		EXO-1	5'-exonuclease	MM DSB repair	Nick gaps blunt ended DNAs[29]
		GEN-1	5'-endonuclease	DSB repair	HJ[30]
		XPG	5'-endonuclease	NER	DNA bubble[4]

Table 1.1. *FEN-1 superfamily members with corresponding structure and function.*
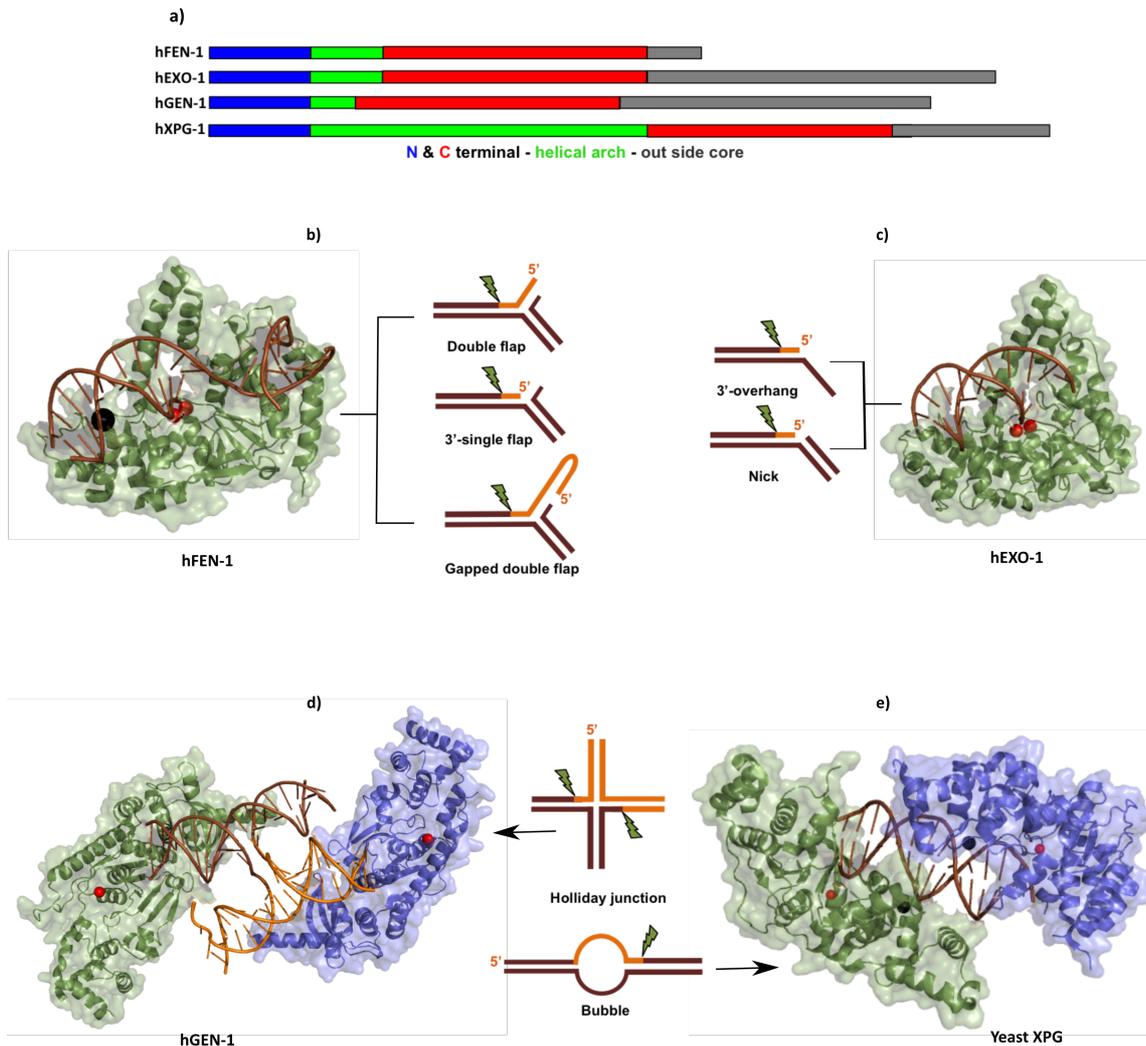
Figure 1.12. *FEN-1 superfamily members and their substrates. **a)** Domain alignment of FEN-1 superfamily members as determined by primary sequence analysis. **b), c), d),** and **e)** The crystal structures of **hFEN-1** (PDB:3Q8K), **hEXO-1** (PDB:3QEB), **hGEN-1** with **other monomer** (PDB:5T9J) and **yeast XPG** with **other monomer** (PDB:4Q0W) in complexes with **DNA**, respectively, to show how they recognise their substrates using their active site **metals** and **K^+ ion**. Also scheme of the preferred substrates of eachmember included and showing the **cleavage site**, **released product** and the **main product**. The substrates shown are: FEN-1 substrates (double flap, 3'-single flap, and gapped double flap), EXO-1 substrates (3'- overhang and nick), GEN-1 substrate (Holliday junction), and XPG substrates (bubble).[2,30,31]*

Flap endonuclease-1 (FEN-1), which typifies the superfamily, is present in all domains of life as a structure specific metallonuclease. It possesses phosphodiesterase activity to increases the rate of phosphodiester bond hydrolysis at least 10^{17} fold. FEN-1 is essential in DNA replication and lp-BER because of its ability to recognize the 5'-flap of DF substrate that arises during the processes *(figure 1.12.b)*. Participation of FEN-1 in DNA replication and repair provides important roles to maintain genomic stability. Differently from other superfamily members, FEN-1 also has 5'-exonuclease (exo) and gap endonuclease (gen) activities because of its recognition of 3'-single

flap (3'-SF) and gap substrates respectively. This multi-functionality allows it to participate in several DNA processes to recognise a wide variety of DNA structures and cleave the target phosphodiester-bond specifically *(figure 1.12.b)*. When FEN-1 acts in vivo it is bound to PCNA (Proliferating cell nuclear antigen). Interaction with PCNA and some other proteins is mediated by a short C-terminal extension.[1,3,5]

Exonuclease-1 (EXO-1) is another member of the superfamily that is very close homologue of FEN-1 that is detected in eukaryotic cells. EXO-1 is a much larger than FEN-1 although it has smaller nuclease domain. The smaller nuclease domain is situated at the amino-terminus. The structural differences between EXO-1 enzyme and the other superfamily members differentiate its functional specialisation. This 5'-structure specific nuclease has critical roles during mismatch repair (MM) and double-strand break repair (DSB). In these processes, EXO-1 can recognise the 5'-termini of different substrates such as nicked, gapped, or blunt DNA and hydrolyse them exonucleolytically *(figure 1.12.c)*. In some cases, the EXO-1 active site can recognise single-stranded 5'-flaps, which undergo endonucleolytic cleavage. Deficiency of EXO-1 activity can cause deep deleterious effects on human health leading to cancer. [3,29,32]

GEN-1 is a member of the superfamily that plays important roles recognising various substrates during DNA replication and recombination *(figure 1.11.e)*. GEN-1 can cleave 5' flap substrates in vitro, but the in vivo substrate is proposed to be Holliday junctions (HJ). The structural features of GEN-1 reveal its specificity for a four-way (Holliday) DNA junction *(figure 1.12.d)*. To resolve these HJs, GEN-1 acts in a dimeric form to generate the bilateral cleavage. Successful removal of HJs averts appearance of chromosome aberrations.[3,30,33]

XPG also belongs to FEN-1 superfamily and is responsible for one of the steps of the nucleotide excision repair (NER) process.[34,35] The human XPG (hXPG) sequence contains of 1186 amino acids and it works with other proteins as a multi-complex.[2] The proper recruitment of the enzyme requires interacting with XPB, XPD, RPA and PCNA.[31,34] In vitro, XPG cleaves various DNA structures such as single stranded 5'-overhangs and splayed arms DNA structures. However, the unique feature of XPG in vivo is the recognition of the DNA repair bubbles *(figure 1.12.e)*. The substrate specificity of the XPG protein could be related to the wide catalytic core that allow egression of the DNA without the free 5'-end.[3,31]

Overall, FEN-1 superfamily members have ability to recognise the 5'-duplex ends, sometimes when they are embedded in more complex structures and to hydrolyse the target phosphodiester bond by cleaving the P-O bridge. Despite the large structural similarity between the members, they provide various cellular activities. While these activities are independent of the

chemical nature of the substrate, they are considerably stimulated by the substrate structure.[36] Although individual superfamily members manage cleavage of different DNA structures, all of these nucleases cleave at one nucleotide into a 5'-double-stranded region of their target substrates.[37]

1.5. FEN-1 protein fold:

Generally, all of the superfamily members are a group of metalloenzymes that have a similar common protein structure with some modifications that occurred in the higher life organisms.[2] Primary sequence analysis of FEN-1 superfamily structure has characterised a number of conserved domains, which are N (N-terminal) and C (C-terminal) regions separated by a variable length of intermediate domain (I). When folded, these domains form a complex structure containing a variety of different structural features.[38] The early reported crystal structures of two members of the superfamily, which are FENs and EXO-1, from bacteriophage through archaeal to human have been highlighted the common FEN-1 superfamily architecture in *figure 1.13 a-d*.[2,33] Despite the differences in details, both members reveal monomeric proteins with binding sites that are created by conserved amino acids to contain at least two metal ions. A part of the nuclease domain feature orders as central parallel mixed and twisted β-sheet surrounded by secondary structures elements consisting almost entirely of α-helices. Although these α-helices sometimes have different positions in proteins from different organisms, they are mostly presented in similar location. The created central hall feature, which is known as the enzyme active site, provides highly conserved Asp and Glu residues to coordinate the requisite divalent metal ions. Moreover, it provides two sites on either sides of the β-sheet to bind the target dsDNA.[1,2]

The very recent crystal structures of GEN-1 and XPG confirmed what was proposed years ago from protein sequence information and showed high structural homology to FEN-1 in the active site *(figure 1.13 e.f)*.[31,33] The active site of FEN-1 and EXO-1 is capped and limited by helical arch, which is shaped by two α-helices from the intermediate domain. This spanned loop provides specific roles to organise specific DNA structures, and it has been divided further to helical gateway and helical cap regions for better understanding to their roles. In spite of the conserved similarity across the superfamily members in the gateway region (the base of the helical arch) that contains some of the active site residues, a huge variability in the cap region was detected. Whereas FENs and EXO-1 have shown variation in the helical cap length, neither XPG nor GEN-1 have helical cap, which makes their active sites wider and therefore they can recognize bigger sizes of specific DNA structures including continuous DNAs like four-way-junctions and bubbles. The superfamily members differ greatly in the C-terminal regions, where the sequence

homology is poor. In addition, the domain alignment of the superfamily members has shown variation in lengths outside the nuclease core (the C-terminal extensions' length) from huge in EXO-1, XPG, and GEN-1 to being smaller in FENs. Although it might be dispensable in the super-family members' activities, it is very important to their interactions with other proteins such as PCNA.[1,2,38-41]

Elucidation of FEN-1 structure and its DNA-binding elements revealed the answer to many ques-tions concerned with the mechanism of substrate hydrolysis and the ability to cleave various DNA structures within a single active site.[2] To present, thirteen crystal structures of FENs are obtained from a wide range of organisms including bacteriophages,[42-44] archaea,[36,45] thermo-philic bacteria,[46] yeast,[47] and human.[5,29] In hFEN-1, the N and C termini of the protein that are connected by I domain form the saddle like structure domain. The nuclease core domain archi-tecture consists of a seven stranded twisted β sheet supported with 15 α-helices from N and C domains creating the active site. This α/β system is connected by multiple loops ranging be-tween 12 and 15 Å to shape the interwoven fold of the protein. The interwoven loops allow res-idues throughout the sequence to form a carboxylate rich active site below a feature known as the helical arch.[5] Seven acidic residues (Asp and Glu residues) that are positionally conserved throughout evolution compose the hFEN-1 active site and surround two tightly bound catalytic metal ion cofactors.[2,5] This active site is gated by the bases of helices α4 and α2, and capped by the top of α4 and the α5 helices and the flexible loop between helices to complete forming of the active site hole. The resultant helical arch feature presents positively charged residues that line the inner surface that might contribute to accommodation of the single stranded DNA flap. Another binding feature is formed by the hydrophobic wedge, which consists of the top of α2 and α3 and the loop between helices. The α2 wedge bends the DNA at the junction point, whereas the α3 wedge and a part of the loop compose a hydrophobic pocket to organise the 3'-terminal nucleotide. This 3'-hydrophobic flap pocket consists of ten amino acids with mostly aliphatic residues. The H2TH region, which forms by α10 and α11, presents an additional bind-ing site to recognise the opposite DNA duplex. This domain binds a potassium ion that interacts with the downstream ds-DNA backbone to assist processing of the DNA.[5]
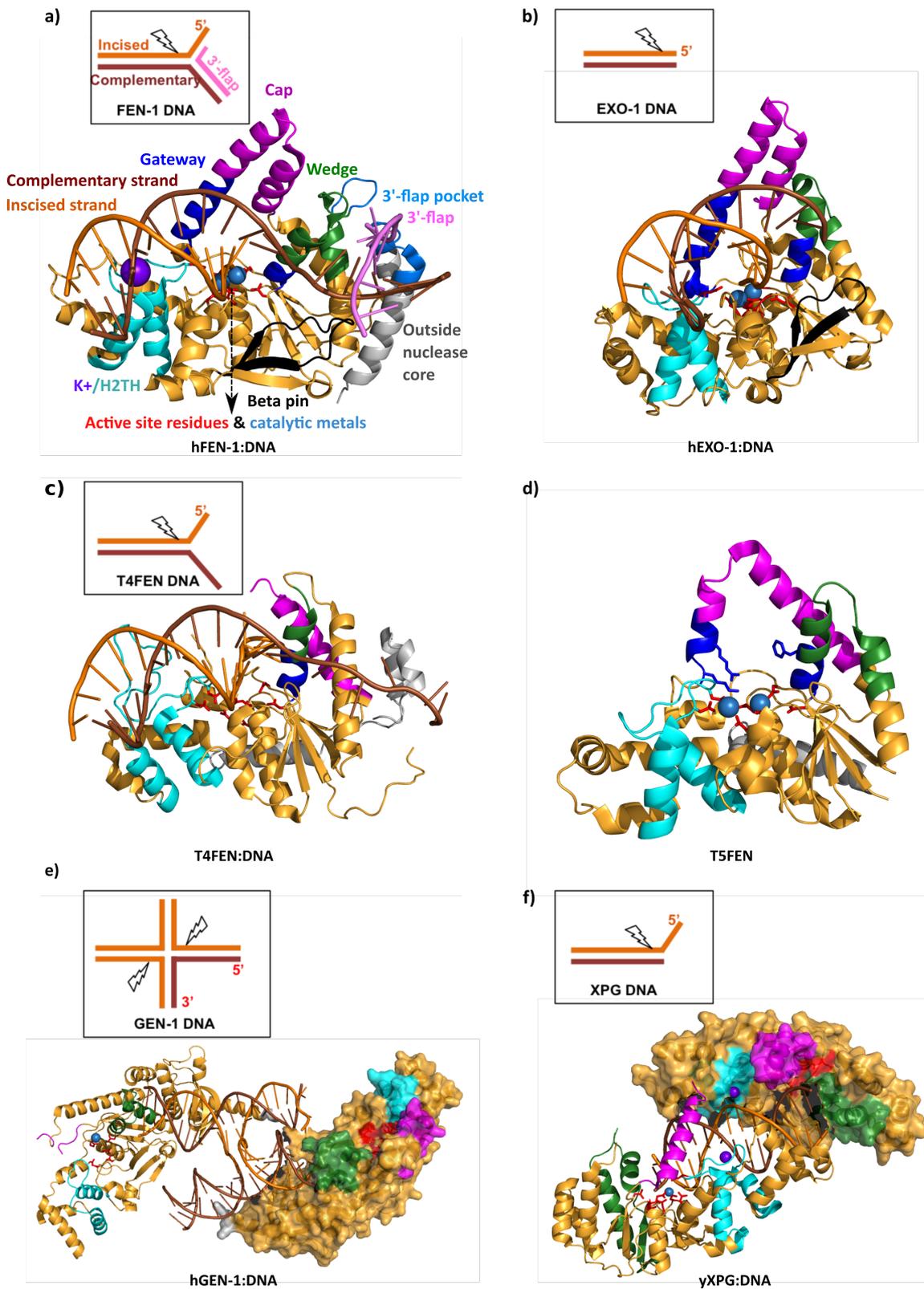
Figure 1.13. *The crystal structural features of FEN-1 superfamily that are created by folding of the key domains, which are indicated in a) using same colours, and showing their participation to recognise their preferred substrates in their catalytic cores by organising the **substrate duplex** to be cleaved in the target (arrows). The structures shown are: **a)** hFEN-1 (residues 2-346) with **DNA** and **active site metals** (PDB: 3Q8K), **b)** hEXO-1 (residues 2-346) with **DNA** and **ac-***

*tive site metals (PDB: 3QEB), **c)** T4FEN with **DNA** (PDB: 2INH) **d)** T5FEN with **active site metals** (PDB: 1UT5), **e)** hGEN-1 with **DNA** and **active site metals** (PDB:5T9J), and **f)** yest XPG with with **DNA** and **active site metals** (PDB:4Q0W). The crystal structure of each enzyme with DNA complex shows how the enzyme recognises the substrate using the main features: the helical arch (**cap** and **gateway**), the **wedge**, the **K^+/H2TH**, and the outside nuclease core to position the target bond into the **active site residues** between the two **metal ions**.*

1.6. Substrate specificity of hFEN-1:

FEN-1 is an essential enzyme that plays significant roles during DNA replication and DNA repair pathways.[1] This enzyme is characterised by its ability to recognise different DNA structures with diverse efficiency. Substrate structure is the factor that differentiates between the so-called various activities of FEN-1. Consequently, FEN-1 is described as a multifunctional and structure specific enzyme.[3,39] The FENs of higher organisms, which require discontinuous duplex DNA with free 5'-termini, recognise substrates with two dsDNA regions (downstream and upstream duplex regions) relative to the cleavage site. The enzyme uses its phosphodiesterase activity to carry out the cleavage that occurs at 1-nt into the downstream duplex *(figure 1.14)*. FEN-1 recognises DNA junctions to identify the structural polarity.[1]
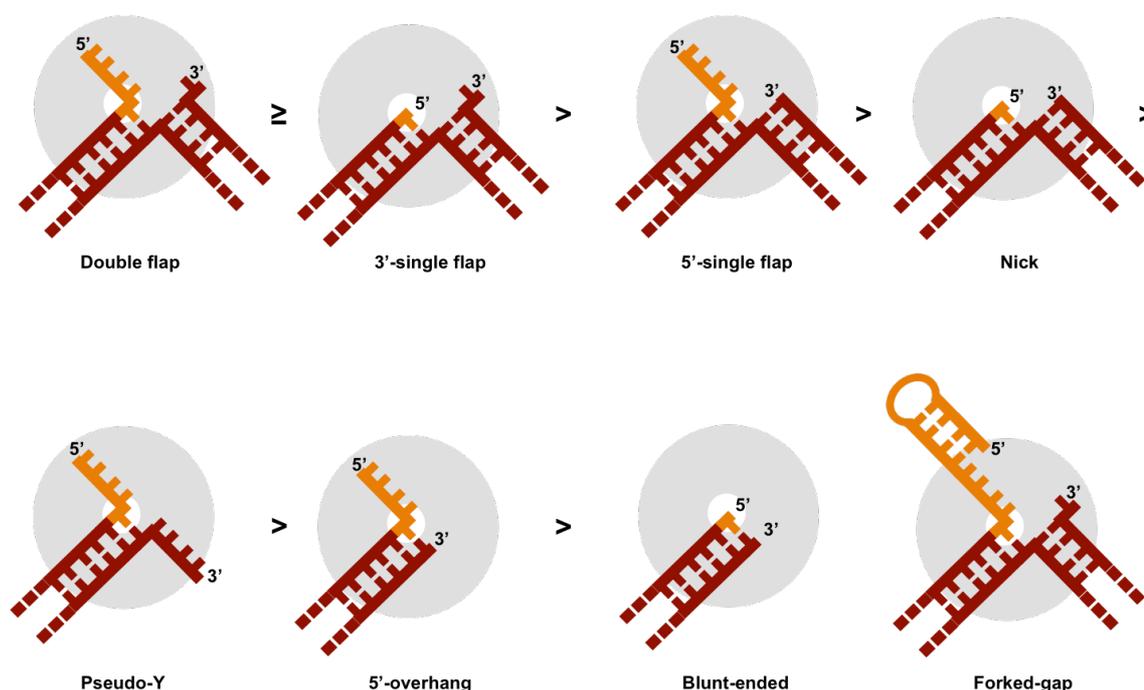
Figure 1.14. FEN-1 binds different DNA structures within its active site to achieve different activity with varying efficiency.[1]

Characterisation of FEN-1 specificity provides insight into its ability to perform endonuclease (fen), exonuclease (exo) and gap endonuclease (gen) activity using a single active site.[1,48] FEN-1 performing these three different activities can recognise different DNA structures that are pro-

duced during metabolic pathways. Flap endonuclease (fen) activity is the most dominate activity of FEN-1 that contributes in 5'-flap removal during DNA replication and in eukaryotes in lpBER from various DNA structures for example double flap, 5'-single-flap, pseudo-Y and 5'-overhang DNAs *(figure 1.14)*. Various factors can affect FEN-1 efficiency to cleave 5'-flaps. First, an increase in flap length of more than 20 nucleotides reduces FEN-1 efficiency in part because this can cause flap secondary structure.[46,49] Second, the modification of the 5'-phosphate group of the reacting 5'-duplex decreases the rate of FEN-1 reaction.[50] Third, a mere occurrence of big non nucleotide molecules attached to the flap can also inhibit FEN-1 activity.[51] The other FEN-1 activity is exo that is used to recognise several other DNA structures such as 3'-single-flap, nick, and blunt ended DNAs to remove 1-nt *(figure 1.14)*. The last characterised activity of FEN-1 is gen, which is responsible for cleavage of DNAs with secondary structures like forked-gap DNA *(figure 1.14)*. In vivo it is proposed that exo and gen FEN-1 activities play a role in apoptotic DNA fragmentation and rescue of stalled DNA replication. [1,48] FEN-1 is considered as a multifunctional enzyme, in vitro, its efficiency varies greatly to cleave multiple structures of DNA *(figure 1.14)*.[1] While the double flap and 3'-single-flap substrates are the ideal DNA structures for eukaryotic FENs, a pseudo Y DNA structure is the optimal substrate for the lower organisms such as T4-FEN and T5-FEN bacteriophages.[1,52,53] But the question is if FEN-1 does the same chemistry to hydrolyse the phosphodiester-bond, how does it still have different functions or does it just recognise different substrates?

1.7. The 3'-flap preference:

An early study observed that the upstream region was important for efficient cleavage by FEN-1s from higher organisms.[46]. Later, the same group used a designed double flap substrate with 19 nt 5'-flap and either 1 nt or 10 nt 3'-flap lengths to demonstrate the importance of the single nt 3'-flap to efficient binding and cleavage of the 5'-flap by FEN-1 in higher organisms. Although the DF substrate with 10 nt 3'-flap bound less efficiently than DF with 1 nt 3'-flap and 5'-SF substrates, both DF structures cleaved more efficiently than 5'-SF DNA *(figure 1.14)*. These results provide evidence for the importance of the 3'-flap and aid understanding of how FEN-1 recognises different DNA structures, although the exact role of 3'-flap plays to recognition the DF substrate by FEN-1 still not clear.[52] Predilection of FEN-1 to cleave the double-flap structure indicated that FEN-1 binds the 3'-flap specifically.[54] More biochemical results suggested that FEN-1 catalysis requires conformational changes in the flexible loop domain as the enzyme binds its substrate.[55,56] As a result of a scarcity of structural information about the conformational changes occurring with FEN-1 interactions, the modelled FEN1:DNA complex did not explain the FEN-1 specificity that was observed biochemically.[36,42,45,57]

Generally, FENs of the higher organisms cleave substrates with 3'-flaps endonucleolytically if they have a 5'-flap or exonucleolytically if they are without a 5'-flap with greater efficiency than any other substrate *(figure 1.14)*.[1] It is of great interest to understand why FEN-1 prefers to bind 3'-flap substrates. The first suggestion is that the 3'-flap increases the first order rate of FEN-1 reaction after initial formation of FEN-1:DNA complex.[1] Secondly, the importance of 3'-flap (1-nt) is that the single nucleotide of the flap is complementary to the template DNA and so it fills and closes the gap after releasing the dsDNA product *(figure 1.15.a)*. Thus, the ligatable nicked product will not require initiation of DNA repair mechanisms that are required in absence of the 3'-flap because of the production of gapped DNA *(figure 1.15.b)*.[1] Structural studies of FEN-1 showed that the higher organisms possess a 3'-extrahelical nucleotide binding pocket.[58] Observation of this a specific binding pocket of 3'-flap (1-nt) in the hFEN-1 structure underscores the great importance of this 3'-flap for FEN-1 substrate recognition. Placement of the single nucleotide 3'-flap in this pocket stabilises the FEN-1:DNA complex. Overall, absence of this 3'-flap decreases the catalytic efficiency of FEN-1 and varies its ability to interact with different substrates.[1,59]
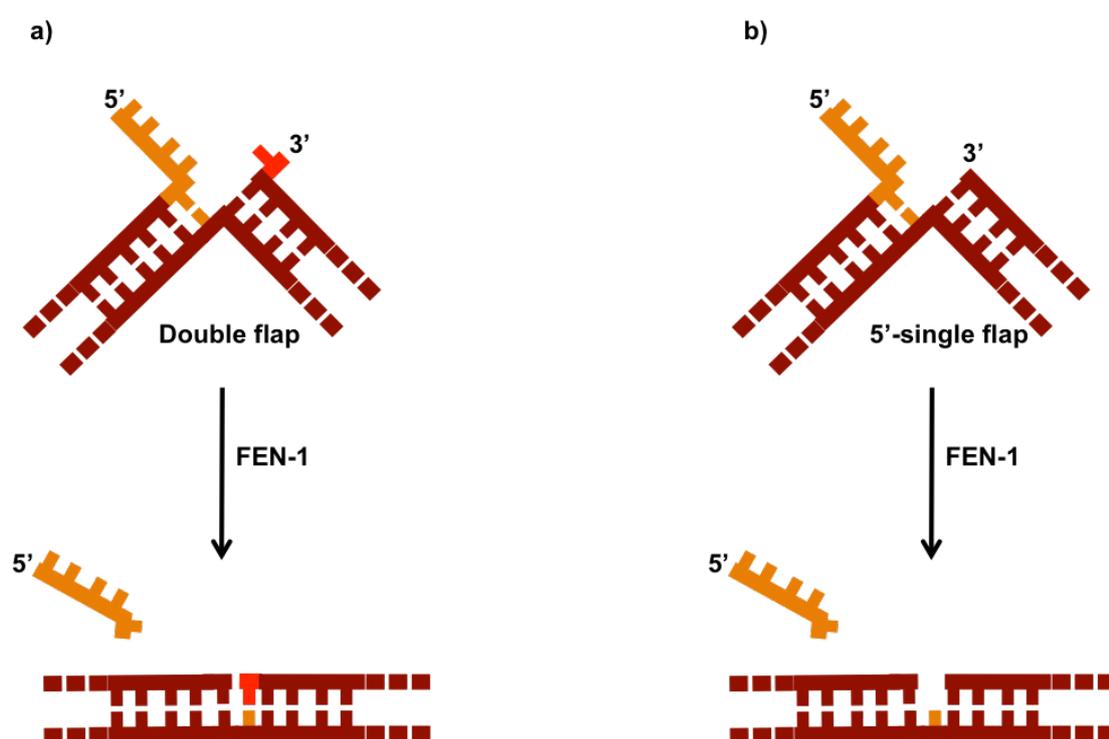
Figure 1.15. *Importance of* the 3'-flap in FEN-1 substrates. **a)** The double flap substrate produces a ligatable nick. **b)** The 5'-single flap produces a gapped DNA.[1]

1.8. hFEN-1 binding elements:

Based on biochemical evidence, FEN-1 proteins of higher organisms prefer substrates with two dsDNA regions (downstream and upstream duplex regions) with a junction between *(figure 1.16.a)*.[1] These structures contact hFEN-1 via several positively charged side chains of amino acids that extend across length of FEN-1 protein. The hFEN-1:DNA interaction is mediated by several binding elements that are the helix-2-turn-helix (H2TH) region with bound potassium ion, the hydrophobic wedge, the 3'-flap binding pocket, the helical arch that is divided to cap and gateway, and the active site that component of seven residues surround two metal ion *(figure 1.16.b.c)*.[5] Each binding element is discussed in detail below:
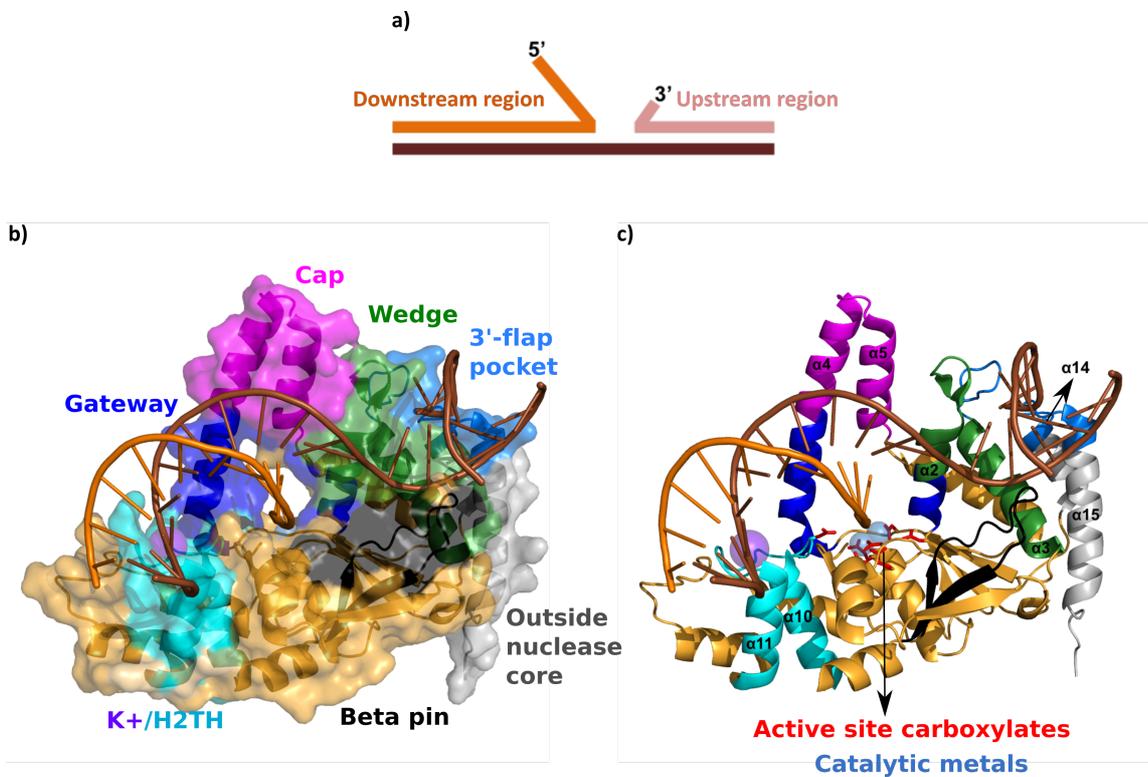
Figure 1.16*. Binding elements of hFEN-1 to interact with dfDN . **a)** DNA substrate with upstream and downstream regions such as dfDNA is the FEN-1 preferred substrate. **b)** hFEN-1 surface shows its key binding elements that are involved in its complex with product (PDB:3QK). **c)** Internal view of hFEN1 product complex shows the binding elements and illustrates the active site residues that directly coordinate the two metal ion*

1.8.1. The K$^+$/H2TH binding site:

The helix-two-turn helix (H2TH) is a classical DNA binding site has been found in many of replication and repair enzymes. In hFEN-1 this motif forms by α-helices 10 and 11 and binds K$^+$ ion, which is preferred because of their typical ion-ligand distances that allow intimate DNA contact.

The backbone carbonyl oxygens of IIe238 and IIe241 and the hydroxyl of Ser237 coordinate the K^+ ion. These residues exist in the turns between the two helices (10 and 11). The four basic residues Arg239, Lys244, Arg245 and Lys267 from the H2TH region protrude from the protein surface creating a furrow in combination with this motif *(figure 1.17)*.[2,5] The importance of this downstream duplex binding site towards the interaction with the substrate has been indicated in human and phage FENs. Each of hFEN-1 double mutations K244A/R245A and K252A/K254A impaired the cleavage significantly.[60] Similarly in T5FEN, K215A and R216A mutations of the extended basic residues resulted a significant deficiency in binding and cleavage.[57]
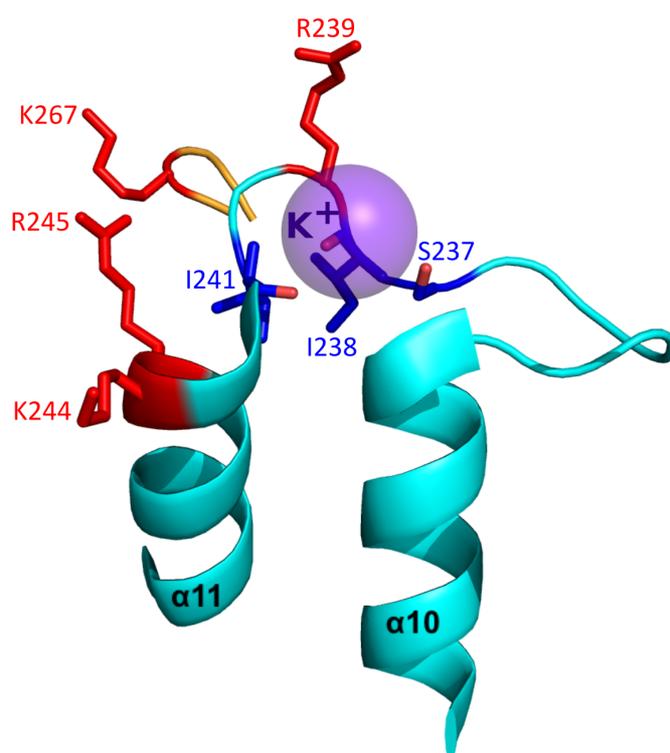
1.17. Domain structure of K^+/H2TH in hFEN-1 shows the residues that coordinate the K^+ using their oxygenic atoms and the residues that create a furrow in this region.

1.8.2. The hydrophobic wedge:

This protruding binding feature is made up from two regions: the α2-helix and α2-α3 loop hydrophobic wedges and the β6-β7 loop (β-pin) *(figure 1.16.c)*. Conformational changes have been observed in the α-helices and connecting loop and the positively charged β-bin side chains by comparison the crystal structure of the enzyme within and without substrate. This motif binds the substrate by interaction with some base pairs from the downstream and upstream regions.[2,5]

1.8.3. The 3'-flap binding pocket:

Structural studies of FENs have illustrated a conserved binding site for the 3'-flap from archaea to human. The first crystal structure of Archaeoglobus flugidus FEN-1 (AfFEN-1) in complex with a 3'-overhang DNA revealed the importance of the single 1 nt 3'-flap.[59] FEN-1 possesses a 3'-extrahelical nucleotide-binding pocket (a 3'-flap pocket) that contacts the 3'-flap to open and kink the bound DNA.[47] Through evolution, the amino acids that create this extrahelical 3'-flap pocket have been observed in archaea and human but not in phage. In hFEN-1 the amino acids L53, Q54, N55, T61, M65, K314, Q315, F316, S317 and R320 form the 3'-flap pocket region *(figure 1.18)*. These ten residues surround the 3' -flap forming non-sequence specific hydrogen bonds to the 3'-hydroxyl and van der Waals interactions to the 2-deoxyribose sugar, that nucleobase and the 3'-phosphate end. This binding motif comprises 1/8 of the hFEN-1:DNA binding surface. The presence of this flap besides the 5'-flap increases the reaction rates of FENs as this binding to this pocket is very important to orient the substrate and position the scissile phosphate close to the active site to facilitate the cleavage specificity.[1,5,58]
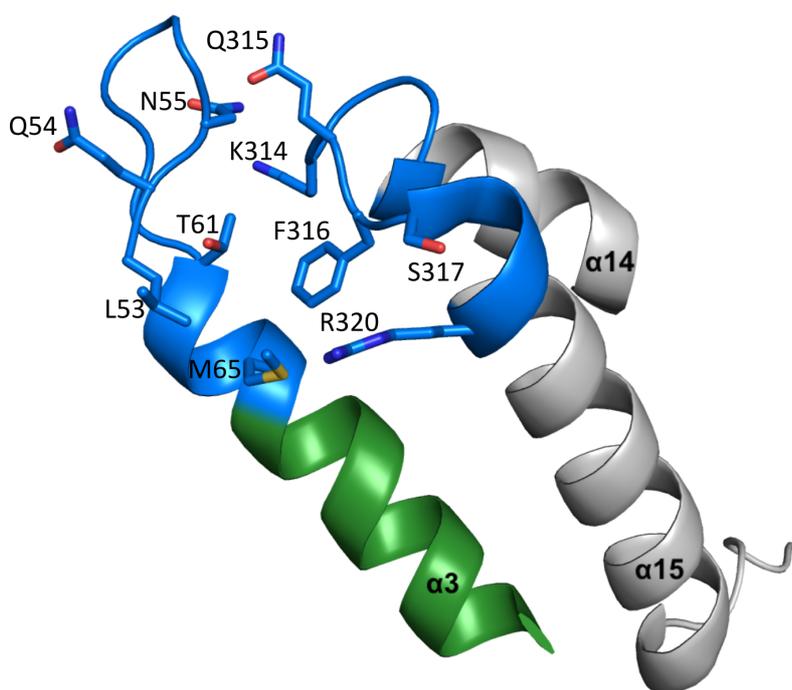
Figure 1.18. Domain structure of 3'-flap pocket shows their main residues that surround the 3'-flap of the substrate to be threaded into the pocket.

1.8.4. The helical arch:

Within the nuclease core domain, a major FEN-1 motif that limits the active site dimensions has been conserved throughout evolution. This DNA binding motif is known as the helical arch region. Its position above the active site mediates the interaction with the single stranded 5'-flap of the DNA substrates. In hFEN-1 this region comprises mainly charged and hydrophobic amino acid residues of the C-terminal part of α2 and all of α4 and α5. This region has been divided to two motifs that are the helical cap (a part of α4 and all of α5) and gateway (parts of α4 and α2) *(figure 1.16.c)*. While the gateway is conserved in all FEN-1 superfamily members, the cap region is only present in FEN-1 and EXO-1.By forming secondary structure, the helical arch creates a hole that allows ss5'-terminus passage but not the dsDNA structure. This mostly disordered region is ordered upon ss5'-terminus threading.[2,5]

1.8.5. The active site:

Like most nucleases, FEN-1 superfamily members are metalloenzymes. They are characterised with a long adaptable loop (arch) above a carboxylate rich active site that coordinate two essential metal ions cofactors. Structural studies of FENs active sites observed eight carboxylate residues in bacteriophages and bacteria, but only seven in all FENs of the higher organisms. These active site residues of FENs provide coordination positions for divalent metal ion cofactors that catalyse rate enhancement. Initial structures for crystallised FENs in bacteriophage T4 and T5, archaea, and human showed that the two metal ions bound to the active sites with varying spaces 6.3, 8, 5, and 3.4 Å respectively *(figure 1.19)*. The N-terminus in higher organism proteins (archaea and humans) seems to have replaced the "second" metal ion position in phage structures with the N-terminus, which seems to push the "second" ion towards the first metal ion in higher organism FENs. Despite of the exact roles of this positive charged N-terminus remaining unknown, their occurrence in the active site of the higher organism enzyme may suggest an important role during catalysis. Some studies suggest a third metal ion is transferred to the active site temporarily during the reaction of lower organism FENs to promote the two metal ion mechanism.[2,3,42,43,61,62]
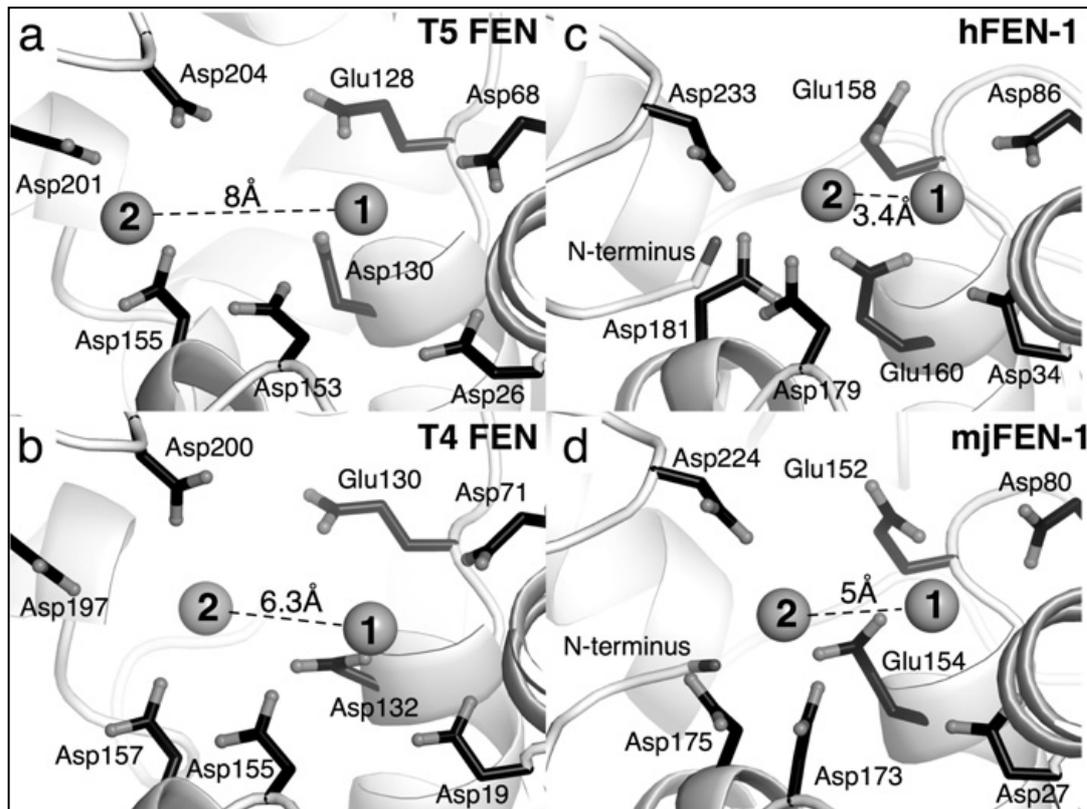
Figure 1.19. *The observed two metal ions in active sites of FENs.* ***a)*** *T5FEN (PDB: 1UT5).* ***b)*** *T4FEN (PDB: 1TFR).* ***c)*** *hFEN-1 (PDB:1UL1).* ***d)*** *mjFEN-1 (Methanococcus jannaschii FEN) (PDB: 1A77). Black arms illustrate positions of the active site carboxylate groups, and* grey spheres *(1 & 2) show the two metal ions of FENs active site. FENs of the lower organisms phage and bacterial (a & b) have an extra active site residue (T5 Asp[201], T4 Asp[197]), while, the N-terminus of the protein in the higher organisms human and archaeal (c and d) engages a similar position to metal ion 2. However, M1 is bound in a similar location in all structures, M2 position is changeable to vary the space between the two metal ions. The spaces between M 1 and M2 in the structures are a) 8, b) 6.3, c) 3.4 and d) 5 Å.*[3]

In hFEN-1 the two metal ions are coordinated by D86, E160, D179 and D181 to make the inner sphere, thus the other three residues of the active site D34, E158 and D233 and Y234 interact with the water molecules to make the outer sphere. All of these features with either N-terminus or a third metal ion specify the target phosphodiester bond to start the hydrolysis. Besides that, later studies suggest more interactions with the gateway aromatic residue Y40 from helix two and the gateway basic residues K93 and R100 from helix four stabilise the substrate during and after the chemistry *(figure 1.20)*. Consequently, they are presented as an active site residues in the new active site definition.[1,2]
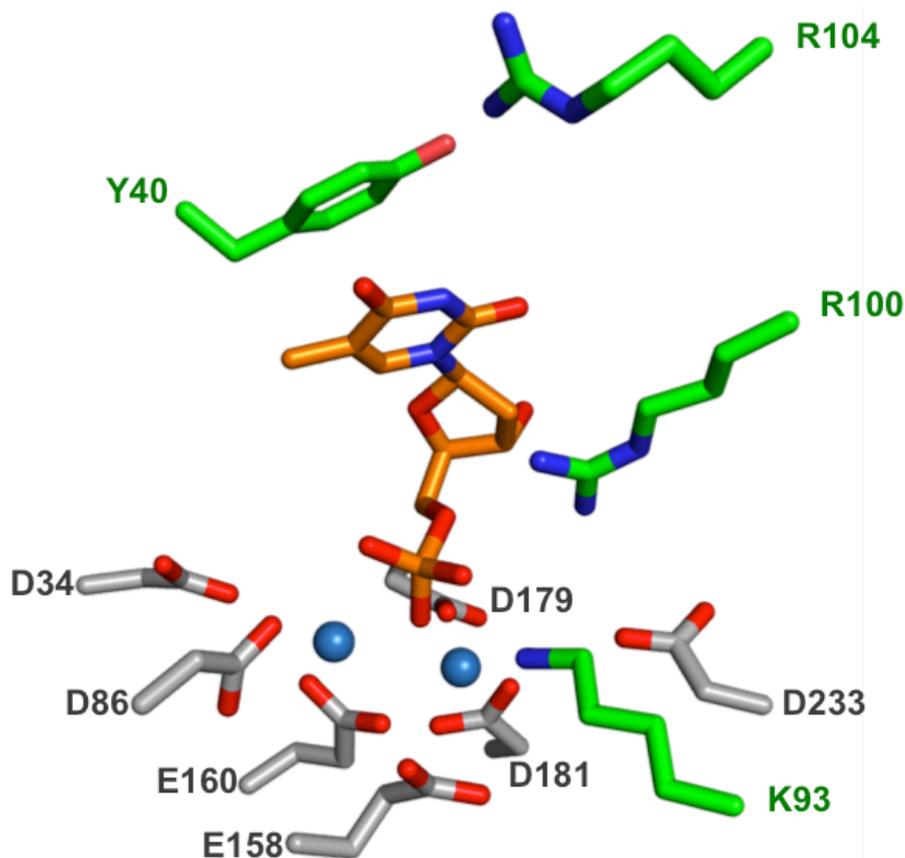
Figure 1.20. The main seven active site residues of hFEN-1 in product complex and their oxygen atoms to coordinate two Sm^{3+} metal ion and to form the sphere (PDB:3QK). Also includes the three basic residues K93, R100 and R104 from α4, and an aromatic residue Y40 from α2 that have roles to organise the 5'-end using their oxygen and nitrogen atoms.

A variety of divalent metal ions have been observed in living organisms such as Mg^{2+}, Ca^{2+}, Mn^{2+}, Fe^{2+}, Ni^{2+}, Cu^{2+} and Zn^{2+}, with Mg^{2+} at the highest concentration inside cells. Each is involved in specific enzymatic catalysis. Nucleases active sites are usually associated with Mg^{2+}.[12,52,63] This preferred binding of Mg^{2+} by nucleases possibly refers to its availability, solubility, redox stability compared to Mn^{2+}, Fe^{2+} and Cu^{2+}, its stringent coordination geometry compared to Fe^{2+}, Ni^{2+} and Cu^{2+}, small size compared to Ca^{2+} and its unusual hydration properties.[12] As indicated before, the divalent metal ion Mg^{2+} is essential for the nuclease function of FEN-1 superfamily as both an exonuclease and a structure specific endonuclease. Similar in vivo, Mg^{2+} is used to catalyse the phosphodiester bond hydrolysis to study these essential roles in vitro. However, the bigger sized ion Ca^{2+} is used to inhibit the hydrolysis.[12] For the structural study of FEN1, it is crystalized with two Sm^{3+} ions that occupy the two Mg^{2+} ions sites. This transition metal ion has similar size to Ca^{2+}.[5]

27

1.9. FEN-1:DNA interaction features:

The FEN-1 superfamily contains multifunctional enzymes with similar active site domains that are involved in nearly all DNA pathways (replication, recombination and repair). Binding of the enzyme to the favoured substrate ensures that the hydrolysis occurs strictly in the right position so that the product can be used for downstream processes.[64] The hydrolysis mechanism of FEN-1 in each organism requires a specific distance between the two metal ions (from their closest sites).[3] Although the absence of the two metal ions does not inhibit formation of the T4FEN:DNA complex as seen in a structure of this protein-substrate in the presence of EDTA, the target site of FEN hydrolysis (1nt into the downstream duplex) is not close enough to the active site for the hydrolysis to occur *(figure 1.13.c)*.[2,44] This suggests that the two metal ions are required to achieve better contact between the target phosphodiester bond and the active site.

Historically, several different mechanisms were proposed for FEN-1 substrate recognition. In spite of the variation between FENs in terms of favoured substrates, binding method, and mechanistic details, all of them share the common mechanism of phosphodiester bond selection using the two metal ions active site. Models of FEN-1 mechanism were proposed in previous studies to address how FEN-1 recognizes substrate. In one model of FEN-1 mechanism, which known as a tracking model, FEN-1 recognizes the 5'-flap end to push it self towards the ss/ds junction.[50] However, this mechanism was ruled out by the observations that beside double flap substrate, FEN-1 possesses 3'-SF and gap substrates with comparable efficiencies *(figure 1.12.a)*.[5,29,58] A second model is known as the threading model suggested that FEN-1 binds the DNA junction first then threads the flap. But also this mechanism raises concerns about the ability of flap threading through a small hole especially when accommodating gapped flaps.[58,65] A third clamping model suggested that FEN-1 folds its helical arch over the 5'-flap allowing interaction with the gateway residues.[29] However, the clamping model is not compatible with the need to confine reaction to free (discontinuous) 5'-ends.

An alternative disorder-thread-order mechanism is currently the preferred one. In this model, binding of the upstream and downstream double strand regions of dfDNA first, followed by 5'-flap threading through the disordered helical arch and 3'-flap engagement into 3'-flap pocket. Then, following ordering of the helical arch and 3'-flap pocket the scissile phosphate bond will be positioned into the active site.[2] In fact, many specific details about this mechanism are still unknown, for example, no evidence can suggest which of the flaps is organized first or if they are organized simultaneously. The overall disorder-thread-order mechanism as shown in *figure 1.22* is described below in steps that do not represent the order:
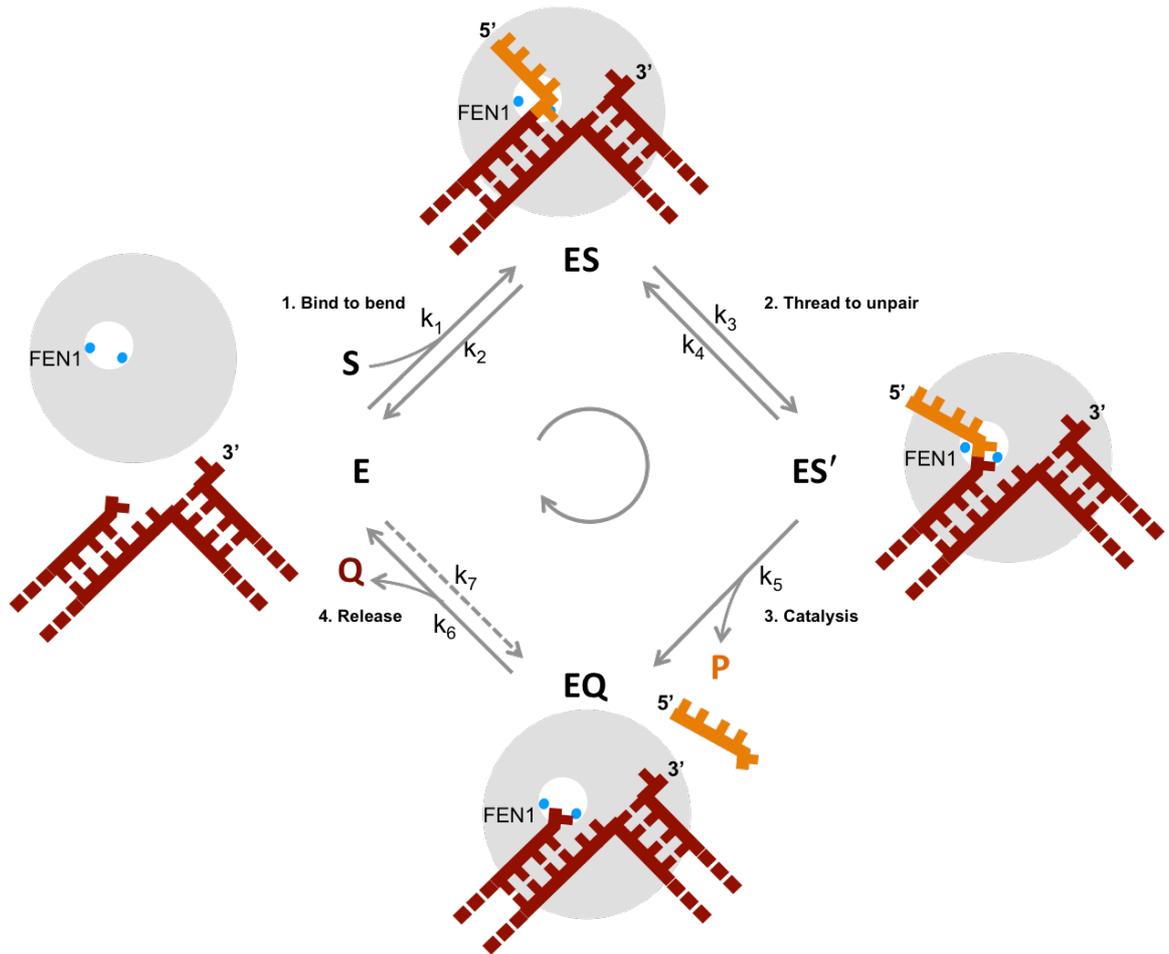
Figure 1.21. *A diagram of the disorder-thread-order mechanism of FEN-1 (E) and DF substrate (S) interaction deduced from FENs studies to date and there may be more steps in the reaction than illustrated.* **1)** *E binds S to bend S and to form the ES complex.* **2)** *To cleave the substrate, ES complex undergoes conformational changes creating a cleavage competent complex ES'.* **3)** *Catalysis creates 5'-flap (P) and dsDNA (Q) products. As the flap product releases the complex immediately upon cleavage, the Q product results as EQ complex.* **4)** *Dissociation of the EQ complex results in nicked DNA (Q) and free FEN1 enzyme.*

1.9.1. Binding:

FEN-1 binds the downstream portion of the DNA that undergoes reaction via the non-specific dsDNA interaction motif K^+/H2TH (*figure 1.22.a*). The majority of DNA interactions occur at the template strand via this region. The interaction occurs through a direct contact between the K^+ ion and the four protruding basic residues of this motif and the phosphate backbone of the complementary strand of the DNA *(figure 1.22.b)*. This interaction may allow fluid slide for the DNA minor groove through the furrow created by four protruding basic residues within the protein surface *(figure 1.22.c)*. After this binding region, the bound template strand returns back to

the protein surface introducing the substrate to more binding interactions that allow for bending of the DNA.[2,5]
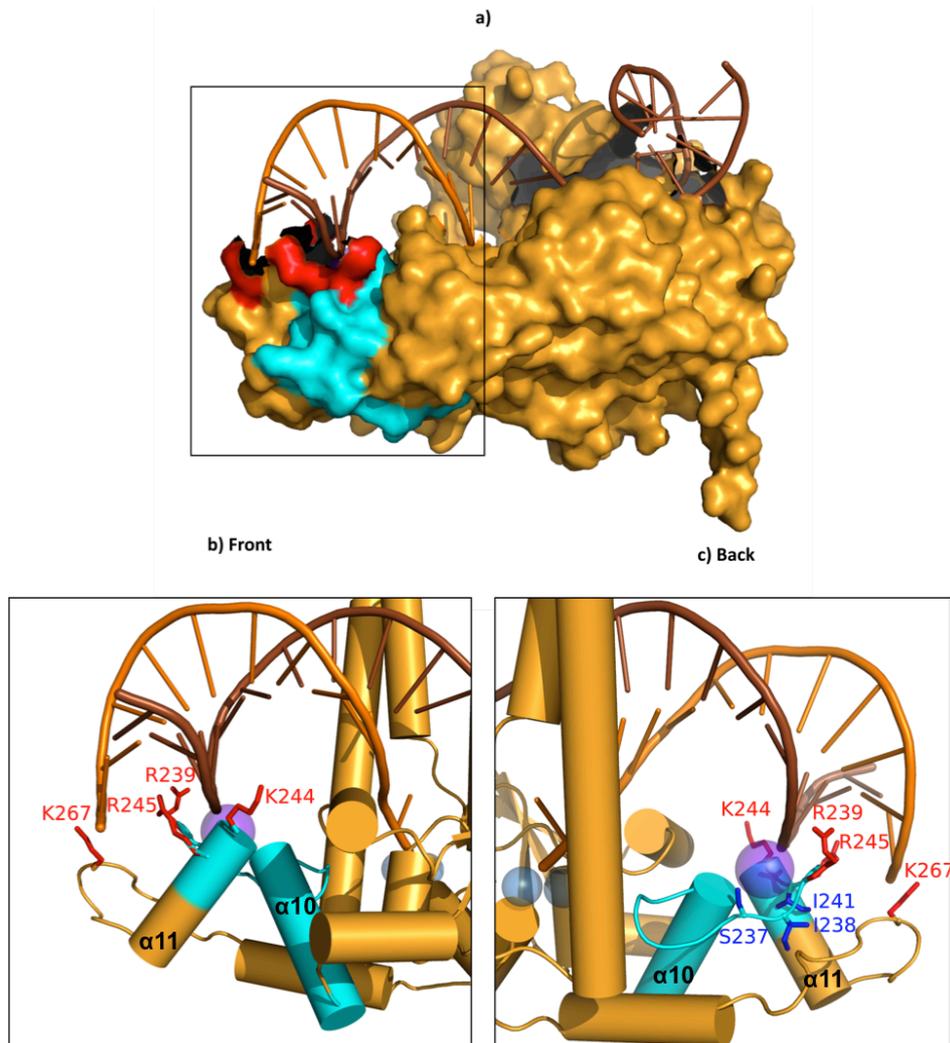
Figure 1.22. *Binding of the DF substrate by* **H2TH** *motif of hFEN1.* **a).** *Structure of hFEN1 product complex (PDB:3QK) shows binding of the* **downstream region of DF DNA** *by* **H2TH** *region and its fluid slide through the* **furrow** *of the region.* **b).** *Front view of the* **H2TH** *region shows interaction of The* **K⁺** *ion with the* **back bond of the complementary strand** *of the DNA and the* **basic residues** *that create the furrow within protein surface.* **c)** *Back view of the* **H2TH** *region shows its* **basic residues** *that create the furrow within protein surface and its* **residues** *that coordinate the* **K⁺** *ion.*

1.9.2. Bending:

Beyond the K^+/H2TH motif, arcs of the template strand become free from the protein surface. They re-join close to the active site by hydrophobic wedge motif. Introducing the substrate induces disorder-to-order transition of the α2-α3 loop and moves the β-pin to position the template. The reacting strand interacts with a basic residue that is provided by β-pin to pass under the template strand towards the hydrophobic wedge. Consequently, the α2 wedging helix

stacks between the base pairs of either side of the dsDNA junction. The hydrophobic wedge (β-pin and α helix amino acids residues) also binds the upstream template by an interaction with the last three base pairs (next to the 3'-flap) to bend the substrate *(figure 1.23.a.b).* This binding feature bends the substrate markedly at a single phosphodiester bond with ~100° between upstream and downstream regions of reacted DNA *(figure 1.23.a).* The bent conformation at the DNA junction is stabilised by an extra interaction between the last base pair face and some of the hydrophobic wedge residues. The bent substrate forms what looks like an arch from the template strand.[2,5] Ensemble FRET experiments confirmed that this bent conformational change of the DF substrate is stabilised by hFEN-1 binding.[47] The big conformational change of this binding interaction presents the two flaps of the DF DNA to the next two binding elements, which are the helical arch and the 3'-flap binding pocket.[5] There is currently no evidence confirming which accommodation of the 3'-flap and/or 5'-flap occurs after this initial binding and bending or even at what point the DNA is threaded nor whether there is equilibrium with a non-threaded form.
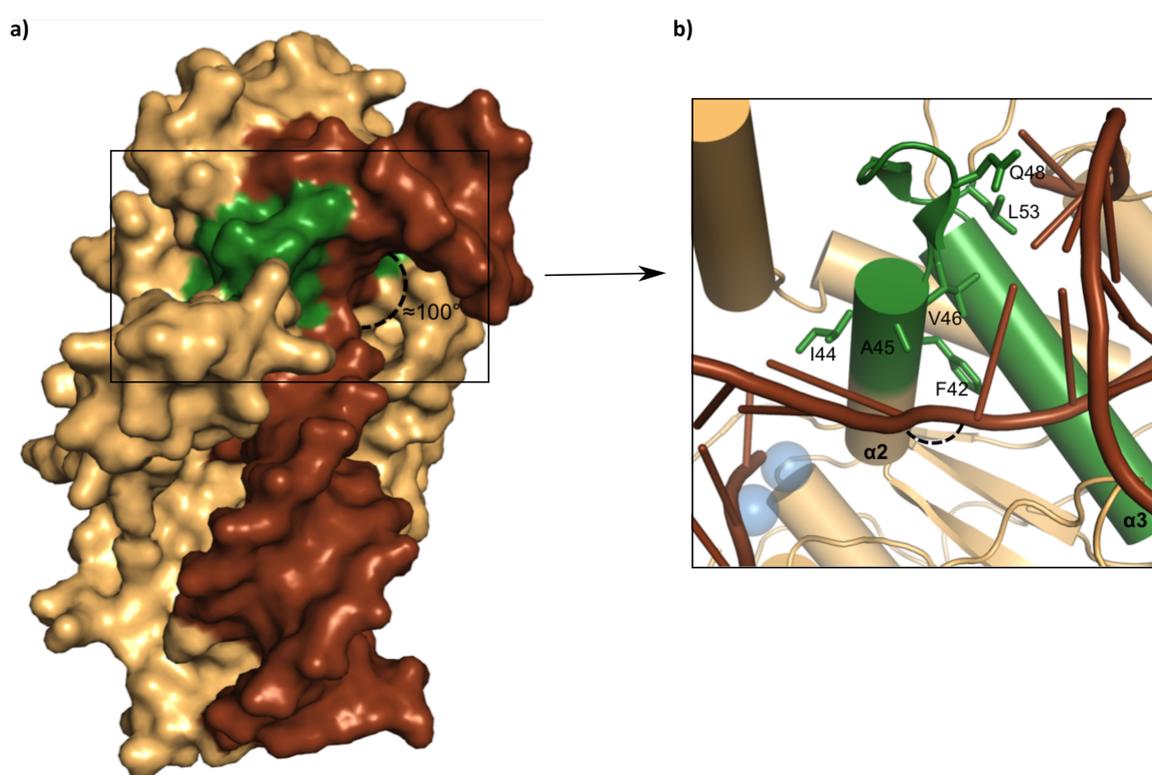
Figure 1.23. *Bending of the DF substrate by hFEN1* **wedge***. **a)** hFEN1 product complex (PDB:3QK) shows bent product at the junction between the* **upstream** *and* **downstream** *DNA regions by hydrophobic* **wedge** *motif.* **b)** *View of the* **hydrophobic wedge residues** *that stabilise bending by their interaction with the junction* **base pairs***.*

1.9.3. The 5'-flap threading:

It is known that the 5'-flap is threaded through the helical arch. After the 5'-flap is threaded, the arch residues (cap and gateway) go from a disordered-to-ordered helical state. The disorder-thread-order transition structures the cap and the gateway to limit the dimensions of the helical arch. The ordered gateway width allows only ssDNA passage while prevents entry of the dsDNA and this is could be another FEN-1 technique to organise specific DNA structures *(figure 1.24.a.b)*.[5] Any bulky addition to the 5'-flap such as streptavidin (SA) prevents the threading through the structured or unstructured helical arch resulting in a drastically slowed FEN-1 reaction.[50]

The requirement of ss-flap for the disordered-to-ordered threading mechanism raised many questions regarding the processing of 3'-SF substrate (without 5'-flap) and gapped DNA (5'-flap containing secondary structure) with FEN-1. Biological results are related to ss5'-flap length effects demonstrated that the 5'-flap length does not affect the binding affinity of the DF substrate.[51] However, the total removal of this 5'-flap creating an exo substrate (3'-SF) has small effects on substrate binding and cleavage,[66] which suggested that the 5'-flap is not required for disorder-to-order transition as this could be brought about by the first two downstream nucleotides. More evidence relating to gapped DNA clarified that a 5'-flap with secondary structure is excised efficiently by the gen activity of FEN-1 via the same mechanism of fen activity.[51,60] Theoretically, the binding of the complementary strand of the gap substrate orients the dsDNA structure of the 5'-flap into the disordered archway region. Consequently, the arch region must be reordered after threading of the ss-flap section.[51]

Regardless of the 5'-flap length or the presence of a short duplex within the 5'-flap, the mostly disordered region is ordered upon substrate binding and threading. This ordered region is stabilised through the interactions with the residues M37 and Y40 in α2 gateway, K93 and R100 in α4 gateway region, R103 and R104 in α4 gateway/cap region, and K125, K128, and R129 in α5 cap region.[2,5] Threading of the 5'-termini into the disordered helical arch positions the scissile phosphate near the active site with neither direct contact to the two metal ions nor the amino acid residues of the active site to undergo hydrolysis *(figure 1.24.c.d)*.[1,5] Accordingly, an extra force is required to position the target bond into the active site to start the chemistry.
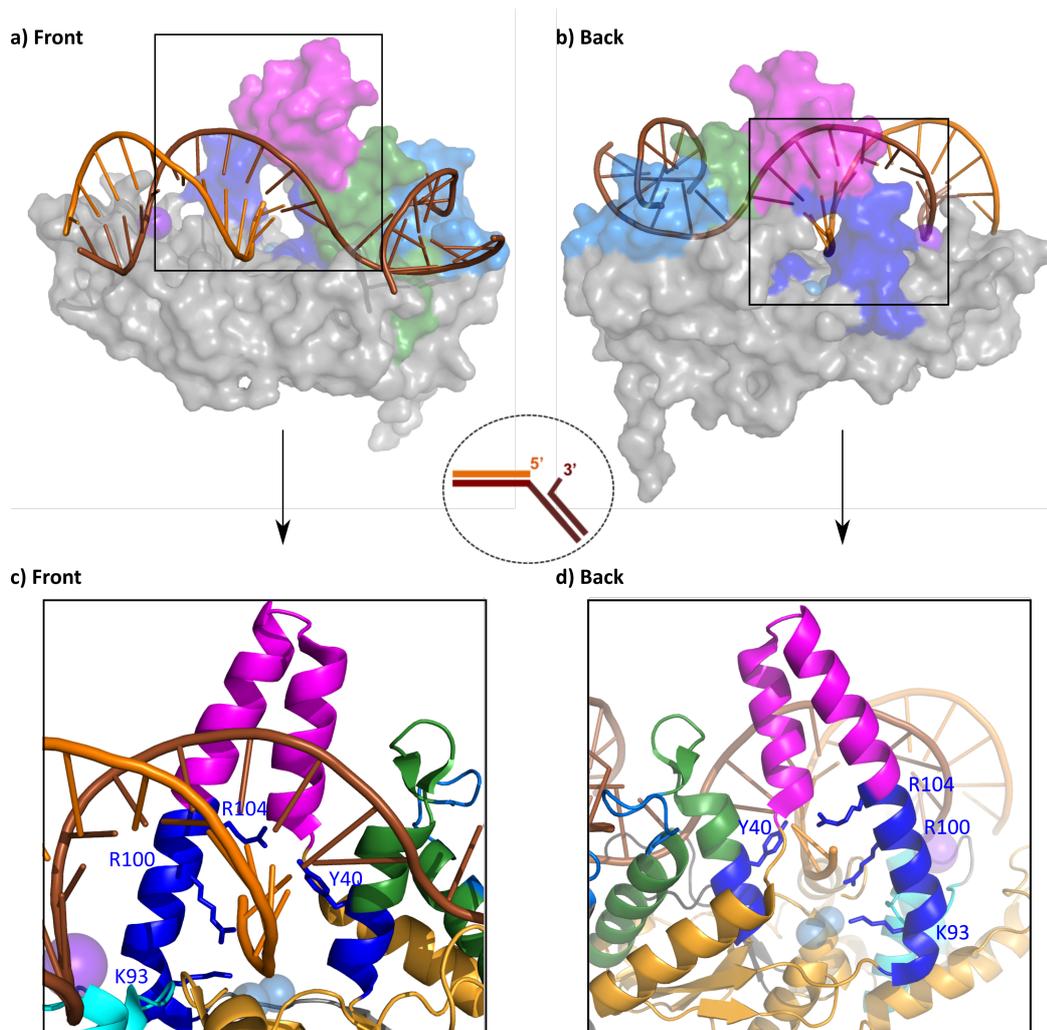
Figure 1.24. *Threading of 5'-end of substrate by hFEN1.* **a)** *and* **b)** *Front and back view, respectively, of hFEN1 3'-flap substrate complex (PDB:3Q8L) shows threading of* **5'-end** *through the helical arch (***cap*** and ***gateway***). c)** *and* **d)** *Front and back view, respectively, of the helical arch shows its* **residues** *from the* **gateway** *that may interact with the* **5'-end** *of the substrate.*

1.9.4. The 3'-flap engagement:

Along with two-way dsDNA junction bending by the hydrophobic wedges, the upstream duplex places the 1nt 3'-flap of the DF substrate near the 3'-flap binding pocket of FEN-1 *(figure 1.25.a)*. The ten amino acid residues of the 3'-flap binding pocket surround and interact with the last nucleotide of the upstream DNA region (1nt 3'-flap). Amazingly, this cleft can only fit a single unpaired nucleotide of the 3'-flap and recognises the single nucleotide regardless of base sequence *(figure 1.25.b)*. Therefore, most of the 3'-flap strand does not interact with FEN-1 although it is base-paired with the template strand. The 3'-flap specificity allows FEN-1 to recognise different DNA structures, which have 1nt 3'-flap such as double flap, 3'-single flap and forked-gap *(figure 1.14)*.[1,5] Presumably, all of these previous binding features together stabilise

the FEN-1:DNA complex and prevent flap migration and cause FEN-1 to be structure specific for dfDNA.[1,5] Recognition of the two-way DNA junction and threading of the two flaps capture the substrate initially to place the selected scissile phosphate close to the active site but not in the exact required position for hydrolysis *(figure 1.25.c.d)*.
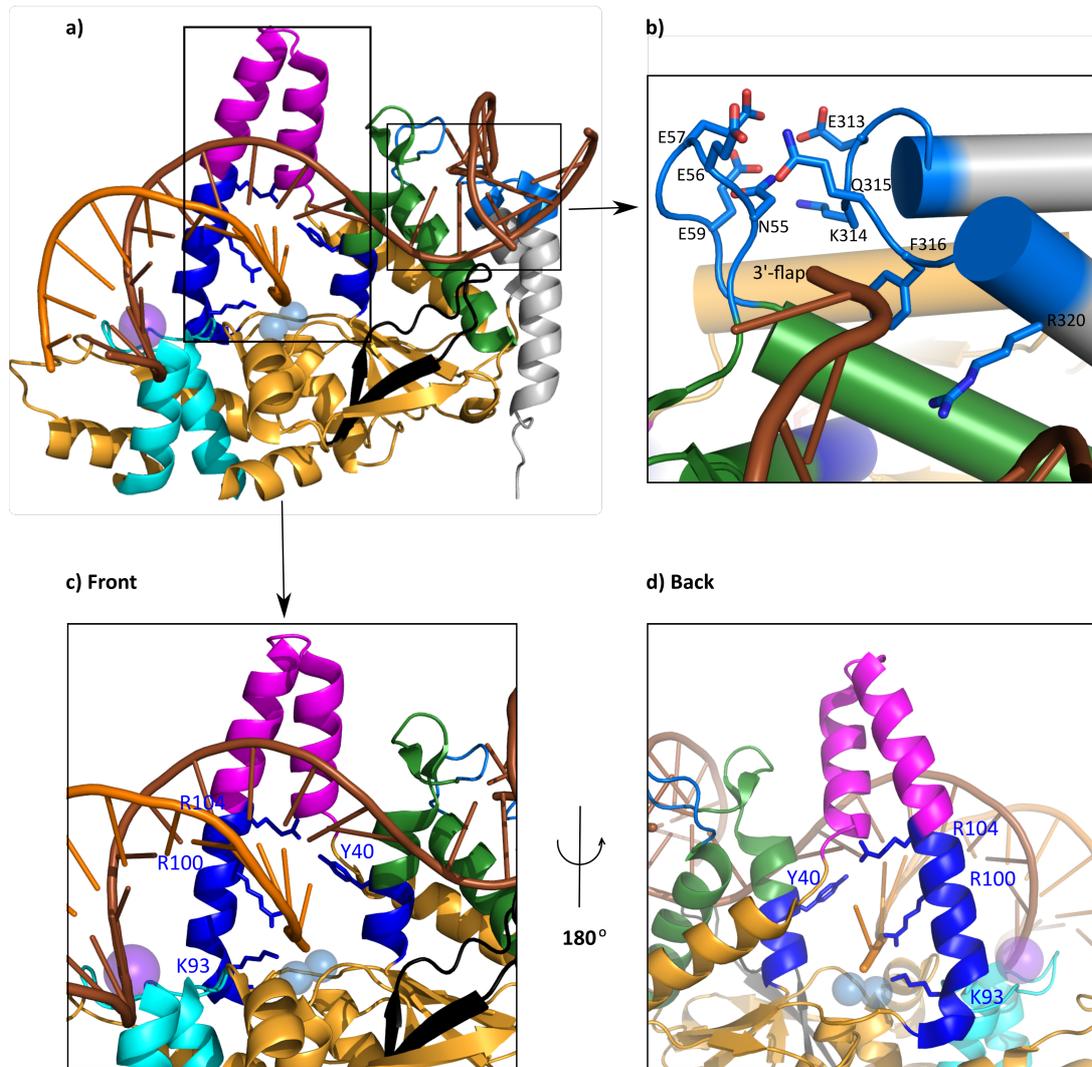
Figure 1.25. *Threading of 3'-flap of double flap substrate by hFEN1.* ***a)*** *Total view for threading of the 5'-end and 3'-flap of df-substrate in hFEN1 product complex (PDB:3QK) into the* ***helical arch*** *and the* ***3'-flap flap pocket*** *respectively.* ***b)*** *View of the 3'-flap pocket and its residues (****oxygen*** *and* ***nitrogen*** *atoms shown****) that surround the 3'-flap.* ***c)*** *and* ***d)*** *Front and back view, respectively, of the* ***helical arch*** *show the different position of the* ***5'-end*** *between its* ***residues*** *because of the* ***3'-flap*** *engagement compared with figure1.22.c and d.*

1.9.5. Shifting of the scissile phosphate bond to reach the active site metal ion:

In fact, all the suggested models of FEN-1 mechanism (tracking, threading and clamping) did not explain how FEN-1 specifies the target incision site and moves it onto the metal ion active site. This led to a focus on the molecular active site changes that occur during the chemistry by comparing between the enzyme substrate and enzyme product complexes. As a result a Double Nucleotide Unpairing mechanism (DNU) was proposed *(figure 1.26.a)*.

A structural study of EXO-1 and hFEN-1 enzyme substrate (ES) and product (EP) complexes showed base pairing of the +1nt and -1nt in the substrate complex. However, the product complex showed unpairing of the -1nt and small differences in the positions of -1nt and template strand *(figure 1.26.b)*.[5,29] The two metal ion sites in the crystal structure of hFEN-1 are occupied by Sm^{3+} ions that support binding but prevent cleavage. Despite neutralisation of the active site negative charge in hFEN-1 complexes, the target phosphate diester bond in the substrate complex was around 5.2 Å from either metal ion, however, the unpaired -1nt of the product complex was coordinated directly with the active site metal ions *(figure 1.26.c)*. In the substrate complex Y40 was stacked with the +1 nt, but in the product complex it was stacked with the unpaired and rotated -1 nt (≈25°). The template strand showed a slight difference in the position in both the substrate and product complexes because of its interaction with the hydrophobic wedge (α2 and β pin).[5] These differences between the substrate and the product complexes suggested a double nucleotide unpairing mechanism occurs at +1 and -1 nts.[5]

This DNU mechanism was suggested to promote localisation of the target phosphate bond (one nucleotide into the downstream) between the two metal ions and toward the carboxylate rich region that stabilises this unpaired state *(figure 1.26.a)*.[3,44,64] Thus, there is no evidence to suggest if the disorder-to-order transition step is a prerequisite for the double nucleotide unpairing. Furthermore, it is unknown if this duplex end unpairing is necessary for the chemistry to release the 5'-flap by locating the scissile phosphate bond into the right position or is it a step to prevent hFEN-1 reaction with the duplex DNA.[2,5,37]

More recently, collaborative structural studies have revealed that unpairing does not take place, but the DNA becomes untwisted. This is dealt with in further detail later, nevertheless there is a marked substrate conformational change at the end of the downstream reacting duplex.
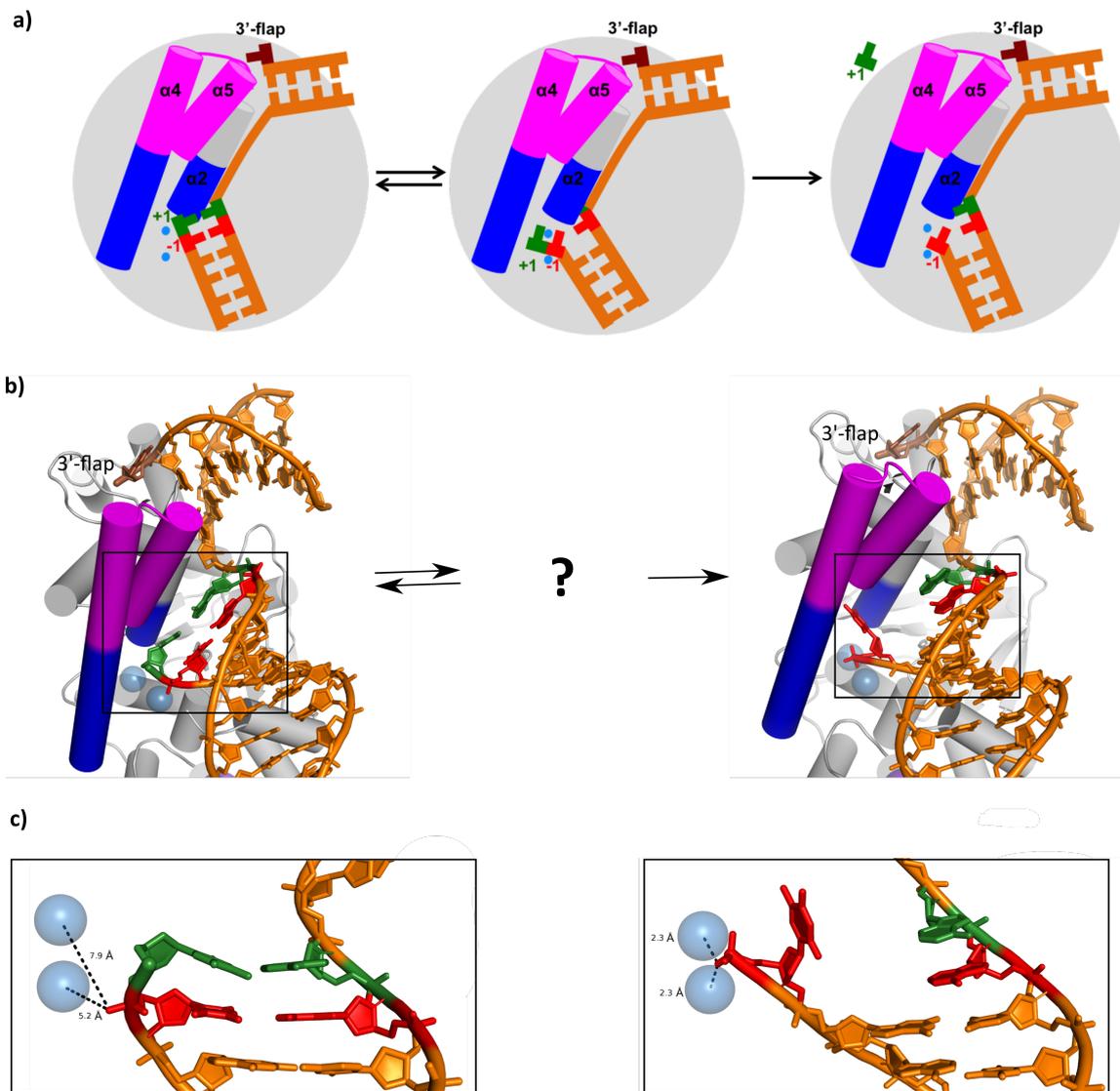
Figure 1.26. *Double Nucleotide Unpairing (DNU) mechanism is conserved across FEN superfamily.* ***a)*** *Cartoon of the DNU mechanism represents orientation of the target phosphodiester bond in the enzyme active site between the* ***two metal ion*** *sites.* ***b)*** *Complexes of hFEN-1 with substrate (PDB:3Q8L) and product (PDB:3QK) that support the DNU mechanism. In the ES complex the two nucleotides,* ***+1*** *and* ***-1*** *that are on either side of the target phosphodiester bond, are still paired. In the EP complex the terminal* ***-1*** *nucleotide is unpaired. Therefore, the* ***+1*** *and* ***-1*** *nucleotides are assumed to be unpaired. The* ***helical gateway*** *is the base of α4 and α2, and the* ***helical cap*** *is top of α4 and α5.* ***c)*** *Comparison of the* ***substrate*** *and* ***product*** *DNA near the active sites shows that double nucleotide unpairing (****+1*** *and* ***-1*** *nucleotides) allows direct interaction between the scissile phosphate and* ***two metal ion*** *active site.*[67]

1.9.6. Catalysis and product release:

The initial binding of FEN-1 to the dsDNA two-way junction by the H2TH and wedge regions, the consequent bending at the DNA junction, accommodation of the two flaps, and the double nucleotide unpairing or untwisting are conformational changes that position the scissile phosphate bond into the active site. As DNA conformational change is proposed to localize the target phosphodiester bond (between +1 and -1 downstream unpaired nucleotides) more close to the

active site, one metal ion faces the 5' side and the other one faces the 3' side of the scissile phosphate. The active site residues Y40 (in α2), K93 and R100 (in α4) are the key for efficient DNA positioning to start the hydrolysis reaction and stabilise the resulted cleavage *(figures 1.27.a.b)*. Structural data position the two metal ions in coordination variously with the attacking nucleophile, leaving group and phosphate oxygens to start the chemistry.[68]
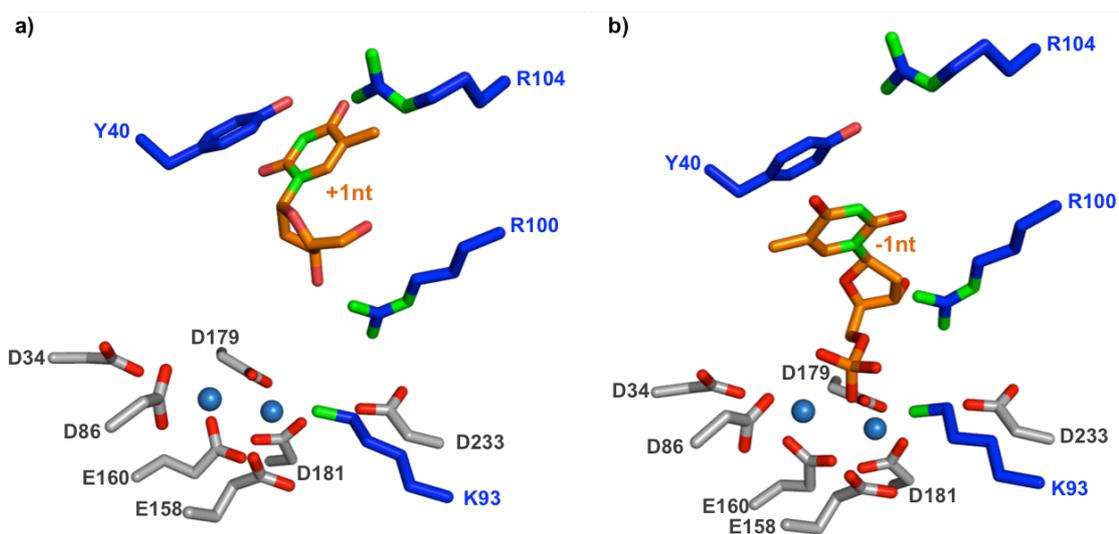
Figure 1.27. *The key residues of the* *active site* *and* *helical gateway* *of hFEN-1.* *a)* *and* *b)* *The different positions of the terminal nucleotides* *+1* *and* *-1* *in hFEN1* *substrate* *and* *product* *complexes respectively. This illustrates the coordination of key residues with the respective nucleotides in the different states. The* *oxygen* *and* *nitrogen* *atoms of the side chains were shown.*

The presence of the two metal ions active site is the common and significant feature of the phosphodiesterases. It has been suggested that hFEN-1 uses the two metal ion mechanism proposed previously for 3'-5'exonuclease.[12,69] In this mechanism, the phosphodiester bonds are hydrolysed using an addition-elimination type mechanism in harmony leading to inversion of configuration at the central phosphorus *(figure 1.7)*.[12,70] The two metal ion chemistry is claimed to have the advantage of high substrate specificity and efficient product release.[12] The active site metal ions interact with the non-bridging phosphate oxygens of the substrate. In such a way, they may catalyse the reaction by acting as a Lewis acid factor (electron-accepting). The Lewis acid effect of one or more Mg^{2+} increases the electrophilic power of the phosphorus atom of the scissile bond *(figure 1.28.a)*. This environment maybe catalyses the reaction by enhancing the nucleophile formation, by locating the nucleophile in-line with the target bond or by stabilising the leaving group.[61,68] Another mechanism suggests a direct interaction with only one of the non-bridging oxygens of the scissile bond with one of the active site metal ions, while the other

one is exposed to the solvent (figure 1.28.b).[12] At all events, the direct and indirect interactions between the active site metal ions and the non-bridging oxygen are suggested to develop the positive charge on the phosphorus with simultaneous with nucleophilic attack and breakage of the P−O bond to the leaving group.[12,68]
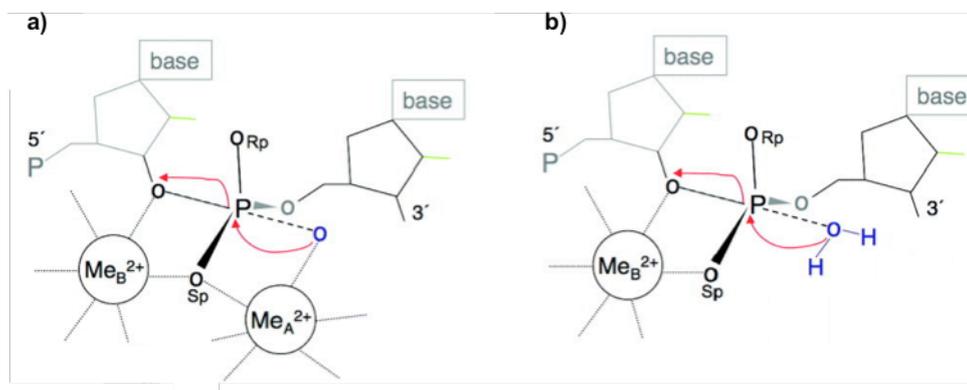
Figure 1.28. *Active site metal-ion- dependent mechanism of the phosphodiesterases.* **a)** *The two metal ion mechanism. Both metal ions coordinate the O_{SP} of the scissile phosphate bond from both sides 5' and 3'.* **b)** *The one metal ion mechanism. While one metal ion coordinates the O_{SP} of the scissile phosphate bond, the metal ion A is eliminated.*[12]

1.10. FEN-1 mutations and their link to cancer:

Unrepaired damaged DNA causes accumulation of genetic mutations and genomic instability that together can result in or exacerbate human cancers.[71,72] Genetic mutations are any change or alteration in the composition of DNA sequence other than recombination that arise from either outside factors (induced) or natural processes (spontaneous). Gene mutations can be manifest at two categories that are insertion/deletion mutations and point mutations.[6] [18] Earlier studies focused on the cause of mutations due to external effects. For example, smoking and asbestos exposure were suggested as main causes to increase the risk of lung cancer. Eventually it was found that the tobacco smoke and the asbestos fibres contained a large amount of DNA-damaging agents and mutagens that increase cell mutations.[73] Later studies focused on alternative causes of genetic instability that underlie increased mutations in cells and cause cancer. DNA mutations can mutate proteins by altering their amino acids sequence and this may affect on protein function.[74]

Protein expression is the means by which cells synthesise the needed particular protein from appropriate nucleotide sequence of DNA. The DNA sequence contains the genetic information that is needed for protein synthesis, but it is used indirectly when it is transcribed into RNA and then translated to protein. In eukaryotic cell nuclei, pre-mRNA is transcribed as an intermediary

molecule of DNA, which consists of both coding (exons) and non-coding (introns) regions. Via RNA splicing, it is modified to remove introns and produce the mature mRNA that is then transported to the cytoplasm to start translation in ribosomes. Each triplet codon is translated to one amino acid to produce the final protein.[6] DNA mutations affect protein expression in different ways. An insertion/deletion mutation could change the amino acids and produce either too short or too long as the reading shift can code the stop codon too early or too late, or it could lead to a frame shift mutation if the inserted or deleted base pairs are not a multiple of three nucleotides and lose protein function (non-functional proteins). Missense and nonsense point mutations affect the resultant protein by changing one amino acid to another or stop its translation respectively, however, a silent point mutation does not affect the produced protein sequence but it can affect its synthesis limiting its amount.[74,75] Indeed, mutated genes are present in all cancer cells.[74] Thus, a single event such as a mutation's effect on function of protein can increase the rate of accumulation and cause tumour progression.[76,77]

Identification of mutations in enzymes that have no impact on gene expression in cancer cells, but that alter enzyme activity and produce mutations due to the impact on DNA metabolism strongly suggest a link between altered DNA repair/replication and cancer.[18,76] FEN-1 is an indispensable factor to maintain genome stability and integrity because of its important roles in DNA metabolism.[78,79] Because FEN-1 mutations such as I39T, Q112R and S317F have been indicated in many human cancer specimens, it becomes very important to understand the exact relationship between FEN-1 mutations and cancer susceptibility.[18] In fact, not all of FEN-1 mutations observed in cancer patients can cause cancer progression, but only those effect the FEN-1 activities and protein-protein interactions appear to produce effects on DNA replication and repair.[48] Functional defects in FEN-1 have been suggested to affect genomic stability and the development of cancer.[18,76,80] More evidence was suggested from mice models to identify the relationship between mutant FEN-1 and tumourgenesis of the gastrointestinal tract. Mice containing E160D mutation showed rapid progression of cancer and decreased life expectancy. Accordingly, defects of FEN1 function does not alter cancer initiation process but has a deep effect on cancer progression because of its important roles in DNA repair.[76] This defect of FEN-1 function was highlighted as a main cause of accumulation of mutations that underlie cancer.[80] Contrary to the previous findings, a later study did not indicate any mutations of FEN-1 or a decrease in protein expression in human lung cancer cells. In this case, defects of FEN-1 function were involved in progression of human lung cancer. Otherwise, cancer cells showed increase of FEN1 expression that attributed to the response to the increase of DNA damage.[81] Further support was introduced from a study to evaluate FEN-1 expression in breast and ovarian human cancer

cells. By indicating of FEN-1mRNA, the enzyme showed over expression in most cases of cancer and this could be an indication of genomic instability in human cancer.[82] In spite of the wide acceptance that accumulation of mutations in the genes that directly control genome stability causes cancer, mechanisms of mutation generation remain the subject of continuing debate.

Attempts to understand the implications of FEN-1 mutations have focused on consideration of FEN-1 as a multifunctional enzyme with fen, exo and gen activities (section 1.6).[48] One study proposed that an E160DmFEN-1 point mutation, which results in lung cancer in mouse models abrogates the exo and gen activities of FEN-1 while retaining fen activity to cleave flapped substrates. Thus, deficiency in the minor activities of FEN-1 is claimed to increase spontaneous mutations and accumulation of unprocessed DNA fragments in cells. It is suggested that, consequently, cells will be predisposed to chronic inflammation that promotes tumour progression.[18] In another study of the E160DmFEN-1 point mutation, reduction of fen activity of FEN-1 was claimed to be the result of the protein alteration. Accordingly, decreased flap cleavage capability was claimed to increase the incidence of cancer, especially lymphoma, due to resulting problems with DNA replication.[19] A later study identified that E359KmFEN-1 mutation, which has been indicated in human breast cancer, in mouse models affects the protein-protein interaction and reduces gen activity of FEN-1 and this leads to defective FEN-1 ability to process bubble structures that are formed during replication.[83] Furthermore, a recent study claims to show that L209PhFEN-1, which was identified in human colon cancer, point mutation develops cancer in humans because of the deficiency in fen, exo and gen activities of FEN-1.[84]

Despite the substantial knowledge of that FEN-1 mutations affect its function and this leads animal models to develop cancer, many issues such as the exact effects of FEN-1 mutations remain to be addressed.

1.11. Aims of project:

Although FEN-1 structure and function have studied widely, still lots of outstanding questions that need explanation. For example how the enzyme can recognise different substrates using the same active site, how FEN-1 recognises the substrate and achieves reaction at a specific bond, how mutations, which were indicated in many cancer cells, affect FEN-1 ability to recognise any substrates. Many answers have emerged through the previous studies. While many suggestions were acceptable and successful, some others were contradictory and not convincing.

In the previous studies, FEN-1 was considered as a multifunctional enzyme with different activities. By observing FEN-1 mutations in numerous different human cancer kinds, many of them were studied to indicate their effects on these different activities. In 2007, effect of FEN-1 mutations in cancers was tested on FEN-1 point mutation generated in mice by using a gene targeting approach. As a result, most of mice with E160D point mutation spontaneously developed cancers, mainly in the lungs. In addition, above 60 % of cancers in these E160D mice have other mutations in *Kras, p53, Apc or Egfr*. According to the nuclease activity assay, it was claimed that many of these mutations such as A159V and E160D led to loss of exo and gen cleavage abilities, but had no effect on fen cleavage ability of mFEN-1 *(figure 1.29.a.b)*. Consequently, this could be the main reason that lead to increase incidence of cancer especially lung cancer. [18] A later study in 2008 showed that the E160DmFEN-1 point mutation leads to reduced fen nuclease activity *(figure 1.29.c.d)*. Accordingly, decreased flap cleavage capability was proposed as the main cause of increasing incidence of cancer, especially lymphoma due to resulting in problems with DNA replication.[19] To facilitate studying of mutations' effects, many FEN-1 mechanisms were suggested. The disorder-thread-order mechanism was the preferred one that facilitates studying of FEN-1. Although many details remain to be understood, the mechanism was summarized in simple steps, which are binding, bending, 5'-flap threading, and 3'-flap engagement, and scissile phosphate bond possition.[2]
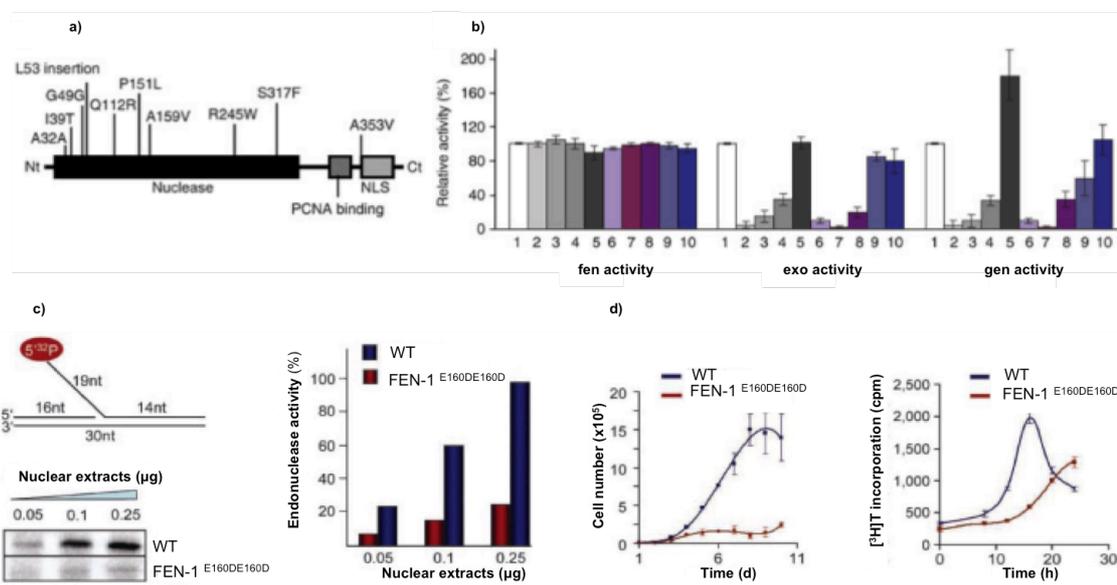
Figure 1.29. From 2007 study on mFEN-1: a) hFEN-1 mutations that were observed in numerous cancer using direct DNA sequencing. b) The profiles of FEN-1 activities and its mutant variants: 1 is WT, 2 is E160D and 7 is A159V. Both mutations reduced exo and gen activities without effect on fen activity.[18] From 2008 study on mFEN-1: c) Left: Structure and cutting of 5'-flap DNA. Right: Comprisal the endonuclease activity of WT and E160D. d) Left: The proliferation of WT and E160D in low density conditions. Right: The cell cycle progression of WT and E160D.[19]

Based on the previous mentioned observations; this project aims to study both mouse and human FEN-1. In mFEN-1 study, WTmFEN-1 will be first studied kinetically with different substrates to verify if the enzyme actually has different activities or just recognises different substrates using same chemistry. In addition, the two active site mutations A159V and E160D mutations, which have been shown to have contradictory effects on the enzyme activities, will be studied to indicate their effects on the enzyme efficiency, substrate conformational changes and substrate bending. Finally, this combined kinetic and structural study may also help to illustrate some other points such as the preferred substrate of FEN-1 and the main requirements for the enzyme mechanism. A study of hFEN-1 will focus on effects of some different mutated residues from the arch and 3'-flap pocket on the local DNA conformational change and substrate bending. This hFEN-1 study will be done to complete the kinetic study, which has been done by our group members Dr. Nikesh Patel, Dr. Jack Exell and Dr. Mark Thompson. As in mFEN-1, also this combined study can help to add some explanations to the general proposed hFEN-1 mechanism.

Chapter 2: Methodology:

2.1. WT and mutant mFEN-1 site-directed mutagenesis and protein expression:

Buffers, media & gels required:

2X YT media (1 L)	16 g tryptone, 10 g yeast extract, 5 g NaCl, H_2O, adjust pH to 7.5 with 5N NaOH.
SOC media (1 L)	20 g trypton, 5 g yeast extract, 0.584 g NaCl, 0.186 g KCl, 1.20 g MgSO4, 3.60 g glucose, H_2O.
TY media (928 ml)	12 g Tryptone, 24 g yeast extract, H_2O.
50X 5052 media (100 ml)	2.5 g glucose, 25 g glycerol, 10 g α-lactose, H_2O.
AI media (500 ml)	464 ml TY media, 25 ml 20XP, 10 ml 50X 5052 media, 1 ml 1 M $MgSO_4$, 100 ul 10^3X Trace metals, supplemented with 500 µl of chloramphenicol (34mg/ml) and 4 ml of kanamycin (50 mg/ml).
LB/Agar plate's media (1 L)	10 g Tryptone, 5 g yeast extract, 10 g NaCl, 15 g agar, H_2O. (+ Antibiotic: 50 µg/ml kanamycin or kanamycin + chloramphenicol).
Cell lysis buffer (100 ml)	100 ml IMAC buffer A with a dissolved tablet of SIGMAFAST Protease Inhibitor Cocktail (EDTA free).
1X PBS (1 L)	8 g NaCl, 0.2 g KCl, 1.44 g Na_2HPO_4, 0.24 g KH_2PO_4, adjust pH to 7.4.
50X TAE buffer (1 L)	242 g Tris base, 57.1 ml glacial acetic acid, 100 ml 500 mM EDTA (adjust pH to 8), H_2O.
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)	1 g agarose, 100 ml 1X TAE, 10 µl SYBR safe DNA gel stain (10,000X).
Acrylamide gel	Resolving (12 %): 3.4 ml H_2O, 4 ml Acrylamide/bis-Acrylamide (30 % solution), 2.5 ml resolving gel buffer, 100 µl APS (10 %), 10 µl TEMED, 50 µl TCE. Stacking (4 %): 3.4 ml H_2O, 830 µl Acrylamide/bis-Acrylamide (30% solution), 680 µl stacking gel buffer, 50 µl APS (10 %), 10 µl, 5 µl TEMED.

2.1.1. Site-directed mutagenesis:

All of the DNA sequences with the target mutant codons were generated using site-directed mutagenesis. The WT, E160D, and A159V mFEN-1 were generated by PCR using a specific pair of primers, designed with mismatching nucleotides, for each mutation. The WTmFEN-1 plasmid (pet-28b-mFEN1) was sequenced (University of Sheffield Genomic sequencing facility) to confirm that it is a free of mutations and a completed sequence to be used in the expression. First, this WT plasmid was used to prepare E160DmFEN-1. Second, this mutated plasmid was used to prepare more WTmFEN-1. Finally, the new WTmFEN1 plasmid was used to prepare the A159VmFEN-1 plasmid.

2.1.1.1. PCR reaction:

The E160DmFEN-1 primers were designed with the Agilent QuickChange Primer Design tool (http://www.genomics.agilent.com/primerDesignProgram.jsp). For each mutation point (insertion or deletion) a different pair of primers was used, as shown in *(Table 2.1)*.

pET-28b mFEN-1- primers	Primer sequence	Length (nt)	From	To
E160D-F **E160D-R**	5'-gcacccagcgaggcagatgccagctgt-3' 3'-cgtgggtcgctccgtctacggtcgaca-5'	27	WT	E160D
D160E-F **D160E-R**	5'-gcacccagcgaggcagaggccagctgt-3' 3'-cgtgggtcgctccgtctccggtcgaca-5'	27	E160D	WT
A159V-F **A159V-R**	5'-cacccagcgaggtagaggccagctg-3' 3'-gtgggtcgctccatctccggtcgac-5'	25	WT	A159V

Table 2.1. The primers were used for site directed mutagenesis of mFEN-1 contain the mutated bases.

Mismatched primer solutions for the required mutation (100 µM) were prepared by re-suspending the dehydrated primers in Milli-Q^{TM} water. The Pair-primer (PP) solutions (100 µl, 3 µM) were prepared by mixing 3 µl of each of the forward and reverse primer solution with 94 µl of H$_2$O. The PP solutions were used for PCR mutagenesis reactions, along with the reagents described in *table 2.2*. The reactants were mixed together and thermo-coupled using a PCR machine (BIO-RAD) with a program as shown in *figure 2.1*. The next day, the solution was incubat-

ed at 37 °C for 1.5 hours after adding 1 µl of DPN1. Mutagenesis reactions were analysed by running 5 µl on an agarose gel and staining with Sybr Safe stain.

Ingredient	Volume (µl)	Ingredient's role
Pair-primer solution (3 µM)	2.5	Needed to amplify the target DNA fragment.
10X PfuUltra HF Reaction Buffer (bought from Agilent)	2.5	To create an environment for optimum activity of Taq DNA polymerase.
PfuUltra Hotstart DNA polymerase (2.5U/µl)	0.5	Enzyme helps catalyse the polymerisation of the deoxynucleotides into a DNA strand.
Template plasmid (79 ng/µl)	1	Contains the target fragment to be amplified.
dNTP (10 mM)	2	Mix of nucleotides to build the new DNA strands.
Milli-Q water	16.5	To adjust the concentration of each component.
Total	25	/

Table 2.2. PCR reaction mixture and role of each ingredient.
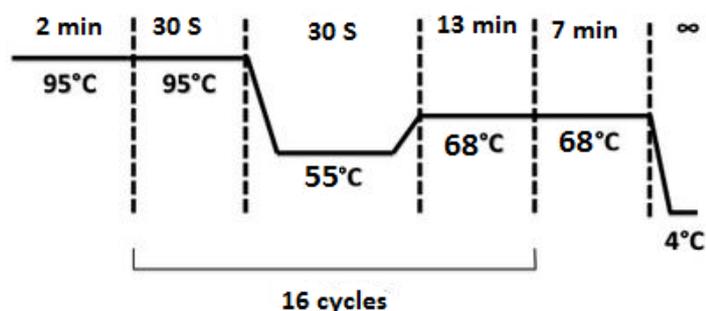
Figure 2.1. The PCR mutagenesis reaction program.

2.1.1.2. Transformation of cells:

After monitoring the DNA formation using an agarose gel, 2.5 µl of the mutagenesis solution was added to 50 µl of DH5α cells, which had been thawed on ice for 10 minutes. The cells were incubated on ice for 45 minutes. Then the cells were heat shocked in a water bath at 42 °C for 90 seconds, and then incubated on ice for 5 minutes. 1ml of SOC media was added to the transformed cells, which were incubated at 37 °C, 200 rpm for 1 hour. Subsequently, the cells were centrifuged at 13,400 rpm for 5 minutes. Next, the media was removed and the cells were resuspended in 250 µl of SOC media. 150 µl of the cell mixture was spread on LB agar plates sup-

plemented with 50 µg/ml of kanamycin. The plates were incubated at 37 ^{o}C overnight. The transformed cells were observed on plates as single colonies.

2.1.1.3. Plasmid preparation:

A single colony of cells was added to 10 ml of 2XYT media supplemented with 10 µl of 50 mg/ml of kanamycin (3 tubes for each plasmid) and left to grow for 14-16 hours overnight at 37 ^{o}C, 200 rpm. The starter cultures were grown until an OD at 600 nm ranging from 0.5 to 1.5 of 1:10 dilution was achieved.

The starter culture from each tube was divided into 6 of 1.5 ml tubes and centrifuged for 10 minutes at room temperature to pellet the cells. A mini-prep kit (Qiagen) was used to collect the plasmids. First, the pelleted cells were re-suspended in 250 µl of P1 buffer. Then, 250 µl of buffer P2 was added and the tube was mixed slowly by inverting until the solution turned blue. 350 µl of buffer N3 was added and the tube again was mixed slowly by inverting until the solution turned colourless. The mixture was centrifuged for 10 minutes to separate the insoluble portions.

The supernatant was applied to the QIA-prep spin column, which was then centrifuged for 1 minute and the flow-through was discarded. The QIA-prep spin column was washed by adding 500 µl of PB buffer, and then centrifuged for 1 minute and the flow-through was discarded. This step was repeated again by washing the column with 750 µl of PE buffer. The column was centrifuged again for 1 minute to remove all remaining wash buffers.

The QIA-prep spin column was placed in a clean 1.5 ml centrifuge tube. 50 µl of MilliQ water was added, and the plasmid was collected by centrifugation for 2 minutes. The plasmid concentrations were determined by absorbance at 260 nm using the Nanodrop UV spectrophotometer. The prepared plasmid solutions were then sent for sequencing. The sequencing results are shown in the appendix from *section 7.1.1 to 7.1.6*.

2.1.2. Protein expression:

BL21DE cells were transformed using 2 µl of the chosen plasmid solution using the same protocol described above.

A single colony of cells transformed with each plasmid was added to 25 ml of 2XYT media supplemented with 25 µl of chloramphenicol (34 mg/ml) and 12.5 µl of Kanamycin (50 mg/ml) and left to grow for 14-16 hours overnight at 37 ^{o}C, 200 rpm. The culture was grown until an OD at 600 nm of 10 % of the solution around 0.6 was achieved.

5 ml of each overnight culture was added to in 500 ml of AI media supplemented with 500 µl of chloramphenicol (34 mg/ml) and 4 ml of kanamycin (50 mg/ml) (2 flasks for each plasmid). The cultures were grown at 37 ^{o}C, 255 rpm for 3 or 4 hours. The cultures were then incubated at 18 ^{o}C overnight until an OD at 600 nm of 10 % of the solution at around 10 was achieved.

The 6 cultures were put in 3 of 1 L centrifuge bottles, and pelleted by centrifugation (9000 x g, 30 min, 3 ^{o}C). After removing the supernatant, each cell pellet was re-suspended in 30 ml of ice-cold PBS. The cells were transferred to 50 ml tubes, and pelleted once more by centrifugation (4000 x g, 1 hour, 4 ^{o}C). After removing the supernatant, each cell pellet was suspended in 25 ml of lysis buffer and then 3ml of lysis buffer containing 2 % lysozyme from chicken was added. The cells were shaken gently at 4 ^{o}C for 1 hour, and 100 µl from each was retained for analysis by SDS PAGE. All the lysates were frozen at -20 ^{o}C.

2.2. WT and mutant mFEN-1 protein purification:

Buffers required:

IMAC A1	20 mM tris (pH 7), 1 M NaCl, 5 mM imidazole, 0.02 % NaN$_3$.
IMAC A2	20 mM tris (pH 7), 500 mM NaCl, 40 mM imidazole, 0.02 % NaN$_3$, 0.1 % (v/v) Tween-20.
IMAC B1	500 mM NaCl, 250 mM imidazole (pH 7.2), 0.02 % NaN$_3$
Anion exchange A1	20 mM tris (pH 8), 1mM EDTA, 0.02 % NaN$_3$.
Anion exchange B1	20 mM tris (pH 8), 1 mM EDTA, 1 M NaCl, 0.02 % NaN$_3$.
Heparin A1	25 mM tris (pH 7.5), 1 mM CaCl$_2$, 0.02 % NaN$_3$.
Heparin B1	25 mM tris (pH 7.5), 1 mM CaCl$_2$, 1 M NaCl, 0.02 % NaN$_3$.
PS/A1	25 mM tris (pH 7.5), 10 % glycerol, 1 mM CaCl$_2$, 0.02 % NaN$_3$.
PS/B1	25 mM tris (pH 7.5), 2M (NH$_4$)$_2$SO$_4$, 2 mM CaCl$_2$, 0.02 % NaN$_3$.
SEC buffer	100 mM HEPES (pH 7.5), 200 mM KCl, 2 mM CaCl$_2$, 10 mM DTT, 0.04 % NaN$_3$.
Stripping buffer	20 mM Na phosphate (pH 7.4), 50 mM EDTA, 0.5 M NaCl, 0.02 % NaN$_3$.

Initially, the presence of the target protein was tested by carrying out SDS PAGE using the 100 μl samples that were taken previously from the cells lysates. Accordingly, each enzyme was purified separately in the order E160D, A159V, and then WT mFEN-1.

The cell lysate was prepared for protein purification by incubating the suspension in cold tap water until it was completely thawed. Then, the viscous suspension was sonicated on ice 3 times at 75 % power with 5 seconds bursts with at least 25 seconds between each burst until it became liquid. Then, the cell lysate was washed by adding buffer IMAC A1 containing 10 % Tween 20 to obtain a final mixture containing 1 % Tween 20. The lysate was centrifuged in 50 ml tubes for 30 minutes (30,000 X g, 4 ^{o}C) and then the supernatant was collected.

All proteins were purified at 4 ^{o}C in the cold room by using an AktaFPLC (GE Lifesciences). According to the protocol of the manufacturer, the Chelating Sepharose Fast Flow (GE Lifesciences) column (1.6 cm ID, 10 cm length) was charged with Ni^{2+} ions. The column was equilibrated with 5 column volumes (CV) of buffer IMAC A1. The previously collected supernatant (~100 ml) was then applied to the column, which was subsequently washed with 7 CV of buffer IMAC A1 and 5 CV of buffer IMAC A2. The target protein was eluted as a single fraction with 5 CV of buffer IMAC B1. 10 μl of the eluate was retained for SDS PAGE for detecting the presence of the protein. The eluted fraction was then diluted with an equal volume of a mixture of anion exchange buffers A1 and B1 (70 % of A1 and 30 % of B1).

The diluted eluate was applied directly to a Hitrap Ionic Exchange column (5 ml) using anion exchange buffers A1 and B1. Fractions containing the target protein were pooled, and 10 μl of the fraction were retained for analysis by SDS PAGE. The ionic strength of the solution was increased by the slow addition of a cold solution of 3 M of $(NH_4)_2SO_4$ with stirring to approximately 2 M final concentration.

The final solution was injected directly to a HiTrap Phenyl Sepharose HP column (20 ml) and then applied using PS/A1 buffer. The protein was eluted from the column by an inverse linear salt gradient generated using buffers PS/A1 and PS/B1 and was collected in 10 ml fractions. Protein-containing fractions were pooled, and 10 μl were retained for analysis by SDS PAGE. The pooled fractions were concentrated by ultra-filtration using a 250 ml Amicon Ultra-filtration cell with 10,000 MWCO PES membrane (Millipore) pressurised with NR R (40 PSI). The volume of pooled fractions was reduced to about 10 ml.

The concentrated fraction was injected into the super-loop using a syringe, and then applied to a HiPrep 26/10 Desalting column (1X 53 ml) using heparin buffers A1 and B1. The protein was eluted using 1CV of heparin buffers A1 (95 %) and B1 (5 %), and NaCl (5 mM).

The mixture of collected fractions was applied directly to a HiTrap Heparin HP column (3X 5 ml in tandem) using buffer heparin A1. The protein was eluted using a 50 CV linear NaCl gradient (0 to 1 M NaCl) using heparin buffers A1 and B1 and collected as 2.5 ml fractions. Protein-containing fractions were pooled and 10 μl of the elate was retained for analysis by SDS PAGE. The collected fractions were concentrated using an ultra-filtration system (Millipore) (as previously described) to collect about 10 ml. The volume of the concentrated fractions mixture was reduced to about 5 ml by using a Vivaspin (10,000 MWCO) spin columns (4000 X g, 30min, 4 ^{o}C).

The retentate from the spin columns was then applied to a Sephacryl S-100 (1.6 cm ID x 60 cm) size exclusion column (GE Lifesciences) and isocratically eluted using SEC buffer and collected in 2.5 ml fractions. Protein containing fractions were pooled, and 20 μl of the eluate was retained for analysis by SDS PAGE. The fractions were concentrated using a Vivaspin (10,000 MWCO) spin column (4000 X g, 30min, 4 ^{o}C).

2.3. Determination of protein concentration and protein folding test:

Initially, the protein concentrations were measured by absorbance at 280 nm using a Nanodrop UV spectrophotometer. The volume of the mFEN-1 sample was adjusted with the appropriate volume of 100 % glycerol (chelexed and filtered) and SEC buffer to adjust the final concentration to 100 μM protein, 50 % (v/v) glycerol, 50 mM HEPES, 100 mM KCl, 1 mM CaCl$_2$, 10 mM DTT, 0.02 % (w/v) NaN$_3$.

Protein concentrations of glycerol stocks were confirmed using a Bradford assay according to the Microtiter Plate protocol (Bradford assay).[85] Eight dilutions of protein standard (BSA) (43 mg/ml) were prepared in water with concentrations ranging from 0.05 mg/ml to 0.5 mg/ml. 1X Bradford Reagent was also prepared by dilution 5X stock in water. From the enzyme stock solution, four different enzyme concentrations were prepared in water (2.5, 5, 10, and 20 μM based on the concentration measured using the Nanodrop). Into separate microtiter plate wells, 10 μl of each protein standard and protein solution were pipetted. Then 200 μl of 1X of Bradford Reagent was added to each well, mixed, and centrifuged for few minutes to remove any bubbles. Finally, the absorbance was measured at 595 nm using a Cary Bio UV-Vis spectrophotometer (Varian). The mFEN-1 concentrations determined by comparison with the standard curve.

The secondary structure and folding properties of FEN-1 protein (WT, A159V and E160D) were analysed using ECCD technique. First, each buffer of enzyme was exchanged to another buffer contains 0.1 M potassium phosphate, 3 M ammonium sulphate, 99 mg magnesium sulphate, 500 mM EDTA and 1 M tris (hydroxypropyl) phosphine (THPP) using micro biospin6 columns-buffer exchange. Each enzyme's concentration in the new buffer was determined using a Nanodrop UV spectrophotometer. From each enzyme, a sample of 300 µl containing 6 µM of protein in the exchanged buffer was prepared and analysed in the ECCD machine. The spectrum of each protein was collected between 190 and 260 and nm.

2.4. Kinetic based assays to compare WT and mutated mFEN-1 (A159V & E160D):

Buffers and solutions were required:

10X RB (50 ml)	1 M KCl, 0.5 M HEPES (pH 7.5), 80 mM MgCl$_2$, 1 mg/ml BSA
10X FB (50 ml)	1 M KCl, 0.5 M HEPES (pH 7.5).
5X RRB (1 ml)	500 µl 10XRB, 5 µl DTT (1 M), 495 µl H$_2$O.
1X RRB (2 ml)	400 µl 5XRRB, 1600 µl H$_2$O.
MM (1.6 ml)	360 µl 5XRRB, 1240 µl H2O.
1X FB (10 ml)	9 ml H2O, 1 ml 10X FB.
ST quench (1 L)	1.5 M NaOH, 50 mM EDTA.
10X RB quench using (1 L)	1 M KCl, 0.5 M HEPES pH=7.5, 80 mM MgCl$_2$, 1mg/ml BSA, 0.02 % NaN$_3$.
1X RRB quench using (1.5 L)	150 ml 10X RB quench using, 1350 ml H$_2$O.
1X RB quench using (1 L)	110 mM KCl, 55 mM HEPES (pH 7.5), 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$_2$O.
Wave buffer B (5 L)	70 % MeCN, 1 mM EDTA, 2.5 mM tbab, H$_2$O.
Wave buffer C (5 L)	8 % MeCN, H$_2$O.
Wave buffer D (5 L)	80 % MeCN, H$_2$O.

2.4.1. Kinetic substrates design:

The kinetic based assays (multiple turnover and single turnover) were conducted using oligonucleotides to form static substrates as shown in *figure 2.2* to enhance formation of the non-productive complexes and simplify the kinetic analysis. The 5'-terminus of the single strands F1, E3, G1 and 2'OH is labelled with fluorescein fluorophore connection to be indicated by dHPLC machine. After dissolving the oligonucleotide stock in appropriate amount of water, the mass of each oligonucleotide was determined by Mass Spectrometer (MS) to be compared with the actual mass *(table 2.4)*. The concentration of the final oligonucleotide solution was determined by UV using Nanodrop at 260 nm and 20 °C using extinction coefficients generated by IDT oligo analyser 3.1 tool (https://www.idtdna.com/calc/analyzer) *(table 2.4)*.

The 50 µM of each substrate stock solution comprised 50 µM fluorescently labelled single strand DNA (F1, E3, G1 or 2'OH) and 55 µM T1 template strand in 1X FB *(figure 2.2) (table 2.3)*. Before substrate addition to any subsequent reaction mixture, it was annealed by heating to 95 $^{\circ}$C for 3 minutes and then leaving to cool at room temperature for 10 minutes to allow correct substrate folding and annealing. All the kinetic substrates), which are double flap DNA (DF), 3'-single flap DNA (3'-SF), forked gap substrate (gap) and double flap 2'OH (DF2'OH), have a junction between tow duplexes (up-stream and down-stream) *(figure 2.2)*. While the 5'-terminal of the junction is the structural difference between the DF (5'-flap), 3'-SF (no flap) and gen substrates (5'-flap with secondary structure), the DF2'OH substrate similar to DF but has RNA feature in the +1 nt possition. In addition, all these substrates had labelled 5'-terminuse with fluorescein fluorophore to indicate the product and unreacted substrate using dHPLC that equipped with a fluorescein detector. In the kinetic study, all these substrates (DF, 3'-SF, gap and DF2'OH) were used to study WTmFEN-1 activities and DF, gap and DF2'OH substrates used to study mFEN-1 mutations on the enzyme efficiency.
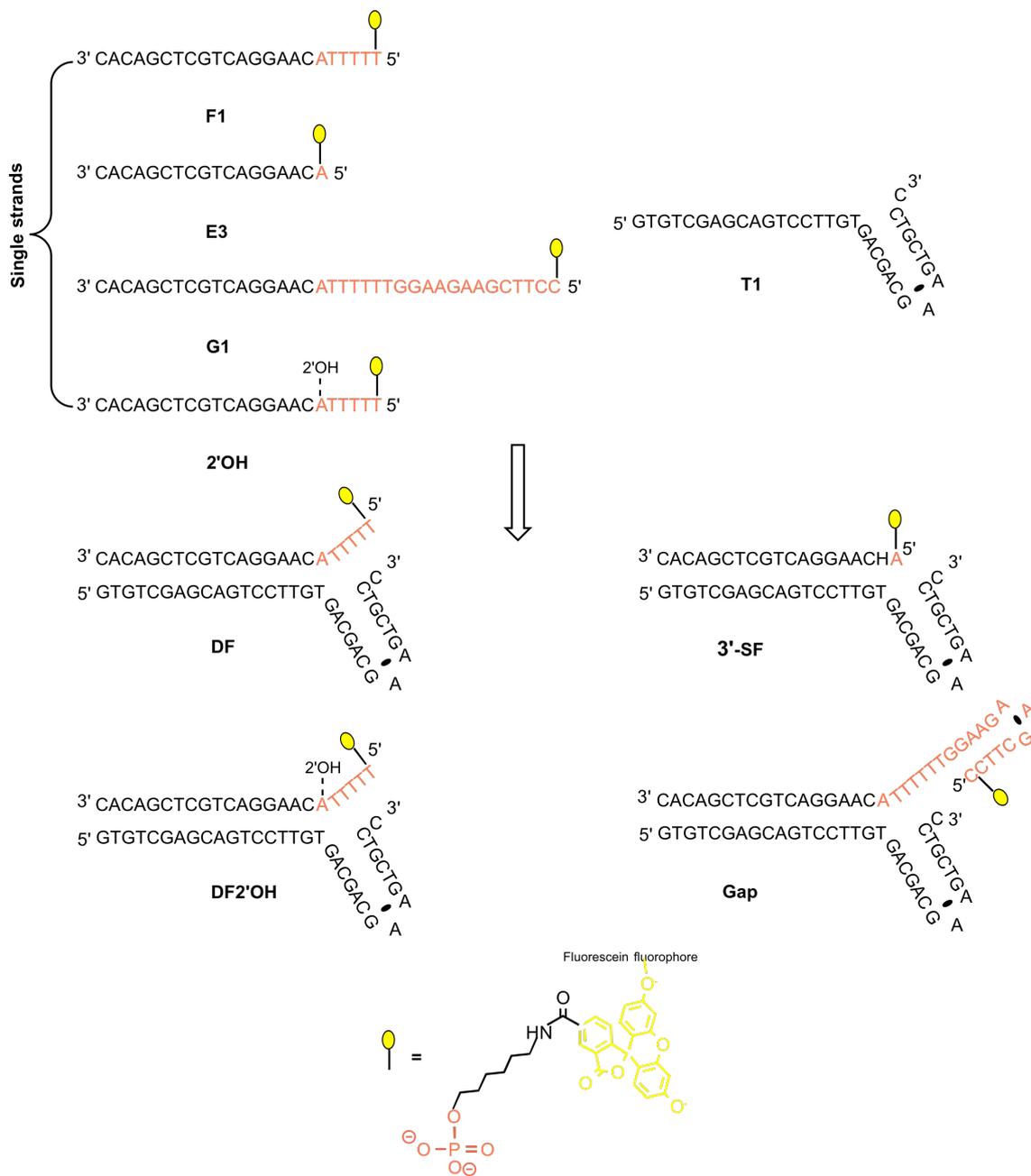
Figure 2.2. *Schematic structures of the synthetic oligonucleotides used to prepare DF, 3'-SF, bap and DF2'OH substrates that used for the kinetic study of WT and mutated mFEN-1. Each substrate structure was formed from a single strand (F1, E3, G1 or 2'OH) and a T1 template. mFEN-1 attack the substrate into **one nucleotide down stream** to release the **product** that is labelled with fluorescein fluorophore (presented as a yellow ball) connection to be indicated by dHPLC machine.*

Construct	Composition
DF	F1 + T1.
3'-SF	$E3_1$ + T1.
Gap	G1+ T1.
DF2'OH	2'OH + T2.

Table 2.3. Oligonucleotide combinations used to make the substrate constructs for MT and ST experiments to study mFEN-1 kinetic. Each substrate structure was formed from T1 and another single strand.

Oligonucleotide code	Actual Mass (g.mol^{-1})	Calculated Mass (g.mol^{-1})	ε_{260} (L.mol^{-1}.cm^{-1})
F1	7536.1	7536	236760
E3	6015.1	6013	196360
G1	11885.9	11886.12	373960
2'OH	7552.1	7552.1	237260
T1	10474.8	10475	323200

Table 2.4. Details of the kinetic oligonucleotides synthesised and purified for work with mFEN-1.

2.4.2. The multiple turnover kinetic (MT):

The multiple turnover assays for WT, E160D, and A159V mFEN-1 were conducted under similar conditions: 110 mM KCl, 55 mM HEPES (pH 7.5), 8 mM MgCl$_2$, 0.1 mg/ml BSA and 1 mM DTT. The same range of substrate concentrations was used for all enzymes. However, in order to reach the same initial rate, different concentrations of each enzyme were used at each specific substrate concentration *(table 2.5)*. While WTmFEN-1 was reacted with all substrates, the mutated enzymes were reacted with all substrates except 3'-SF substrate, which was expected to be exactly as DF according to the WT results. The MT experiments tables of WT and mutated mFEN-1 are shown in the appendices *(from table 7.1 to table 7.10).*

Each substrate stock solutions were made up from the 50 µM fluorescently labelled stock (DF, 3'-SF, gap or DF2'OH) annealed as described before. Buffer components were derived from MM solution to maintain the same final concentration between all the reaction mixtures. For each experiment an appropriate amount of 5X RRB was prepared by adding a fresh 5 mM DTT to diluted 10X RB. Then the 5X RRB was used to prepare MM and 1X RRB and kept on ice until re-

quired. All the enzyme dilutions were prepared in 1X RRB using the stocks of WT, E160D, and A159V mFEN1 enzymes, which had concentrations of 100 µM in 50 % (v/v) glycerol.

The multiple turnover assay was performed by site up 12 micro centrifuge tubes (1.5 ml) contain the substrate, MM, and 1X FB in ice. A control tube was prepared by taking 4.5 µl from the 6 first tubes and 18 µl from the other ones. The first 6 samples were incubated in a water bath (37°C) for 8 minutes. The reactions were initiated by adding the enzyme for each tube every 20 S (and mixing by pipetting the solution up and down). Then, the reactions were quenched by removing 5 µl at each time point and adding it to 250 mM of EDTA (pH 7.5). The volumes of EDTA used were determined to give an overall substrate concentration of 700 pM. The same previous steps were repeated with the other 6 samples except that the amount of the quenched reaction was 20 µl.

For each reaction, aliquots were taken at 0 (the control), 2, 4, 6, 8, and 20 minutes time points. The positive control was taken in the absence of enzyme to detect any baseline fluorescence that occurred before enzyme substrate binding. These time points were chosen so that about 10 % of product was formed in 10 minutes to allow true initial rates to be determined. The product was detected by measuring fluorescence at λ_{510} coupled to reversed-phase high pressure liquid chromatography (dHPLC). The wave buffers (A, B, C, and D) were used to separate the product from starting material on a DNAseq. While the retention time of the 5'-terminal product varied depending on its size, the remaining substrate (23 mer) always at around 8 minutes.

All of the reactions were independently repeated 3 times. The initial rates (V_o, nMmin^{-1}) were measured by plotting the amount of product against time for the first 10 % of product formed during the reaction. The measured initial rates were normalized to the concentration of enzyme. The kinetic parameters k_{cat} and K_M were determined by applying the *Michaelis and Menten Equation (equation 3.2)* by plotting the normalized initial rates versus the different initial substrate concentrations. All graphs and statistics were fitted and analysed using *GraphPad Prism 6 Software*.

2.4.3. The single turnover kinetic (ST):

A RQF-63 quench flow device (Hi-Tech Sci Ltd, Sailsbury, UK) was used to perform the single turnover assays discontinuously with a huge excess of enzyme at 37 °C. Triplicate sets of assays were carried out with small timeframes using quench flow machinery. The assays were carried out with final concentrations of enzymes and substrates (annealed as described before) equal to

10X and 1/40X the corresponding K_M that was determined previously in the multiple turnover assays *(table 2.5)*. The enzyme and substrate concentrations were initially prepared at 2X the final concentrations by using 5X RRB and 1X FB (8 mM MgCl, 110 mM KCl, 55 mM HEPES (pH 7.5), 0.1 mg/ml BSA and 1 mM DTT) and kept on ice.

| DNA | a) Concentrations in MT (PM) | | | | | | b) Concentrations in ST (nM) | | | | | |
| | [WT] | | [A159V] | | [E160D] | | WT | | A159V | | E160D | |
	From	To	From	To	From	To	[s]	[E]	[s]	[E]	[s]	[E]
DF	150	2	40000	400	1200	64	5	680	5	1820	5	2100
3'-SF	800	20	-	-	-	-	2.5	680	-	-	-	-
Gap	300	3	80000	1200	1800	100	2.5	920	7	2620	7	2820
DF2'OH	500	3	3000	50	1000	30	2	720	4	1436	8	3310

Table 2.5. The *substrate (S) and enzyme (E) concentrations that* used in MT and ST assays. **a)** *Ranges of mFEN-1 enzyme concentrations that used for each MT reaction, while the substrate concentrations always ranges from 5000 nM to 5 nM.* **b)** *The initial concentrations of S and E for each ST reaction.*

The enzyme substrate reactions were carried out using the range of time points. All the time points for each enzyme can be found in the appendices *(from table 7.11 to table 7.13)*. At each time point 100 μl of the enzyme solution was injected with an equal volume of the substrate solution. Reactions were quenched with 80 μl of the standardised quench solution (ST quench). 170 μl of each quenched reaction was placed in a small tube. All the samples were analysed using HPLC as described for the steady state measurements. The kinetic parameter $k_{ST\,max}$ was determined by plotting percentage of product formed versus time point initially by the one phase exponential *(equation 3.3)*, but they had better fitted with the two phases exponential *(equation 3.4)*. All graphs and statistics were fitted and analysed by using *GraphPad Prism 6* Software.

2.5. Observation of local conformational changes of substrate reacted duplex in the WT and mutated FEN-1 by circular dichroism of 2-aminopurine (ECCD):

Buffers and stocks were required (all were chelexed before using for ECCD assays):

10X buffer (Ca^{+2})	100 mM CaCl$_2$, 1 M KCl, 0.5 M HEPES (pH 7.5).
10X FB	1 M KCl, 0.5 M HEPES (pH 7.5).
Storage buffer (SB)	1X FB, 50% glycerol, 0.02% NaN$_3$.
DTT stock	1 M chelexed DTT solution.
Master mix (MM)	1X buffer (Ca^{+2}), 1 mM DTT.

2.5.1. ECCD substrates design using 2-AP substitution:

The aim of this ECCD experiments to probe specific structural changes of adjacent basis at the ss-ds junction of FEN-1 substrates. Accordingly, substrates containing 2-AP dimers were designed carefully from oligonucleotides to have structures as shown in *figure 2.3*. ECCD of FEN-1:substrate complexes were recorded using 3'-SF, 5'OH3'-SF or DF substrates that constructs of a 1 nt 3'-flap, 1 nt 3'-flap with 5'OH or 5 nt 5'-flap and 1 nt 3'-flap respectively, analogues to the substrates construct that used for kinetic study. 2-AP dimer was placed at +1,-1 and -1,-2 positions. The positions numbers of 2-AP dimer are relative to the scissile phosphate. All substrates had static nature construction and that was important to eliminate flap migration and allowing the ECCD spectrum to be definitely assigned to a specific position of 2-AP dimer within DNA. Each substrate was bio-molecularly formed from a template strand (T1 or T2), which contains the 1 nt 3'-flap, and a single strand with or without flap have or lack a 5'-phosphate *(figure 2.3) (table 2.6)*. Each oligonucleotide was dissolved in chelexed Milli-Q water to prepare a 500 μM stock solution. Their molecular weights confirmed using mass spectrometry, and the exact concentration of the final oligonucleotide solution was determined by UV using Nanodrop at 260 nm and 20 °C using extinction coefficients generated by IDT oligo analyser 3.1 tool (https://www.idtdna.com/calc/analyzer) *(table 2.7)*.

For the ECCD experiment, each substrate and free non-template oligonucleotide mixture was prepared as a 100 μM solution in chelexed 1X FB using the 500 μM oligonucleotide stocks. Before the substrate addition of any enzyme mixture, it was annealed by heating to 95 °C for 3 minutes and cooling to room temperature for 10 minutes to allow correct substrate folding and annealing.
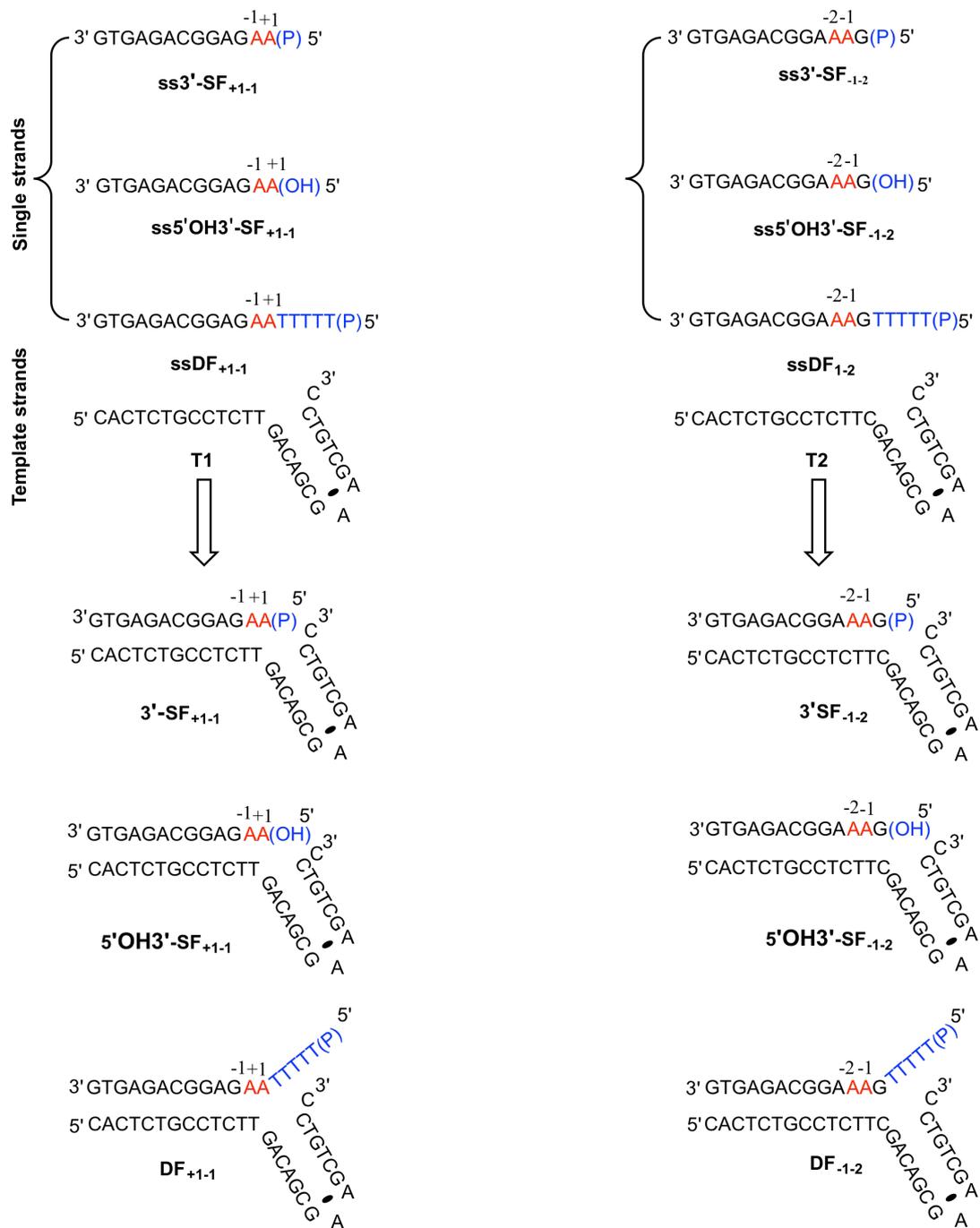
Figure 2.3. *Schematic structures of the ECCD synthetic oligonucleotides used to study the conformational changes of mFEN-1. Each substrate structure was formed from a single strand and a template.* **A** *represents 2-aminopurine nucleotide,* **(P)** *represents a 5'-monophosphate, and* **(OH)** *represents a 5'-hydroxyl. The scissile phosphate connects the +1 and -1 nucleotides.*

Construct	Composition
3'-SF$_{+1-1}$	ss3'-SF$_{+1-1}$ + T1.
3'-SF$_{-1-2}$	ss3'-SF$_{+1-1}$+T2.
5'OH3'-SF$_{+1-1}$	ss5'OH3'-SF$_{+1-1}$+ T1.
5'OH3'-SF$_{-1-2}$	ss5'OH3'-SF$_{-1-2}$+T2.
DF$_{+1-1}$	ssDF$_{+1-1}$+ T1.
DF$_{-1-2}$	ssDF$_{-1-2}$+T2.

Table 2.6. Oligonucleotide combinations used to make the substrate constructs for ECCD experiments to study the conformational changes of FEN-1. Each substrate structure was formed from T1 and a +1-1 single strand or T1 and a -1-2 single strand.

Oligonucleotide code	Actual Mass (g.mol^{-1})	Calculated Mass (g.mol^{-1})	ε_{260} (L.mol^{-1}.cm^{-1})
ss3'-SF$_{+1-1}$	4152.7	4152.71	118500
ss3'-SF$_{+1-1}$	4152.7	4151.7	118500
ss5'OH3'-SF$_{+1-1}$	4072.7	4071.74	118500
ss5'OH3'-SF$_{-1-2}$	4072.7	4073	118500
ssDF$_{+1-1}$	5593.7	5593.96	158400
ssDF$_{-1-2}$	5593.7	5593.96	158400
T1	8789.7	8789.9	257500
T2	8789.7	8793	258100

Table 2.7. Details of the ECCD oligonucleotides synthesised and purified for work with FEN-1.

2.5.2. ECCD experiment technique:

2-aminopurine (2-AP) absorbs light in a different region from bases and proteins. Accordingly, correlation of this group with the terminal nucleotides can provide a typical spectrum. This typical absorption can be affected by the presence of the 5'-terminal phosphate group and enzyme mutations.

Carefully designed substrates containing 2-AP dimers (3'-SF$_{+1-1}$, 3'-SF$_{-1-2}$, 5'OH3'-SF$_{+1-1}$, 5'OH3'-SF$_{-1-2}$, DF$_{+1-1}$ and DF$_{-1-2}$) were used to study the substrate structural changes upon WT, A159V and

58

E160D mFEN-1 binding in ±Ca^{2+} as shown in *table 2.8*. Same substrates except the DF were used to study the structural changes upon WT, Y40A, K93A, R100A, L130P, D181A and K93AR100A hFEN-1 binding in ±Ca^{2+} *as shown in table 2.8*. Samples containing 10 µM of DNA, 100 mM KCl, 1 mM DTT, 12 µM FEN-1 protein (mWT, mA159V, mE160D, hWT, hY40A, hK93A, hR100A, L130P, hD181A or hK93AR100A) and either 10 mM CaCl$_2$ or 10 mM CaCl$_2$ + 25 mM EDTA were prepared. More details about the reaction conditions can be found in the appendices *(table 7.14 and table 7.15)*. CD spectra (300-480 nm) were acquired at 20 ^{o}C using a JASCO J-810 CD spectrophotometer with hFEN-1 and chirascan plus CD spectrophotometer with mFEN-1. CD spectra were an average of 5 scans recorded in 0.5 nm steps (0.5 S response time). The base line was corrected using spectra recorded on samples containing the same components, but storage buffer rather than FEN-1 or 1X FB rather than DNA. The baseline-subtracted spectra were then smoothed using the Means-Movement option with a convolution width of 5 using JASCO Spectra analysis software (version 1.53.07) or chirascan plus analysis software depending on the used CD spectrophotometer. All of the processed data were normalised and then fitted and analysed using *GraphPad Prism 6* Software.

	E	3'-SF$_{+1-1}$	3'-SF$_{-1-2}$	5'OH3'-SF$_{+1-1}$	5'OH3'-SF$_{-1-2}$	DF$_{+1-1}$	DF$_{-1-2}$
mFEN-1	WT	✓	✓	✓	✓	✓	✓
	A159V	✓	✓	✓	✓	✓	✓
	E160D	✓	✓	✓	✓	✓	✓
hFEN-1	WT	✓	✓	✓	✓		
	Y40A	✓	✓	✓	✓		
	K93A	✓	✓	✓	✓		
	R100A	✓	✓	✓	✓		
	L130P	✓	✓	✓	✓		
	D181A		✓	✓	✓		
	K93A R100A	✓	✓	✓	✓		

Table 2.8. Details of ECCD experiments to illiterate which substrates were studied with WT and mutated mouse and human FEN-1.

2.6. Measuring substrate bending upon FEN-1 binding by Fluorescence Resonance Energy Transfer analysis (FRET):

Buffers and stocks were required (all were chelexed before using for FRET assays):

10X Ca^{+2} buffer	100 mM CaCl$_2$, 1 M KCl, 1 mg/ml BSA, 0.5 M HEPES (pH 7.5).
10X EDTA buffer	20 mM EDTA, 1 M KCl, 1 mg/ml BSA, 0.5 M HEPES (pH 7.5).
1X FB	100 µM KCl, 50 mM HEPES (pH 7.5).
DTT stock	1 M chelexed DTT solution.

2.6.1. FRET substrates design using donor and acceptor connection:

DNA constructs-oligonucleotide sequences are shown in *figure 2.4* and *table 2.9*. DNA oligonucleotides including those containing internal TAMRA and fluorescein were purchased with HPLC-purification from DNA Technology A/S. The phosphoramidite synthons used for 5'-biotin, internal TAMRA dT and internal fluorescein dT modifications were N-DMT-biotinyl-2-aminoethoxyethanol amidite, 5-DMT-T(TEG-TAMRA) and fluorescein T amidite respectively, and purchased from Biosearch Technologies Inc. After dissolving the oligonucleotide stock in appropriate amount of water, the mass of each oligonucleotide was determined by Mass Spectrometer (MS) to be compared with the actual mass *(table 2.10)*. The concentration of the final oligonucleotide solution was determined by UV using Nanodrop at 260 nm and 20 °C using extinction coefficients generated by IDT oligo analyser 3.1 tool (https://www.idtdna.com/calc/analyzer) *(table 2.10)*.

FRET substrates have labelling sites are chosen to maximise the FRET change upon enzyme binding *(figure 2.4)*. These substrates were assembled by heating 3-flap, 5-flap/exo and template strands in 1:1.1:1 ratio in 50 mM HEPES pH 7.5 and 100 mM KCL (1X FB) to 80 °C for 5 minutes and then cooling at room temperature. DNA duplex created as above with Tcdonor and template strands in a 1:1 ratio *(figure 2.4) (table 2.9)*. The DF substrate used in this study has similar structure to that used previously, however, it differs in donor and acceptor positions. In addition, substrate lacking 5'-flap with 5'-phosphate was used for first time. All substrates containing internal TAMRA and fluorescein in sites that were selected to maximise the FRET change upon bending resulted from FEN-1 binding *(figure 2.4)*. The FRET substrates (3'-SF and DF) assembled by 3'-flap (contains donor), 5'-flap/exo and template (contains acceptor) strands *(table 2.9)*. All FRET changes were compared with DNA duplex FRET that created from Tcaccept and TcdonorD *(figure 2.4 & table 2.9)*. The donor and acceptor were attached to the thymine of the 5th nt and 6th nt of 5'-terminus of each donor and acceptor obliges respectively. These substrates were

static and similar to the other substrates in this thesis except the DF substrate has 6 nts flap rather than 5 nts.

Single strands

5' GGTCC/**TAMRA**dT/ACTCTGCCTCAAGACGGTCTGCTGCACTGG 3'
Tcaccept

5' CCAG/**Fluor**dT/GCAGCAGACCGTCC 3'
Tcdonor

5' CCAG/**Fluor**dT/GCAGCAGACCGTC 3'
Tcdonor2

5' CCAG/**Fluor**dT/GCAGCAGACCGTCTTGAGGCAGAGTAGGACC 3'
TcdonorD

5' GGTCCTACTCTGCCTCAAGACGGTCTGCACTGG 3'
Tcca

5' CCAGTGCAGCAGACCGTCC 3'
Tccd

5' CCAGTGCAGCAGACCGTC 3'
Tccd2

5' /5**Bio**/TTTTTTTTGAGGCAGAGTAGGACC 3'
Tcflap

5' /**P**/TTGAGGCAGAGTAGGACC 3'
EP-Fret

5' /**HO**/TTGAGGCAGAGTAGGACC 3'
EO-Fret

5' /**P**/TTTTTTATGAGGCAGAGTAGGACC 3'
MM₊₁-Fret

DF(DAL)

3'-SF(DAL)

MM₊₁DF(DAL)

5'OH3'-SF(DAL)

5'-SF(DAL)

Nicked DNA(DAL)

3' CCAGGATGAGACGGAGTTCTGCCAGACGACGTGACC
5' GGTCCTACTCTGCCTCAAGACGGTCTGCTGCACTGG
DNA duplex(DAL)

*Figure 2.4. Schematic structures of the FRET synthetic oligonucleotides and substrates that were used to study the substrate bending upon FEN-1 binding. Each **DAL**, which is **donor** (**Fluor**) and **acceptor** (**TAMRA**) labelled substrate, structure was formed from two or three single strands. **Fluor** represents internal fluorescein, **TAMRA** represents tetramethylrhodamine, **B** represents biotin, **P** represents a 5'-monophosphate, **OH** represents a 5'-hydroxyl, and **AA** represents mismatch.*

Construct	Composition
DF (**NL**)	Tcflap + Tcca +Tccd.
DF (**DOL**)	Tcflap +Tcca +Tcdonor.
DF (**AOL**)	Tcflap + Tccd + Tcaccept.
DF (**DAL**)	Tcflap + Tcdonor + Tcaccept.
3'-SF (**NL**)	EP-Fret + Tcca +Tccd.
3'-SF (**DOL**)	EP-Fret +Tcca +Tcdonor.
3'-SF (**AOL**)	EP-Fret + Tccd + Tcaccept.
3'-SF (**DAL**)	EP-Fret + Tcdonor + Tcaccept.
5'OH3'-SF (**NL**)	EO-Fret + Tcca +Tccd.
5'OH3'-SF (**DOL**)	EO-Fret +Tcca +Tcdonor.
5'OH3'-SF (**AOL**)	EO-Fret + Tccd + Tcaccept.
5'OH3'-SF (**DAL**)	EO-Fret + Tcdonor + Tcaccept.
MM_{+1}DF (**NL**)	MM_{+1}-Fret + Tcca +Tccd.
MM_{+1}DF (**DOL**)	MM_{+1}-Fret +Tcca +Tcdonor.
MM_{+1}DF (**AOL**)	MM_{+1}-Fret + Tccd + Tcaccept.
MM_{+1}DF (**DAL**)	MM_{+1}-Fret + Tcdonor + Tcaccept.
5'-SF (**NL**)	Tcflap + Tccd2 + Tcca.
5'-SF (**DOL**)	Tcflap + Tcdonor2 + Tcca.
5'-SF (**AOL**)	Tcflap + Tccd2 + Tcaccept.
5'-SF (**DAL**)	Tcflap + Tcdonor2 + Tcaccept.
Nicked DNA (**NL**)	EP-Fret + Tcca + Tccd2.
Nicked DNA (**DOL**)	EP-Fret + Tcca + Tcdonor2.
Nicked DNA (**AOL**)	EP-Fret + Tcaccept + Tccd2.
Nicked DNA (**DAL**)	EP-Fret + Tcaccept + Tcdonor2.
DNA duplex (**DOL**)	TcdonorD +Tcca.
DNA duplex (**DAL**)	TcdonorD + Tcaccept.

Table 2.9. *Oligonucleotide combinations used to make the substrate constructs for FRET experiments. **NL** is non-labelled substrate **DOL** is **donor** only labelled substrate **AOL** is **acceptor** only labelled substrate, **DAL** is **donor** and **acceptor** labelled substrate, , and. **AA** represents 2-aminopurine nucleotides, **(P)** represents a 5'-monophosphate, and **(OH)** represents a 5'-hydroxyl. The scissile phosphate connects the +1 and -1 nucleotides.*

Oligonucleotide code	Actual Mass (g.mol^{-1})	Calculated Mass (g.mol^{-1})	ε_{260} (L.mol^{-1}.cm^{-1})
Tcaccept	11570.8	11571	352000
Tcdonor	6271.3	6271	190000.
Tcdonor2	5982.1	5982.09	182800
Tcflap	7819.3	7807	231500
Tcca	11004.1	11004	322900
Tccd	5758.8	5759	176300
Tccd2	5469.6	5468.95	169100
Ep-Fret	5668.7	5668.83	182900
EO-Fret	5588.7	5588.84	182900
MM$_{+1}$-Fret	7422.9	7422.24	237400
TcdonorD	11632.7	11632.5	364800

Table 2.10. *Details of the FRET oligonucleotides synthesised and purified for work with FEN-1.*

2.6.2. FRET experimental technique to characterise DNA bending upon FEN-1 binding:

Florescence Resonance Energy Transfer (FRET) efficiencies (E) were measured using the (ratio)$_A$ method [86] by determining the enhanced acceptor fluorescence at 37 °C. The steady state fluorescent spectra of 10 nM non-labelled (NL) trimolecular, donor-only-labelled (DOL) and douply-labelled (DAL) substrates (figure 2.4) were recorded using a Horiba Jobin Yvon FluoroMax-3® fluorometer. For direct excitation of the donor (fluorometer, DOL) or acceptor (TAMRA, AOL), the sample was excited at 490 nm or 560 nm (2 nm slit width) and the emission signal collected from 515-650 nm or 575-650 nm (5 nm slit width). Emission spectra were corrected for buffer and enzyme background signal by substracting the signal from the non-labelled (NL) DNA sample. In addition to 10 nM of the appropriate DNA construct, samples contained 10 mM CaCl$_2$ or 2 mM EDTA, 110 mM KCL, 55 mM HEPES pH 7.5, 0.1 mg/mL bovine serum albumin (BSA) and 1 mM DTT. The first measurement was taken prior to the addition of protein with subsequent readings taken on the cumulative addition of the appropriate enzyme in the same buffer *(table 2.11)*. With corrections made for dilution. Transfer efficiencies (E) were determined according to *equation 3.8*. Energy transfer efficiency (E) was fitted by non-linear regression in the Kaleidagraph program version 4.5.2 to *equation 3.9*. All experiments were repeated in triplicate.

Donor (fluorescein) was excited at 490 nm with emission sampled as the average value of the signal between 515-525 nm, and acceptor (TAMRA) was excited at 560 nm with emission average between 580-590 nm. For FRET experiments involving substrate bound to streptavidin, 5 molar equivalent of streptavidin were pre-incubated with the biotinylated substrate in buffer containing 10 mM CaCl2, 55 mM HEPES pH 7.5, 110 mM KCl, 1 mg/ml BSA and 1 mM DTT for 10 minutes at room temperature before proceeding as above.

	E	DF	MMDF	3'-SF	5'OH3'-SF	SADF	5'-SF	Nicked
mFEN-1	WT	✓	-	✓	-	-	-	-
	A159V	✓	-	-	-	-	-	-
	E160D	✓	-	-	-	-	-	-
hFEN-1	WT	✓	✓	✓	✓	✓	✓	✓
	Y40A	✓	-	-	-	-	-	-
	L53A	✓	-	-	-	-	-	-
	K93A	✓	-	-	-	-	-	-
	R100A	✓	-	-	-	-	-	-
	R104A	-	-	✓	-	-	-	-
	L130P	✓	-	-	-	-	-	-
	K132A	-	-	✓	-	-	-	-
	D181A	✓	-	-	-	-	-	-
	K93AR100A	✓	-	-	-	-	-	-
	R103AR129A	✓	-	✓	-	-	-	
	R103ER129E	✓	-	-	-	-	-	
	R104AK132A	✓	-	✓	-	-	-	
	R104EK132E	✓	-	-	-	-	-	
	Quad E	✓	-	-	-	-	-	

Table 2.11. *Details of FRET experiments to illiterate which substrates were studied with each WT and mutated mouse and human FEN-1 enzymes. The Quad E is the quadruple mutated R103ER104ER129EK132E hFEN-1*

Chapter 3: Results and Discussion of mFEN-1:

As indicated in *chapter 1*, the FEN-1 enzyme can best be described as a multifunctional enzyme with different activities to recognise different substrates. In previous studies, FEN-1 mutations have been connected to cancer because of the presence mutations reduce enzyme efficiency by affecting FEN-1 activities.[18,19,87] FEN-1 catalysis has been suggested to proceed by a specific mechanism and it is therefore possible that mutations alter specific reaction steps. In this part of project, wild-type mouse FEN-1 (WTmFEN-1) will be studied with different substrates kinetically to verify the impact on different activities and the ability to recognise the different substrates. In addition, mFEN-1 active site mutations (A159V and E160D) were studied and compared to WTmFEN-1 to illustrate if these mutations affect protein chemistry, substrate conformational change and/or substrate bending.

3.1. WT and mutant mFEN-1 expression, purification, adjustment concentration and folding test:

WT, A159V and E160D mFEN-1 enzymes expressed, purified, adjusted to 100 µM concentration and their secondary structure was tested. The obtained results are summarised as below:

3.1.1. PCR reaction, cells transformation and plasmid preparation:

Site-directed mutagenesis was used to create a mutation at a defined site of the known template sequence using PCR. As the primers were extended to replace the original sequence, the resulting amplified product incorporated the mutation. A159V, E160D and D160E primer pairs were used to produce A159V, E160D and WT mFEN-1 plasmids, respectively. The exponential amplification in PCR produced a fragment containing the desired mutation (E160D, WT and A159V mFEN-1) in sufficient quantity. This was tested by observing separation from the original (non-mutant) plasmid on an agarose gel electrophoresis stained with Sybr Safe. The target DNA for each protein was seen as a band using the ChemDoc[TM] MP Imaging System *(figure 3.1)*.

DH5α cells were transformed with pET-28b-mFEN-1 plasmids containing the WT and mutant DNA sequences. The transformed cells were spread on agar gel plates and observed as single colonies.

The starter culture for each enzyme, which was obtained by inoculation of a single colony in 2XYT media, grew successfully and cloudy solutions after overnight growth were obtained. The plasmids for each protein were collected using a mini-prep kit (Qiagen), and their concentrations were determined using a Nanodrop spectrophotometer. All of the concentrations ob-

tained were higher than 50 ng/µl (*table 3.1*). All of the prepared plasmid solutions were sent for sequencing. The sequencing data indicated that the mutagenesis had worked correctly as shown in the appendix *(section 7.1.1 to 7.1.6)*.
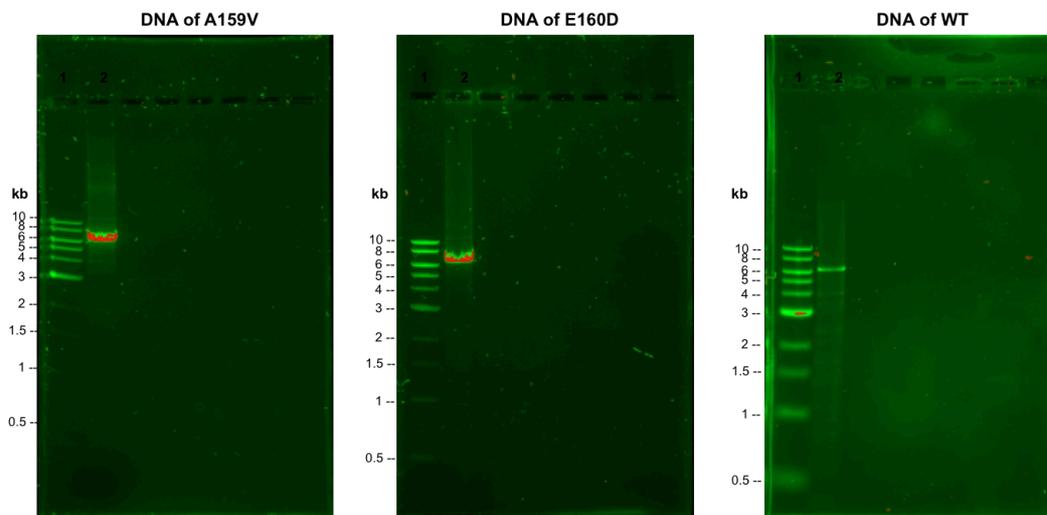
Figure 3.1. *Bands corresponding to target DNA for each enzyme on agarose gels in the second lane, and the molecular weight markers are in the first lane of each gel photo. The gels were visualised using Bio RAD chemiDoc^{TM}MP imaging system.*

mFEN-1 Plasmid	A159V	E160D	WT
The concentration obtained (ng/µl)	60	114	78

Table 3.1. *The concentrations of the obtained plasmids measured using Nanodrop Spectrophotometer.*

3.1.2. Protein expression:

BL21DE cells were transformed with the plasmids. Transformed cells were spread on agar plates and grew as single colonies. A single colony containing each plasmid was used for initial growth in 2XYT media, followed by grow-up in AI media. The cultures were pelleted by centrifugation, washed with PBS, and re-suspended into lysis buffer.

The protein synthesis was tested by running samples from the pellet and supernatant of each culture on SDS gel. A band corresponding to the correct molecular weight each target protein (WT, A159V and E160D) was seen with most of the protein in the supernatant lane *(figure 3.2)*. The bacterial cell supernatants were frozen at -20 $^{\circ}$C, and after that, they were subjected to lysis for purification application.
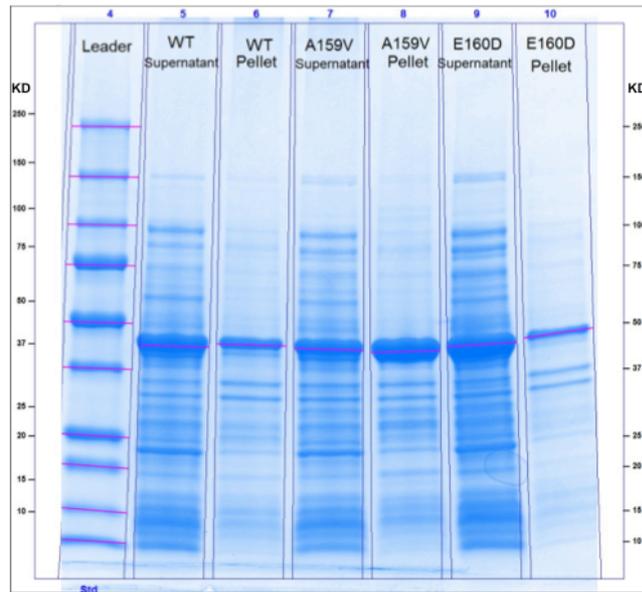
Figure 3.2. *The supernatant and pellet samples of WT, A159V, and E160D mFEN-1 overexpression analysed by SDS-PAGE. The gel was visualised using Bio RAD chemiDoc^{TM}MP imaging system.*

3.1.3. Protein purification:

Purification of all mFEN-1 proteins (WT, A159V and E160D) was performed as described in *section 2.2*. Each purification system required six chromatography media columns (IMAC, Ionic Exchange, Phenyl Sephorose, Desalting, Heparin, and Size Exclusion). The presence of the target protein, which was tested by SDS PAGE using 10 μl samples that were taken from the elution after each of the first four columns, was indicated as a band on the SDS gel. Purity of the final collected protein was indicated from the SDS gel of different concentrations of the final elution (SEC column eluate) as shown in *figure 3.3*.
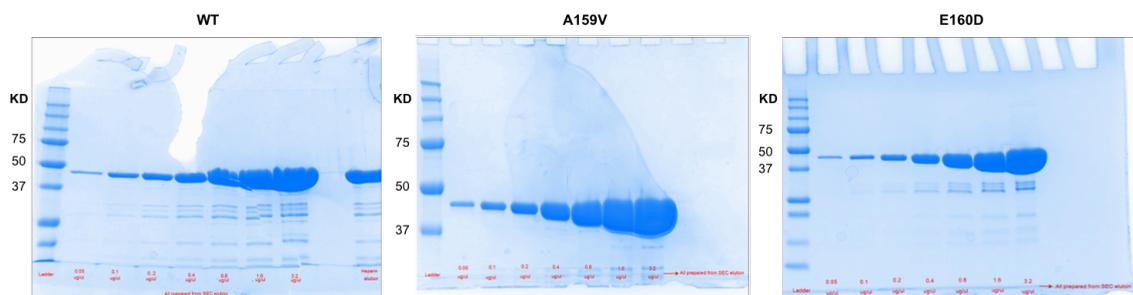
Figure 3.3. *The SDS gels of purified mFEN-1 proteins (WT, A159V and E160D). In each gel: the first lane is protein marker with molecular weights bands shown in KD, and the other lanes correspond to different protein concentrations (0.05, 0.1, 0.2, 0.4, 0.8, 1.6 and 3.2 μg/μl). The gels were visualised using Bio RAD chemiDoc^{TM}MP imaging system.*

3.1.4. Adjustment of the protein concentration and protein folding test:

By using a Nanoview spectrophotometer and the calculated extinction coefficient (mFEN-1: 22920 $M^{-1}cm^{-1}$), each protein concentration was determined by absorbance at 280 nm. The protein concentration was adjusted with the appropriate volume of 100 % glycerol (chelexed and filtered) and SEC buffer to adjust the final concentration to 100 μM FEN1, 50% (v/v) glycerol, 50 mM HEPES, 100 mM KCl, 1mM $CaCl_2$, 10 mM DTT, 0.02 % (w/v) NaN_3. Then, the proteins' concentrations were confirmed using a Bradford assay, in which the measured absorbance was compared with a standard curve *(figure 3.4.a)*.

Examining the structure of proteins in solution (WT, A159V and E160D) using CD technique obtained complementary structural information to show that the WTmFEN-1 and mutant enzymes have similar secondary structures *(figure 3.4.b)*. Therefore, loss of FEN-1 efficiency in the mutant cannot be due to an incorrectly folded structure.
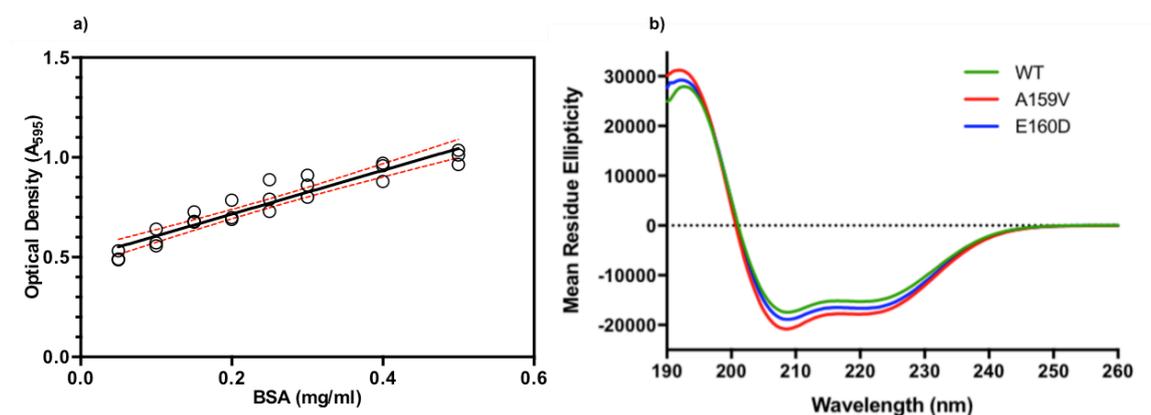
*Figure 3.4. Protein concentration and folding tests. **a)** The Bradford assay for WT, A159V and E160D. BSA standard curve used to determine the concentration of protein in the storage buffer. **b)** The protein fold test of **WT**, **A159V** and **E160D** mFEN-1 using CD machine.*

3.2. Kinetic analysis of WT and mutant mFEN-1:

Kinetic information is useful to understand how FEN-1 interacts with DNA and how the mutations can affect its activity. In fact, it is impossible to understand enzymes without understanding the kinetic properties of their reactions. In *vivo* FEN-1 specifically interacts with DNAs that have a junction between two duplexes to hydrolyse the phosphate bond that is located 1nt into downstream duplex of the 5' end of the DNA. Based on this, FEN-1 can be described as an enzyme with different activities to recognise a double flap (DF) substrate with *fen* activity, a 3'-single flap (3'-SF) substrate with *exo* activity, and gap substrate with *gen* activity *(figure 3.6.a)*.[1]

With DF substrate, the 5' end will be released as a smaller product (P) as shown in *figure 3.6.c*. The single strand product (P) is released directly after the hydrolysis chemistry. However, under multiple turnover conditions (substrate saturating [E] < [S] > [K_M]) release of the double strand product (Q) is rate limiting *(figure 3.5)*.[66] Previous studies compared the catalytic activity of WT and mutant FENs under k_{cat}/K_M conditions.[5] Since substrate-limiting conditions are likely to be diffusion-controlled reactions and substrate-saturating conditions are rate limited by product release, multiple turnover (MT) reactions may underestimate the effects of mutations.

On the other hand, determination of maximal rate of single turnover (ST) reactions can be more useful when characterising mutation, as this does not measure product release or substrate association. Accordingly, the effect of mutations that are associated with cancer on the enzyme activities was studied under MT and ST conditions. Previous studies on E160D and A159V mFEN-1 differed on how active site mutations affect the enzyme activity.[18,19] This catalysis study has two purposes. First, to ascertain whether WTmFEN-1 has different activities or just recognise different substrates by using DF, 3'SF, gap, and DF2'OH substrates. Secondly, to study the effects of A159V and E160D active site mutations on mFEN-1 activities compared with the WT using different substrates.
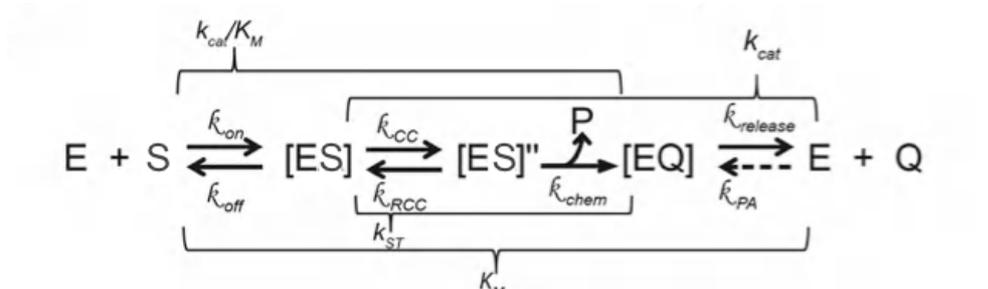
Figure 3.5. The proposed reaction mechanism of FEN-1 and DF substrate. The used relative rate constants are enzyme (E) and substrate (S) association (k_{on}), ES complex dissociation (k_{off}), reversible conformational change (k_{CC} and k_{RCC}), the chemistry step (k_{chem}) and product release step ($k_{release}$). The parameters of Michaelis-Menten (k_{cat} and K_M) and maximal single turnover (k_{ST}) presenters are presented.

For the catalysis study analogous experiments (multiple and single turnover) were conducted for WT, A159V and E160D of mFEN-1. All of these enzymes were prepared within one month to make sure they have similar ages and rule out any impacts of protein degradation with time. Kinetic assays were determined using static substrates (no flap migration) to minimise the chance of non-productive complexes and hence simplify the kinetic analysis. All the substrates are partly structurally similar, as they have two double strands of DNA up-stream and down-stream with a junction between. The 5'-end is the structural difference between all the sub-

strates. While the 3'-SF introduces a mononucleotide to FEN-1, DF and gap substrates introduce flap and flap with secondary structure respectively *(figure 3.5.a)*. All of the substrates were constructed from a template strand (T1) with specific sequence to avoid homodimer formation. T1 was annealed with labelled 5'-end strand (F1) that differs from one substrate to another. As a result, the substrates will have some differences in their 5'-flap sequence. All the substrates are labelled with a fluorescein fluorophore at the 5'-end. This allowed the products of mFEN-1 and DNA reaction to be analysed using dHPLC equipped with a fluorescence detector after separation of the fluorescent product and unreacted substrate *(figure 3.6.b.c)*.
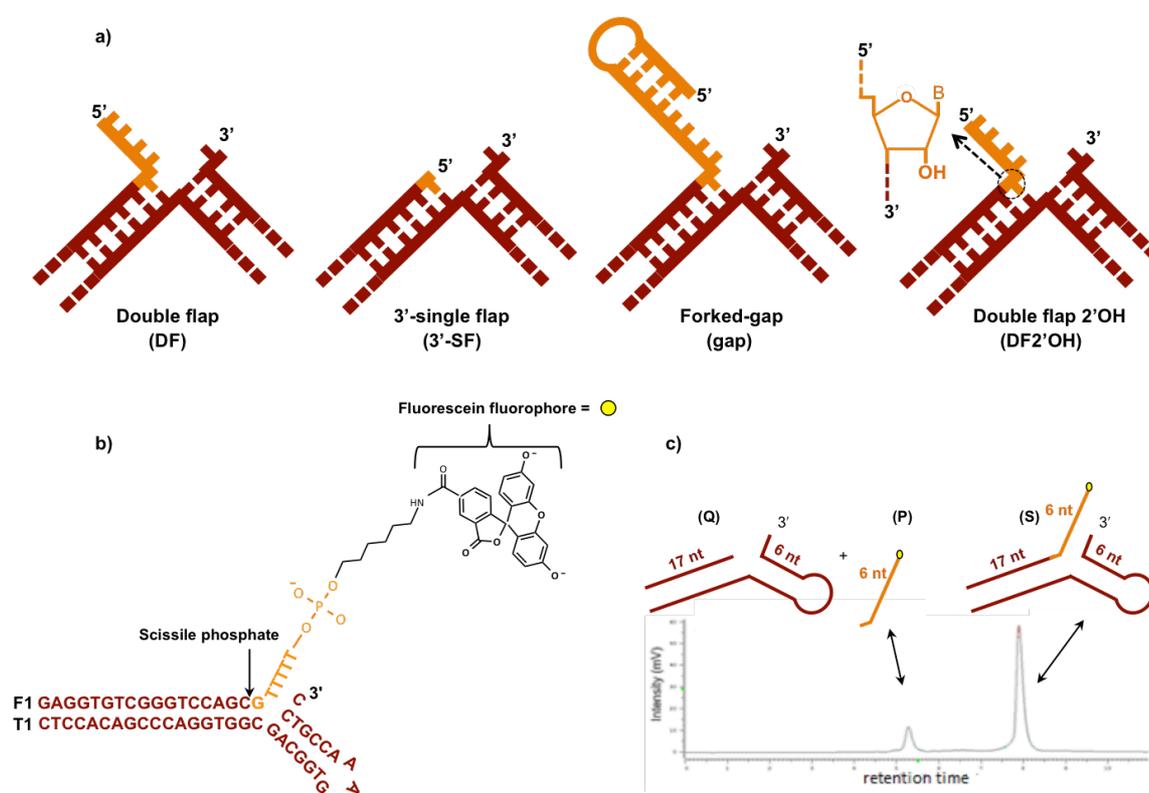
Figure 3.6. *Substrates used for kinetic analysis of mFEN-1.* ***a)*** *Cartoon structure of substrates that used for this kinetic study.* ***b)*** *The double flap substrate (DF) structure with fluorescein fluorophore connection.* ***c)*** *The HPLC trace of the substrate (S), the* ***6 nt flap product*** *(P) and the* ***dsDNA product*** *(Q), which cannot be obtained by dHPLC machine.*

Multiple turnover reactions (MT) were performed as discontinuous assays when *[E] << [S]* for a valid *Michaelis-Menten* relationship. Each reaction, which is triplicate, is measured with 12 different substrate concentrations allowing 10 % of the final product to form within the first 10 minutes of the reaction time to make sure that the measured rate is the initial rate of the product formation. Importantly, this avoids effects of the product accumulation. As the substrate concentration ranges from 5000 nM at the first point to 5 nM at the last point, enzyme concentrations varies from one substrate to another and from WT to mutated enzyme *(table 2.5)*, but

for analysis all derived rates were normalised for enzyme concentration. All reactions were quenched using 250 mM EDTA solution. The quenched reactions and control sample containing only substrate were analysed using dHPLC. Product concentration *[P]* was calculated by using *equation 3.1* where P_o and S_o are the total areas under the product peak and substrate peak respectively, and *[S]* is the substrate concentration. The initial rates of the reaction were measured when about 10 % of product had formed at 10 minutes. For an extended period ranging from 2 to 20 minutes, the product concentration increased linearly with time *(an example is shown in figure 3.7)*. The slope in the initial rate period is the initial rate of the reaction (V_o, nM.min^{-1}). To avoid the effect of substrate diminution or product accumulation, only initial rates were measured for each reaction. The initial rates were used to calculate the kinetic parameters k_{cat} *(V_{max})* and K_m by fitting the *Michaelis-Menten equation 3.2*, where V_o is the initial rate, *[S]* is the substrate concentration, and *[E]* is the total enzyme concentration. Using GraphPad Prism software, the normalized initial rates *($V_o/[E]_o$, min^{-1})* were fitted against substrate concentrations *[S] (nM)*. The multiple turnover reactions were performed in appropriate buffers containing Mg^{2+} ions at pH 7.5 *(section 2.4.2)*. For each protein, the data sets represent three or more experimental repeats.

$$[P] = \left(\frac{P_o\ area}{P_o\ area + S_o\ area} \right) \times [S] \qquad \textbf{\textit{Equation 3.1}}$$

$$\frac{V_o}{E_o} = \frac{k_{cat}[S]}{K_m + [S]} \qquad \textbf{\textit{Equation 3.2}}$$
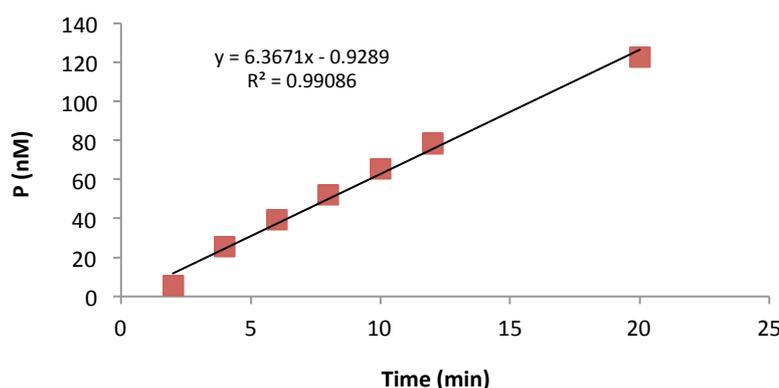
Figure 3.7. *An example of the initial rate curve for E160DmFEN-1 when [E]=480pM and [S]=750nM.*

Maximal single turnover rates (ST) were determined on WT and mutant mFEN-1s. These experiments used quench-flow machinery or manual sampling on bench for slower reactions. The conditions chosen were where *[E] >> K_M >> [S]* to reduce the chance of non-productive complexes and so that at the beginning of reaction all substrate would be in the productive inter-

mediate form *[ES]*. While DF and 3'-SF reactions were quenched using solution with 1.5 M NaOH and 50 mM EDTA concentrations, gap and DF2'-OH substrates were quenched using 80 mM EDTA solution. NaOH was not used to quench the DF2'-OH reaction to avoid transesterifcation of the substrate. The quenched reactions and control sample containing only substrate were analysed using dHPLC. As wll, the use of an EDTA quench (without NaOH) was used with gap substrate to produce better dHPLC separation of the gapped DNA product from its substrate. Although all ST data were initially fitted using a one phase exponential *(equation 3.3)*, they were better fitted using a two phase exponential *(equation 3.4)*, where P_t is the amount of product at time t, P_∞ is the amount of product at time ∞, and k_{STmax} is the single turnover rate of reaction. The single turnover rate was measured at [WTmFEN-1] 10-fold higher and *[S]* 40-fold lower than the respective K_M value for each substrate. For each protein, all data sets represented three or more experimental repeats.

$$P_t = P_\infty(1 - e^{-k_{STmax}t})$$ ***Equation 3.3***

$$P_t = P_{max1}(1 - e^{-k_{ST1}t}) + P_{max2}(1 - e^{-k_{ST2}t})$$ ***Equation 3.4***

3.2.1. WTmFEN-1 catalysis upon binding of different substrates:

Based on the previous description of FEN-1 as an enzyme with different activities, which are fen, exo and gen activities, WTmFEN-1 was reacted with different substrates *(figure 3.8.a)*. MT and ST assays were used to test WTmFEN-1 ability to hydrolyse all these 5'-ends. Curves of the MT and ST assays for the WT with DF, 3'-SF, gap and DF2'-OH substrates that were produced using *GraphPad Prism 6* software showed slight differences between the substrates in each graph *(figure 3.8)*.

The MT curves are the result of fitting of the normalized initial rates ($V_o/[E]_o$, min^{-1}) against the substrate concentrations (*nM*) of each reaction of WTmFEN-1 *(figure 3.8.b)*. Results auditing showed that interaction of WTmFEN-1 with all of the substrates (DF, 3'-SF, gap and DF2'OH) follows the classic Michaelis-Menten behaviour. In all curves, $V_o/[E]_o$ increased linearly with increase in the substrate concentration in the low range. At sufficiently high *[S]* (≈50 nM), the normalized initial rate followed saturation kinetics and tended towards a limiting value with respect to *[S]*. This limiting value, which is known as k_{cat} or the turnover number, was calculated for all of the substrates and summarized in *figure 3.8.f* and *table 3.2*. When *[S]* << K_M the slope of the curve equals k_{cat}/K_M that represents the rate constant of the second order reaction to convert free *E* and *S* to free *E* and *P*. While curves of the gap and DF2'OH substrates were very close and slightly lower than DF substrate's curve, 3'-SF substrate's curve was slightly lower than all. The residuals plot of the weighted non-linear regression plot of the multiple turnover data
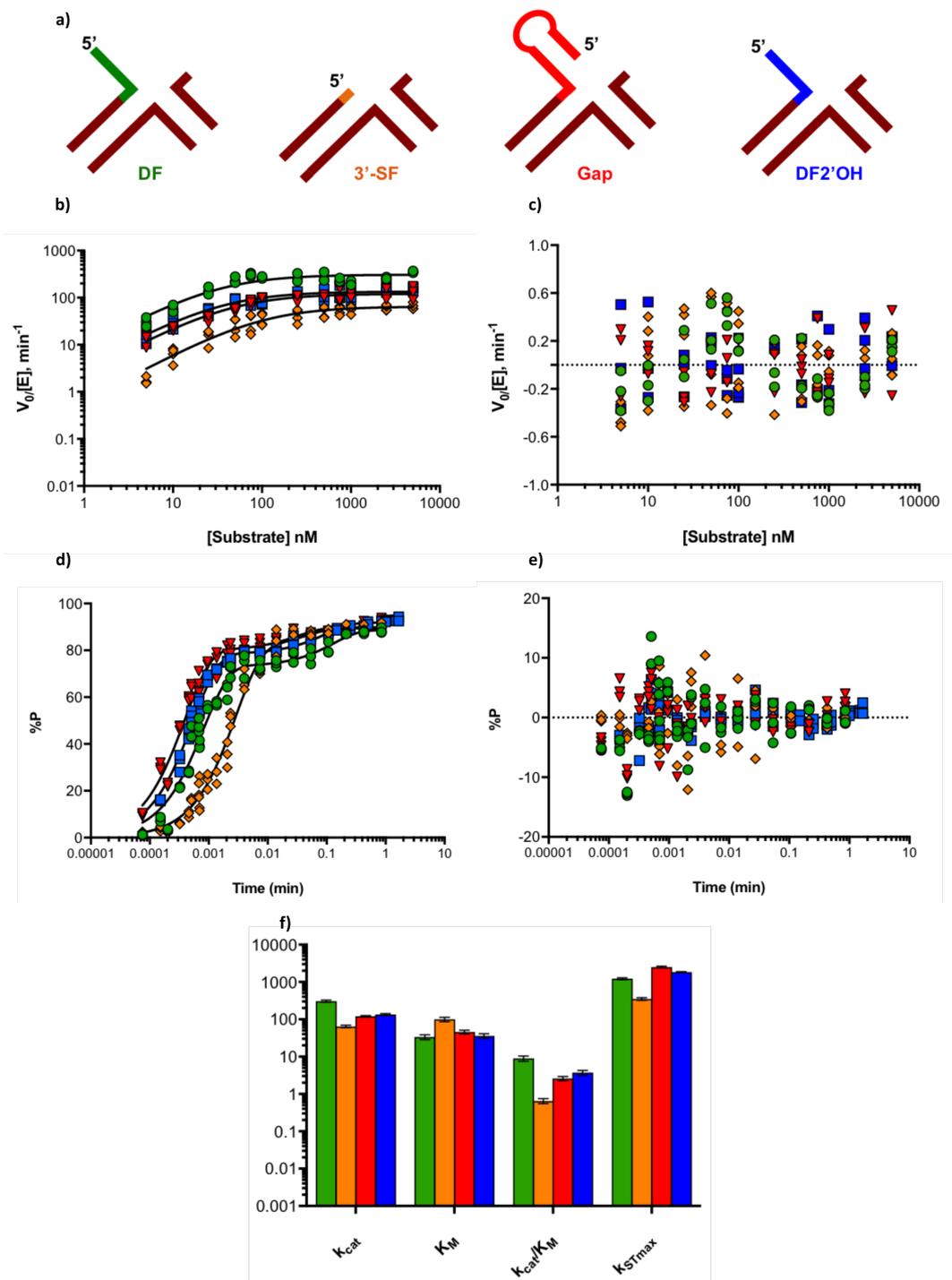
Figure 3.8. *Multiple and single turnover kinetic analysis of WTmFEN-1 upon different substrates.* ***a)*** *Cartoon of **DF**,* ***3'SF***, ***gap*** *and **DF2'OH** substrates that reacted with WTmFEN-1. The minor product is coloured differently in each substrate.* ***b)*** *Comparison of Michaelis-Menten plots for cleavage of **DF**, **3-'SF**, **gap** and **DF2'OH** substrates by WTmFEN-1.* ***c)*** *Result of non-linear regression of the MT data when variance is weighted 1/Y2.* ***d)*** *Comparison of single turnover rate profiles for **DF**, **3'SF**, **gap** and **DF2'OH** substrates by WTmFEN-1.* ***e)*** *Residual plot of the ST un-weighted data.* ***f)*** *Comparison between the obtained parameters from MT and ST kinetics of WTmFEN-1 with **DF**, **3'SF**, **gap** and **DF2'OH** interaction.*

showed disordered scatter plot *(figure 3.8.c)*. The variance of the normalized initial rate (Y-axis) changed randomly and independently on the substrate concentration (X-axis). In total the normalized initial rate variance was small and this suggests high quality of non-linear regression of the MT curve.

The ST curves result from fitting of the product percentage (% P) against time (min⁻¹) for each reaction of WTmFEN-1. Results processing showed that interaction of WTmFEN-1 with all of the substrates (DF, 3'-SF, gap and DF2'OH) follows the two-phase exponential fitting *(figure 3.8.d)*. Rate curves of all substrates appeared time dependent. While the curve of the 3'-SF substrate showed a slight difference in the first phase, all of the substrates showed significant similarity in the second phase. Curves showed that the overall product amount is linear with time reaching to around 75 % in the early phase of each reaction. The amount of product continued to increase with time and reached a final stable state in the second phase where the product was around 90 %. It can be seen that the amplitude of the first phase is bigger so most of the product forms in the first phase of each curve. The obtained single turnover rates of the reactions (k_{STmax}) were compared to MT parameters and summarized in *figure 3.8.f* and *table 3.2*. The residuals plot of the non-linear regression plot of the single turnover data showed disordered scatter plot *(figure 3.8.e)*. The normalized initial rate (Y-axis) had random variance by increase the substrate concentration (X-axis). Generally, this variance was small especially with higher *[S]*, which suggests the high quality of the non-linear regression of the ST curves.

Overall, the kinetic parameters for ST and MT of WTmFEN-1 enzyme with the different substrates (DF, gap, 3'-SF, and DF2'OH), which were summarized in *table 3.2*, were compared to confirm existing details about the enzyme interaction. In combination, they showed small differences in catalytic parameters between the various substrates. The greatest, albeit modest, impact was seen with 3'-SF where the catalytic efficiency was seen to be 9-fold lower.

Substrate	k_{cat} (min⁻¹)	K_M (nM)	k_{cat}/K_M (nM⁻¹.min⁻¹)	$k_{ST\,max}$ (min⁻¹)	$k_{ST\,max}/k_{cat}$
DF	308 ± 18	34 ± 5	9 ± 1	1229 ± 74	4
3'-SF	65 ± 5	100 ± 13	1 ± 0.1	353 ± 26	5
Gap	121 ± 6	46 ± 5	3 ± 0.3	2522 ± 120	21
DF2'OH	135 ± 8	36 ± 5	4 ± 0.5	1847 ± 45	14

Table 3.2. *The obtained MT and ST parameters of WTmFEN-1 with the different substrates.*

From the data in *figure 3.8.f* and *table 3.2* it is clear that mFEN-1 accommodates all of the substrates (DF, gap, 3'-SF, and DF2'OH) with similarity in both the affinity (K_M) and the catalytic efficiency (k_{cat}/K_M), except the 3'-SF substrate that showed slightly lower affinity (~3-fold) and lower catalytic efficiency (~9-fold) compared to DF substrate. Under saturating multiple turnover conditions *([S] >> K_M >> [E])*, WTmFEN-1 showed similar turnover-limiting rates (k_{cat} values) with all substrates, except the 3'-SF substrate that showed ~5-fold decrease compared to DF substrate. Hence, this confirms the similar biological efficiency of the enzyme towards the substrates DF, gap and DF2'OH, and the slightly lower affinity and the catalytic efficiency with the 3'-SF. Furthermore, WTmFEN-1 appears possibly subject to diffusion control (catalytic perfection – k_{cat}/K_m ~10^7-10^9 $M^{-1}.sec^{-1}$) with all of the substrates. The exact values of the measured maximal single turnover rates (k_{STmax}) of the four substrates complement the steady state parameters to identify the rate-limiting step of the WTmFEN-1. The single turnover rate (k_{STmax}) of the WTmFEN-1 was higher than the corresponding multiple-turnover rate (k_{cat}) with all of the substrates especially with gap and DF2'OH substrates. As a result, the enzyme is rate limited by dsDNA product release, because the single turnover rates of all substrates hydrolysis are faster than MT rates. Hence, this is consistent with kinetic data on T5FEN-1 and hFEN-1.[64,67] Furthermore, under single turnover conditions, only the 3'-SF substrate showed 3-fold decrease of the k_{STmax} value compared to the DF substrate. In addition, the enzyme uses the same active site and the same chemistry to interact with all of the substrates, which means that FEN-1 does not have different activities. This conclusion contradicts the earlier studies that suggested that FEN-1 is a multifunctional enzyme.[1,3,39]

3.2.2. Effects of mFEN-1 active site mutations (A159V & E160D) on substrate catalysis:

As two previous results from 2007 and 2008 about how FEN-1 mutations affect its various activities and cause cancer contradict each other, the two mutated mFEN-1 (A159V and E160D) that correspond to the previous studies were evaluated kinetically. According to WTmFEN-1 results from the previous section *(section 3.2.1)*, FEN-1 is not a multifunctional enzyme as thought previously, but it just can recognise different substrates using same chemistry. The individual active site mutations A159VmFEN-1 and E160DmFEN-1 were reacted with DF, gap, and DF2'OH to be compared with the WTmFEN1 *(figure 3.9.a)*. MT and ST assays were used to test the effect of the mutations on WTmFEN-1's ability to hydrolyse the 5'-terminus 1-nt into the downstream duplex of these substrates. Curves of the MT and ST assays for the A159V and E160D mutations with DF, gap and DF2'-OH substrates, which were produced using *GraphPad Prism 6* software, showed considerable deficiency compared to the WT enzyme *(figure 3.9)*.

The MT curves resulted from fitting of the normalized initial rates ($V_o/[E]_o$, min^{-1}) against the substrate concentrations (*nM*) for A159V and E160D with DF, gap, and DF2'-OH substrates and were combined with WTmFEN-1 results from the previous section *(figure 3.9.b,d,f)*. Results auditing showed that the interaction of the mutant mFEN-1 (A159V & E160D) with the different substrates follows the classic Michaelis-Menten behaviour similar to the WT. In all curves, $V_o/[E]_o$ increased with increase in the substrate concentration in the low range. At sufficiently high *[S]* (*≈50 nM*), the normalized initial rate followed saturation kinetics and tended towards a limiting value with respect to *[S]*. However, WTmFEN-1 always has greater response than the mutated mFEN-1s. With DF and gap substrates the rate decreased significantly from WT to E160D to A159V mFEN-1, while with the DF2'OH both of E160D and A159V appeared similar and slightly lower than WT. This variance affected the turnover number (k_{cat} value) and k_{cat}/K_M number that represents the rate constant of the second order reaction to convert free E and S to free E and P *(figure 3.9.h.i.j) (table 3.3)*. All of the residual plots of the weighted non-linear regression plot of the multiple turnover data showed disordered scatter plot *(figure 7.1.b.d.f)*. The variance of the normalized initial rate (Y-axis) changed randomly and independently on the substrate concentration (X-axis). In total, the normalized initial rate variance was small and this suggests high quality non-linear regression of the MT curves.

The ST curves resulted from fitting of the product percentage (*% P*) against time (min^{-1}) of each reaction of mFEN-1. Results processing showed that similarly to WTmFEN-1 interaction, mutant mFEN-1 also followed the two-phase exponential fitting with all of the different substrates *(figure 3.9.c.e.g)*. Rate curves of all mFEN-1 enzymes (WT, A159V, and E160D) with the three substrates (DF, gap, and DF2'-OH) appeared time dependent and *% P* reached to around 90 % in the final stable state. The obtained k_{STmax} values were compared to the MT parameters *(figure 3.9.h.i.j) (table 3.3)*. However, velocity of product formation significantly reduced from WT to E160D to A159V similarly with both DF and gap substrates, while DF2'OH showed less variation in the rate decrease. The residuals plot of the non-linear regression plot of the single turnover data showed disordered scatter plot *(figure 7.1.c.e.g)*. The *% P* (Y axis) had random variance by increase the time (X-axis). Generally, this variance was small especially with higher *% P*, which suggests the high quality of the non-linear regression of the ST curves.
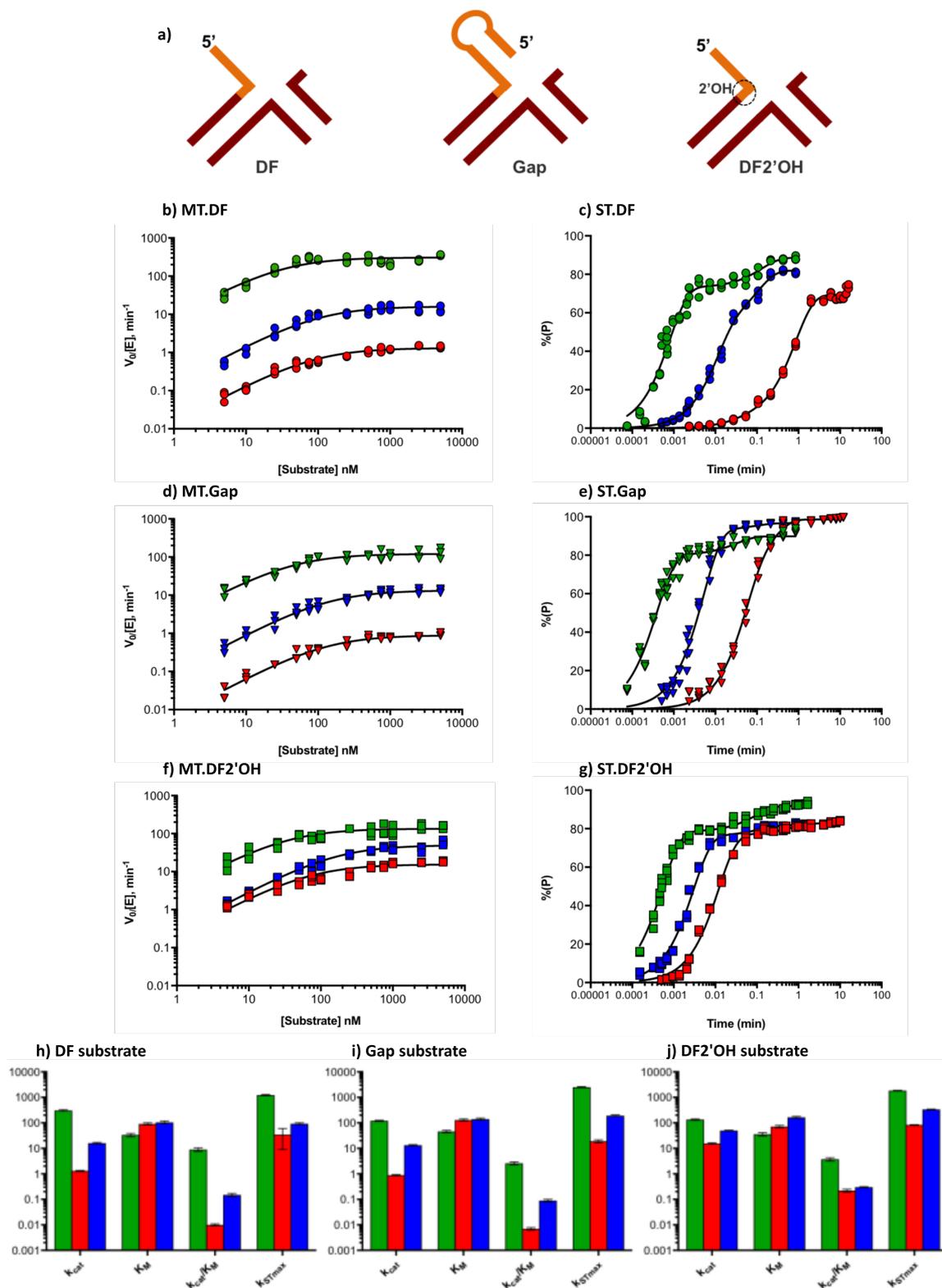
Figure 3.9. *Multiple and single turnover kinetic analysis of **WT**, **A159V**, and **E160D** mFEN-1 upon different substrates.*

*a) Cartoon of DF, gap and DF2'OH substrates that reacted with mFEN-1 to product **minor product**. b) d) and f) Comparison of Michaelis-Menten plots for cleavage of DF, gap, and DF2'OH substrates respectively by **WT**, **A159V**, and **E160D** mFEN-1. c) e) and g) Comparison of single turnover rate profiles for DF, gap and DF2'OH substrates respectively*

by WT, A159V, and E160D mFEN-1. Comparison between the obtained parameters from MT and ST kinetics of WT, A159V, and E160D mFEN-1 interaction are shown in h) DF substrate. i) Gap substrate. j) DF2'OH substrate.

S	E	k_{cat} (min^{-1})	K_M (nM)	k_{cat}/K_M (nM^{-1}.min^{-1})	$k_{ST\,max}$ (min^{-1})	$k_{ST\,max}/k_{cat}$
DF	WT	308 ± 18	34 ± 5	9 ± 1	1229 ± 74	4
	A159V	1 ± 0.06	91 ± 8	0.01 ± 0.001	34 ± 25	34
	E160D	16 ± 1	105 ± 13	0.2 ± 0.02	93 ± 7	6
Gap	WT	121 ± 6	46 ± 5	3 ± 0.3	2522 ± 120	21
	A159V	1 ± 0.05	131 ± 13	0.01 ± 0.001	19 ± 2	21
	E160D	13 ± 1	141 ± 14	0.1 ± 0.01	193 ± 12	14
DF2'OH	WT	135 ± 8	36 ± 5	4 ± 0.5	1847 ± 45	14
	A159V	15 ± 1	72 ± 8	0.2 ± 0.03	83 ± 3	5
	E160D	50 ± 2	166 ± 14	0.3 ± 0.02	336 ± 12	7

(Table 3.3). The obtained MT and ST parameters for WT and mutant mFEN-1.

Comparison all of the ST and MT kinetic parameters of A159V and E160D to WTmFEN-1 show that these active site mutations affect enzyme efficiency when acting upon DF, gap and DF2'OH substrates *(figure 3.9.h.i.j* and *table 3.3)*. Comparison of the multiple turnover numbers (k_{cat}) of A159V and E160D mutant proteins to WTmFEN-1 showed rate decrease of 300- and 20-fold with DF, 120- and 10-fold with gap, and 10- and 3-fold with DF2'OH, respectively. Surprisingly, both mutations increased the K_M values of all substrates very slightly (~3-fold) compared to the WTmFEN-1. Furthermore, the k_{cat}/K_M ratio of A159V and E160D mutant proteins showed decreases 900- and 40-fold with DF, 300- and 30-fold with gap, and 20- and 10-fold with DF2'OH compared to the WT protein, respectively. In addition, the exact values of the measured maximal single turnover rates (k_{STmax}) of A159V and E160D mutations showed decreases in their rate of reactions of 30- and 10-fold with DF, 100- and 10-fold with gap, and 20- and 5-fold with DF2'OH compared to the WT protein, respectively. According to the MT parameters it is clear that these active site mutations decrease protein turnover rates (k_{cat}) and enzyme substrate affinity (K_M values). Unlike WTmFEN-1, mutant proteins are outside the usual range for diffusion control (catalytic perfection – k_{cat}/K_M ~10^7-10^9 M^{-1}.sec^{-1}), which could means that the formation of the ES complex can be the quickest step with the mutated mFEN-1s because the mutations presence reduce the catalytic efficiency of the enzyme. The measured k_{STmax} values complement the steady state parameters to identify the rate-limiting step of mutant mFEN-1s. Although both mutations decreased the k_{STmax} values, it is still higher than the corresponding k_{cat} to confirm that the mutant enzymes also are rate limited by dsDNA product release (Q). By comparisons

between all of the kinetic parameters, A159V and E160D mutations reduce the mouse enzyme efficiency by altering the enzyme substrate affinity, diffusion control property, and rate of the final product release. Overall, the enzymatic activity of the A159V and E160D mutations reported here casts doubt on reports that this mutant had increased incidence of cancer (especially lung cancer) due only to the loss of exo and gen cleavage abilities.[18] Instead, this study supports the previous study[19] that concluded that the increased incidence of cancer was due to decreased flap cleavage capability resulting in problems with DNA replication. Interestingly, A159V mutation has greater effect than E160D mutation. Both mutations reduce mFEN-1 efficiency towards all substrates similarly, except DF2'OH substrate that reduced the effects of the mutations especially with the A159V mutation.

Although the differences observed are not well understood, one can speculate that substitution of glutamic acid with aspartic acid (E160D) or alanine with valine (A159V) affects the mFEN-1 active site structure, which means that the substrates (DF, gap, DF2'OH and expected 3'-SF) cannot be precisely positioned into FEN-1 active site to allow efficient catalysis especially with A159V mutation. Although A159V and E160D mutations do not affect the chemical environment as both still have aliphatic and acidic residues respectively, they result in longer and shorter side chains respectively. Considering the positions of the mutations, E160 has direct contact with active site metal ions, presumably the shorter side chain of E160D impairs the direct interaction with the active site metal ion and increases the spacing of the requisite divalent metal ions in catalysis. Although A159V does not have direct contact with the two metal ions active site as E160D, it has stronger effect to reduce mFEN-1 activity than E160D. This could be because A159V has bigger side chain than E160D and this can cause higher steric hindrance and affect the enzyme secondary structure that can influence substrate entrance and positioning *(figure 3.10)*. It seems unlikely that changes in the metal ion positions or steric hindrance would only affect exo and gap or only DF substrate scissile phosphate-diester placement and not affect scissile phosphate-bond hydrolysis of the other substrates, which casts doubt on previous reports.[18,19]

Decreases in the effects of mutations towards DF2'OH relate to the chemical features of the enzyme active site. The ribonucleotide of the +1 nt possesses an extra 2'OH and mimic the substrate encountered in vivo during RNA primer removal. Although, the presence of this extra 2'OH did not affect WTmFEN-1 reaction compared to DF substrate *(section 3.2.1)*, it enhances the mutant enzyme catalysed reaction, particularly with A159V mutation. Accordingly, the 2'OH produces a chemical environment that potentially facilitates reaction by decreasing the pK_a of the leaving group (pK_a of the leaving group is $\approx$ 12.98 with 2'-OH and $\approx$ 14.30 with 2'-

deoxyuridine).[88] These results suggest that the slower mutant reactions are more sensitive to this substrate change. A potential explanation of this is that the rate limiting step in the mutant is solely or partially chemical, whereas in that of the WT reaction is not. These findings have some implications for the consequences of mutation in vivo, where RNA primer removal might be expected to be more modestly impaired compared to function of FEN-1 that involve substrates that are solely composed of DNA (alpha segment editing and repair processes).
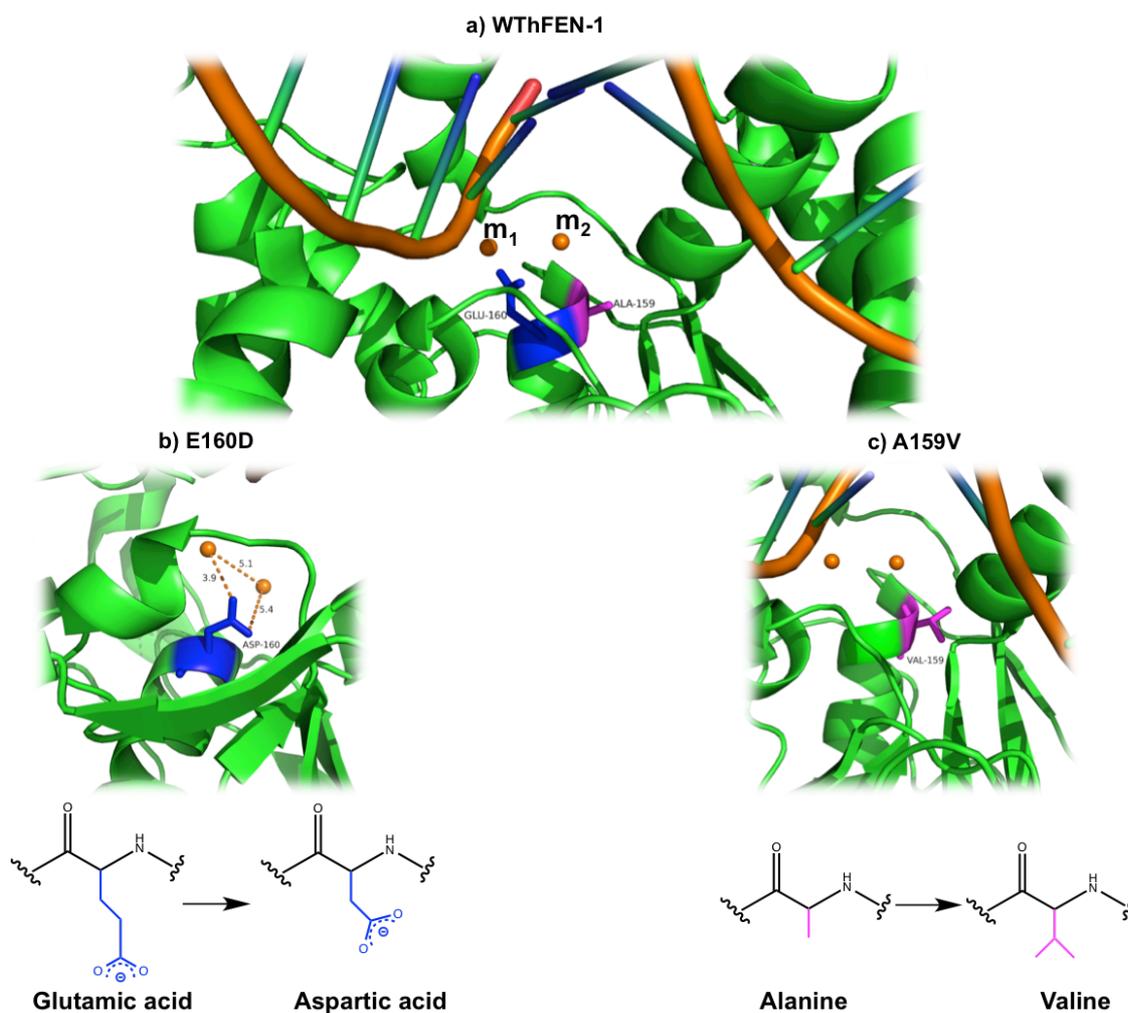
Figure 3.10. *Effects of hFEN-1 active site mutations on its structure. a) The hFEN-1 active site involved in its complex with product (PDB:3QK) to show the position of* **A159** *and* **E160** *residues.* **b)** **E160D** *mutation reduces the side chain length and increases the distance between the two metal ions of the active site.* **C)** **A159V** *mutation increases the side chain length causing higher steric hindrance that influences substrate iterance and position by effect the enzyme secondary structure comparing to* **A159** *residue.*

3.2.3. Summary and conclusion of the kinetic study of mFEN-1:

The kinetic experiments revealed that FEN-1 accommodates DF, gap and DF2'OH substrates endonucleoytically with similar biological efficiency, and 3'-SF substrate exonucleoytically with

slight lower efficiency. The enzyme performs with similar properties such as affinity, catalytic efficiency, diffusion controlled reactions and rate limited by dsDNA product release towards all the substrates, except the 3'-SF substrate that is recognised with slightly lower affinity and catalytic efficiency. In addition, mFEN-1 interacts with all of these substrates using same chemistry (phosphodiester bond hydrolysis) at the same active site, which means that FEN-1 can recognise different substrates but not different activities.

Active site mutations such as A159V and E160D reduce the enzyme efficiency by affecting the active site structure that reduces the enzyme ability to precisely position DF, gap and DF2'OH substrates in the active site to allow efficient catalysis. This is especially so with the A159V mutation. Because A159V and E160D mutations result in longer and shorter side chains respectively, presumably E160D increases the spacing of the requisite divalent metal ions in catalysis and A159V causes higher steric hindrance that influence substrate positioning. Accordingly, these active site structural changes cannot only affect on 3'SF and gap substrates or only DF substrate recognition without affecting the other substrates, which casts doubt on previous reports. DF2'OH reduces the impact of mFEN-1 mutations, particularly with the A159V mutation.

3.3. Investigation local conformational changes of substrate reacted duplex in mFEN-1 active site by low energy circular dichroism of 2-aminopurine (ECCD):

Since analysis of the impact of the mutations using crystal structures suggest that they may produce impacts on substrate positioning it was decided to study the ability of substrate to deform in response to FEN-1 and divalent metal ions. Circular dichroism (CD) of 2-aminopurine (2-AP) is used to study the local nucleic acid structural distortions. Because many of the DNA interactions implicate local conformational changes at specific junctions, only the conformational changes of the nucleotides of these junctions are of relative importance. The normal CD of any polynucleotide structure reveals the total summation of all base-base interactions. However, exciton coupling circular dichroism (ECCD) spectrum of 2-AP, more specifically two adjacent 2-APs, shows a positive CD signal at 326 nm that only reflects the interaction of the substituted residues because no natural nucleic acids or proteins absorb in this region. This method can be used to detect the conformational changes that depend on the exciton coupling between 2-AP bases of two nucleotides.[89]

Fluorescent 2-AP is readily incorporated into DNA constructs in place of adenine.[89] This substitution of 2-AP, which is an adenine structural isomer, does not significantly alter structure and stability of DNA in spite of the difference of the hydrogen bond pattern that results from repositioning of the exocyclic amino group *(figure 3.11)*.[90,91]

major groove   major groove

Adenine   Thymine   2-AP   Thymine

minor groove   minor groove
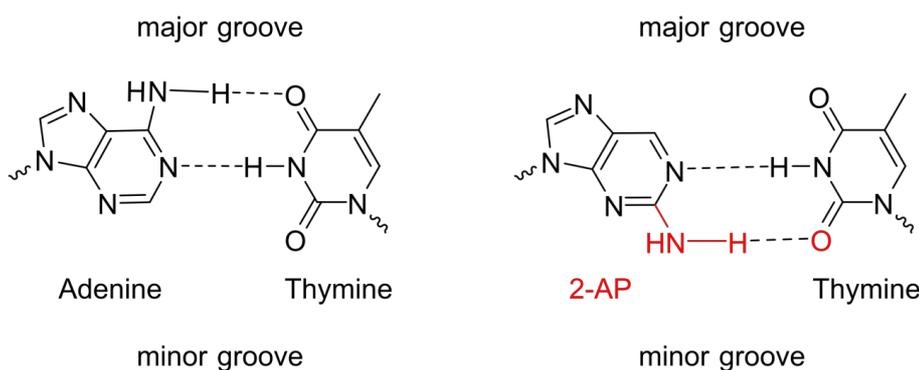
Figure 3.11. Positions of the hydrogen bonds between adenine:thymine and 2-aminopurine:thymine.[92]

Each 2-AP molecule has an electric dipole transition moment (EDTM). Substitution of two adjacent nucleotides in a DNA structure with two 2-AP molecules leads to interaction between the chromophores that result from exciton coupling of the electronic fields and splitting of the energy levels *(figure 3.12.a)*. The high sensitivity of this interaction to the magnitude, of separation and orientation of the 2-AP molecules is used to monitor the base stacking and conformational changes of the replaced dinucleotide with two 2-APs in the protein-DNA complex spectroscopically. For illustration, a non-interacting dimer of 2-APs (such as non-adjacent two 2-APs monomers) has similar one monomer absorption frequency but with doublet intensity. However, the tandem dimer $(2\text{-AP})_2$ in a DNA structure has two different energies of absorption frequency with singly excited electronic transitions *(figure 3.12.b)*.[92]
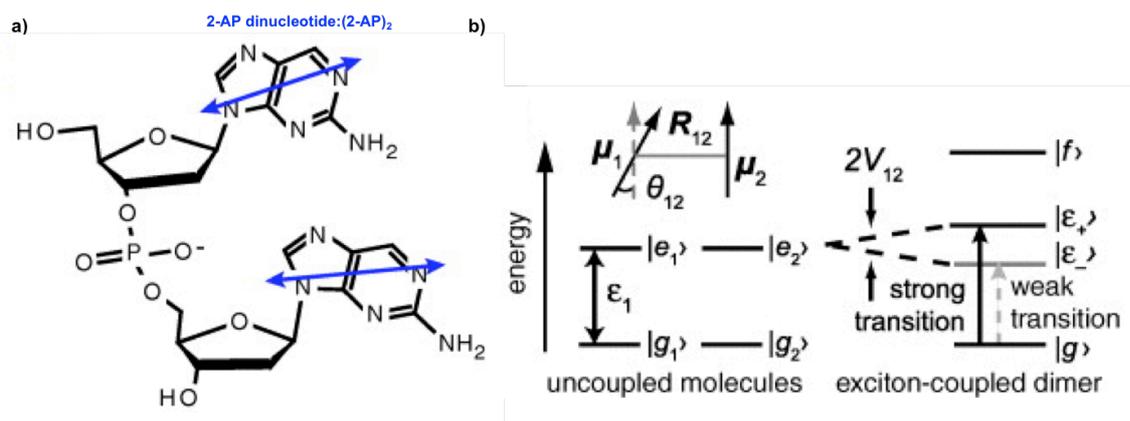
*Figure 3.12. Two adjacent 2-APs in DNA structure. **a)** The molecular structure of the 2-AP dinucleotide. **b)** Energy level diagram of two chemically identical two-level molecules.*[92]

Any two adjacent 2-APs have a significant CD signal caused by a transfer of excitation that resulted from the interaction between the fluctuations in electric field in one monomer with the other one. This electronic interaction between two adjacent 2-APs is affected by separation,

relative orientation of the 2-AP electric dipole transition moments (EDTMs), and magnitude of the excitation.[92] In addition, this ECCD spectra of two adjacent 2-AP residues is sensitive to the base stacking or unstaking in a DNA duplex.[89] The maximum stacking between two adjacent 2-APs depends on the twist angle to induce the maximum surface occlusion of the bases using the hydrophobic forces. Increasing twist angle reduces the degree of stacking (surface occlusion) between the two adjacent bases *(figure 3.13).* [92]
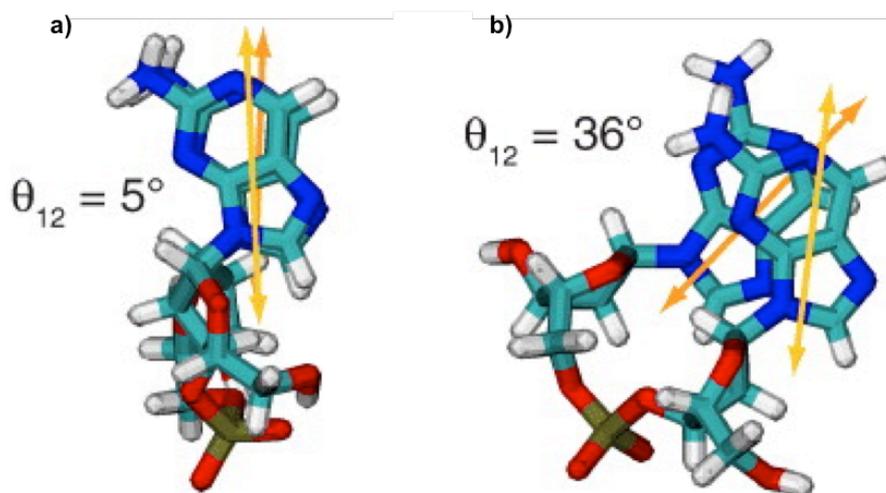
Figure 3.13. *Schematic representation of the stacking surfaces degree of two neighboring nucleic acid bases in DNA.* **a)** *The closely maximal stacking with twist angle* $\vartheta_{12}=5°$. **b)** *Much less stacking degree with twist angle* $\theta_{12}=36°$.[92]

The spectra of two adjacent two 2-APs in the context of single strand (ss) and double strand (ds) of DNA have been characterised previously. The absorption maximum of ds-DNA showed at 326-328 nm with intensity of $\Delta\varepsilon = 3.2 \pm 0.4$ (M AP)$^{-1}$.cm^{-1}. Comparatively, the ss-DNA showed 2.5-fold decrease in this intensity. Depending on the 2-AP dimer position within the DNA construct, its interaction can be used to indicate the local conformational changes of the DNA by affording base stacking and probably base pairing information.[89] In addition, the ECCD technique had been used to indicate effects of hFEN-1 mutations on the local structural changes of DNAs. Accordingly, hFEN-1 was proposed to carry out selective hydrolysis of DNA structures using a novel double nucleotide unpairing mechanism, which was thought to be required to place the scissile phosphate diester bond on active site metal ions. Support for the hypothesis that the terminal nucleotides were unpaired came from structural studies which showed that the 5'-phosphate monoester product was directly coordinated to the metal ions within the active site of hFEN-1.[37] More recent studies of FEN-1:substrate complexes where the scissile phosphodiester bond is in contact with metal ions suggest that the DNA is not unpaired, but is in-

stead untwisted introducing significant distortion to the base stacking arrangements.[93] Nevertheless, ECCD spectra reflect the change that occurs as substrate transfers to the active site upon addition of divalent metal ions.

According to the kinetic results *(section 3.2)*, A159V and E160D mFEN-1 mutations reduced the enzyme efficiency to recognise of DF, gap, 2'OH-DF and also expected 3'-SF substrates. In addition, the A159V mutation had a greater effect than E160D mutation. Thus the ECCD technique was used to further illuminate two main points. First, ECCD measurements of WT complexes with 3'-SF, 5'OH3'-SF, and DF substrates were compared to indicate importance of active site metal ion(s), 5'-flap, and 5'-phosphate of the exo substrate. These illustrated the substrate requirements for the ideal structural change. Second, ECCD measurements of mutant mFEN-1 bound to substrate were used to obtain information about the local structural changes that occur in the reacting duplex, and compared to those with the corresponding WT complex to find out any differences can explain the kinetic results.

To establish analogue 2-AP conformations in +1,-1 and -1,-2 positions in mFEN-1:DNA complexes, ECCD spectra of free ss and ds DNA and bound substrates were investigated for 3'-SF, 5'HO3'-SF and DF substrates. All substrates have a static construction to eliminate flap migration and allow the ECCD spectrum to be definitely assigned to a specific position of 2AP dimer within DNA. Each substrate was formed from two oligomers, T1 (for +1,-1) or T2 (for -1,-2) templates, which also contain the 1nt 3'-flap and the upstream duplex, and a single strand with or without flap that have or lack a 5'-phosphate *(figure 2.3) (table 2.6)*.

The experiments were performed using samples containing 10 µM DNA and when present 12.5 µM mFEN-1 protein (WT, A159V, or E160D) in Ca^{2+} and in EDTA. Like the reactions of many nucleases that use metal ion, Ca^{2+} does not support FEN-1 catalysis. However, Ca^{2+} ions have been shown to be a competitive inhibitor of FEN reactions, demonstrating they can occupy similar sites to viable cofactors like Mg^{2+}. ECCD spectra were recorded between 300–480 nm. Proteins and natural DNA bases do not absorb in this region of the CD spectrum. However, no 2-AP absorbance was observed above 360 nm with any unbound or bound DNA containing the spectral probe, $(2-AP)_2$. Therefore, only the signal between 300-360 nm is shown. Furthermore, to simplify and complement the analysis of the CD traces the ellipticity differences per $(2-AP)_2$ residue at 326 nm was calculated from three independent repetitions. The CD traces and the ellipticity difference of each complex in Ca^{2+} were compared with that of the same sample in EDTA.

3.3.1. Base-pair effects on local conformational changes of free ss and ds DNAs:

The ECCD spectra of $(2\text{-AP})_2$ inserted into single strand and double strand of 3'-SF, 5'OH3'-SF and DF substrates, which were analysed without enzyme as a basis for comparison with the subsequent experiments, were similar to those previously published of ss and ds DNAs.[37] ECCD traces of both ss and ds DNAs containing neighbouring $(2\text{-AP})_2$ in Ca^{2+} were not changed by adding EDTA *(figure 3.14.a.b)*. In addition, comparison the ECCD intensity at 326 nm of three repeats of each ss and ds DNAs in Ca^{2+} *(figure 3.14.c)* and in EDTA *(figure 3.14.d)* were also consistent with typical ECCD traces. However, ssDNA traces had weaker ECCD signals than dsDNAs. All ss traces showed small blue shifts resulted lower λ_{max} values *(figure 3.14.a.b)*.

Reviewing and comparing the free ss and ds ECCD results allows assessment of the juxtaposition of the adjacent 2-APs interaction in absence of the enzyme, which can be used later as a basis to interpret the impact of the enzyme interaction. Therefore, the impact of metal ions, 5'-flap presence, 5'-phosphate monoester, the position of the adjacent $(2\text{-AP})_2$ in the DNA and the presence or absence of basepairing could be assessed. The similar ECCD response in Ca^{2+} and EDTA suggests similar interactions between the adjacent two 2-APs with and without divalent metal ion present. Furthermore, no differences on ECCD response have been indicated to reflect any effect of adjacent 2-APs positions, 5'-flap, or 5'-phosphate monoester. On the other hand, the observed enhanced exciton coupling between the neighbouring 2-APs by transition from ss to ds indicates that the 2-AP nucleobases are restricted to fewer conformations in dsDNA compared with ssDNA. The base-pair opening reduces the interaction between two adjacent 2-APs as a result of the weaker alignment of electronic transition dipoles in the single strand DNA, and this is in good agreement with previous spectra of ssDNA having 2-APs.[89] Taken together, these results show that base-pair opening is the only influence that can increase the freedom of movement of 2-AP bases to reduce the interaction between them.
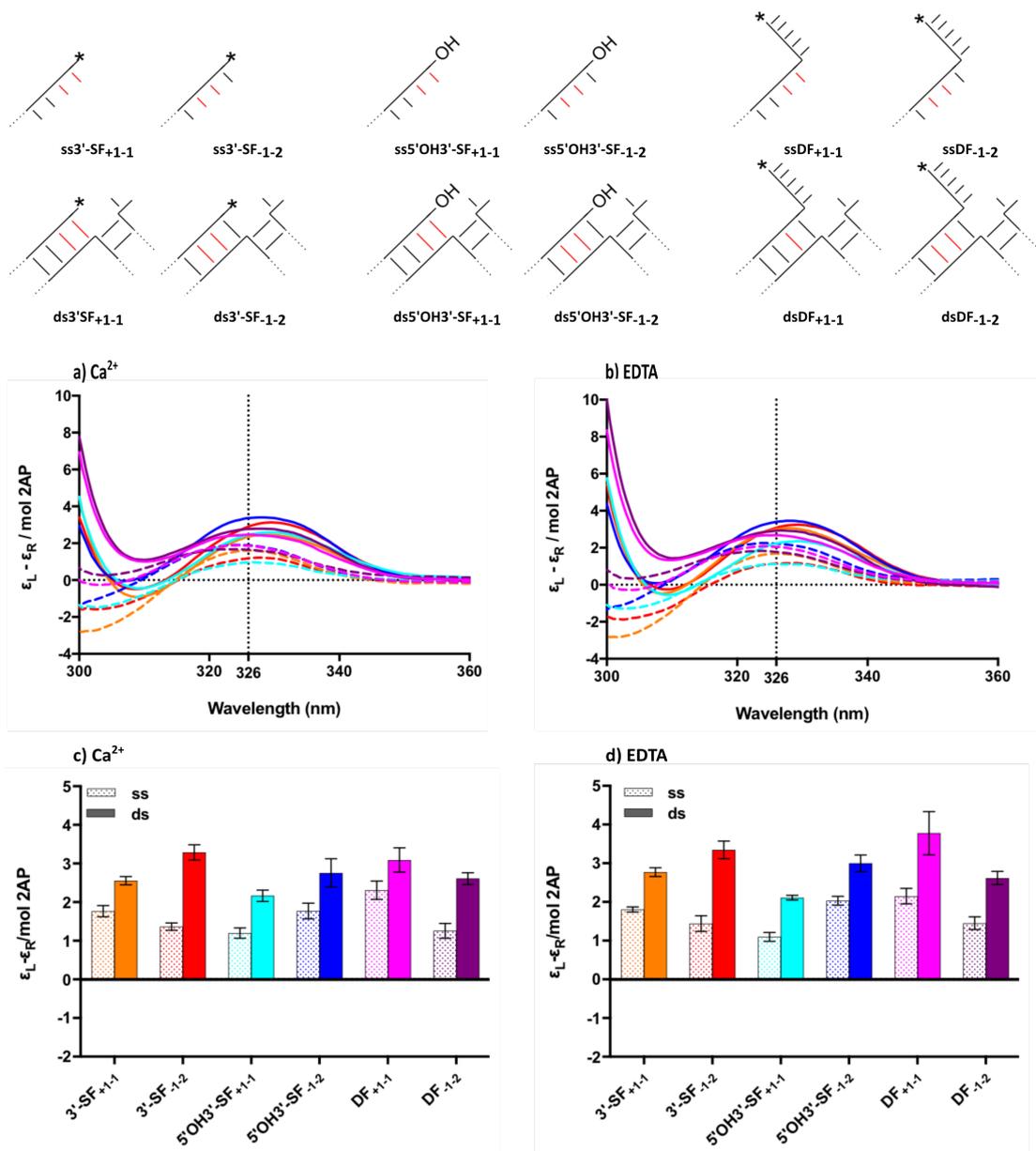
Figure 3.14. ECCD spectra of free single strands (ss) and double strands (ds) of DNAs containing 2-AP dimer. **a)** and **b)** The ECCD traces in Ca^{2+} of the unbound ds represented in solid lines, while their corresponding ss represented in dashed lines in Ca^{2+} and EDTA respectively. The colours as matched as in c and d. **c)** and **d)** Ellipticity changes at 326 nm for ss and ds in Ca^{2+} and EDTA respectively. All data sets are three experimental repeats and standard errors are shown. DNA constructs are shown schematically. 2-AP was represented in **red**, a 5'-monophosphate was represented as ***** and a 5'-hydroxyl was represented as **(OH)**.

86

3.3.2. Importance of mFEN-1 metal ion active site, the 5'-flap and the 5'-phosphate monoester for local conformational changes of the DNA reacting duplex upon WT protein binding:

The kinetic study of WTmFEN-1, which required Mg^{2+}, illustrated that the absence of a 5'-flap reduced the protein catalytic efficiency slightly (~10-fold) compared to the DF substrate *(table 3.2)*. In addition, other early reports (from our group) concentrated on the importance of the 5'-phosphate indicated that lack of the 5'-phosphate monoester of the exonucleolytic substrate decreased hFEN-1 catalytic efficiency.[5] To indicate the importance of the 5'-flap and the 5'-phosphate monoester for local conformational changes of the DNA reacting duplex upon protein binding, ECCD was used to probe the specific conformational changes of WTmFEN-1 complexes with DF, 3'-SF and 5'OH3'-SF.

WTmFEN-1 showed an analogous response with either 3'-SF or DF complexes. In Ca^{2+}, WT complexes with either 3'-SF or DF substrate within $(2\text{-AP})_2$ in +1 and -1 positions showed higher ECCD signals than corresponding complexes within $(2\text{-AP})_2$ in -1 and -2 positions *(figure.3.15 a,b,e,f)*. The calculated ECCD intensity at 326 nm identified the total differences for three repetitions *(figure 3.15.g)*. While the ECCD spectrum of WT:3'SF$_{+1-1}$ and WT:DF$_{+1-1}$ complexes in Ca^{2+} had an exciton-coupling intensity at 326 nm of $\Delta\varepsilon = 1.7 \pm 0.1$ and 1.2 ± 0.3 (M AP)$^{-1}$.cm^{-1} respectively, the lower ECCD intensity at 326 nm of both WT:3'SF$_{-1-2}$ and WT:DF$_{-1-2}$ complexes in Ca^{2+} were 0.3 ± 0.2 (M AP)$^{-1}$.cm^{-1}. The ECCD signal of the 3'-SF and DF complexes with WT were noticeably enhanced in EDTA. The signals of the 3'-SF and DF complexes with $(2\text{-AP})_2$ in +1 and -1 positions resembled the corresponding free ds3'-SF$_{+1-1}$ and dsDF$_{+1-1}$ respectively *(figure 3.15.a.e)*. However, signals of same complexes with $(2\text{-AP})_2$ in -1 and -2 positions increased to have a similar magnitude to that observed for corresponding unbound ss3'-SF$_{-1-2}$ and ssDF$_{-1-2}$ respectively *(figure 3.15.b.f)*.

However, when the 3'-single flap lacked a 5'-phosphate the resultant complexes were not able to fully reproduce the required DNA conformational change. The 5'OH3'-SF complex showed similar ECCD signal to the analogous response of both corresponding 3'-SF complex and DF complex with $(2AP)_2$ in -1 and-2 positions, however, the 5'OH3'-SF complex displayed a greater ECCD signal with $(2AP)_2$ in +1 and -1 positions. Similarly, this substrate complex in Ca^{2+} also showed a higher ECCD signal within $(2\text{-AP})_2$ in +1,-1 position than corresponding complex within $(2AP)_2$ in -1,-2 positions *(figure 3.15.c.d)*. ECCD intensity at 326 nm in Ca^{2+} showed that WT:5'OH3'-SF$_{+1-1}$ complex had similar intensity to the corresponding free ds, while WT:5'OH3'-SF$_{-1-2}$ complex had lower intensity than the corresponding ss. The intensity at 326 nm increased

dramatically by adding EDTA. While intensity of the 5'OH3'-SF$_{+1-1}$ complex exceeded the corresponding free ds, the 5'OH3'-SF$_{-1-2}$ intensity identified the corresponding free ds *(figure 3.15.g)*.

Consideration of all of the WT protein complexes with 3'-SF, 5'OH3'-SF or DF substrates allows identification of the importance of the two metal ion active site, the 5'-flap, and the 5'-phosphate for the local conformational changes of the DNA reacting duplex. First, in accord with what is already known about the importance of FEN-1 metal ion active site, increase of the ECCD intensity at 326 nm of all WT:DNA complexes upon adding EDTA suggests that the presence of the metal ion(s) in the active site (Ca^{2+}) is essential for the local conformational changes of all substrates. This increase indicates the conformational changes with Ca^{2+} result from a decrease in the stacked population of +1 and -1 nucleotides also -1 and -2 nucleotides. Together the data imply the divalent metal ion dependence for local conformational changes. Second, the similar behaviour of WT complexes with either 3'-SF or DF substrates in Ca^{2+} and EDTA demonstrates that lack of the 5'-flap does not affect the local conformational changes of the reacted duplex. This means, the 5'-flap is not required for ideal local conformational changes. This behaviour is consistent with what was observed previously from a DNA conformational change study with DF$_{-1-2}$ and P$_{-1-2}$.[37] In addition, this result is in agreement with the slight decrease of the catalytic efficiency of mFEN-1 with 3'-SF substrate compared to DF substrate that observed kinetically *(table 3.2)*. Third, the different ECCD response of 5'OH3'-SF$_{+1,1}$ substrate from both 3'-SF$_{+1,1}$ substrate, which has a terminal 5'-phosphate, and DF$_{+1,1}$ substrate, which has a 5'-phosphate diester in the corresponding position indicates the importance of the 5'-phosphate monoester of the +1 nt to produce the ideal local conformational change of +1 and -1 nts upon mFEN-1 binding. The higher ECCD response of the substrate complex (WT:5'OH3'-SF$_{+1,1}$) in Ca^{2+}, which is further increased when EDTA was added, suggests that lacking of the 5'-phosphate monoester causes only imperfect conformational change of the +1 nt leading to reduced electronic interaction between the two nucleotides +1 and -1 nts, but not as required. However, lack of this 5'-phosphate does not effect the conformational change of -1 and -2 nts as their electronic interaction reduced ideally as 3'-SF and DF substrates. This provides an explanation for what was previously reported about 5'OH3'-SF (that lacks a 5'-phosphate monoester) where large decreases of hFEN-1 catalytic efficiency were observed.[5]
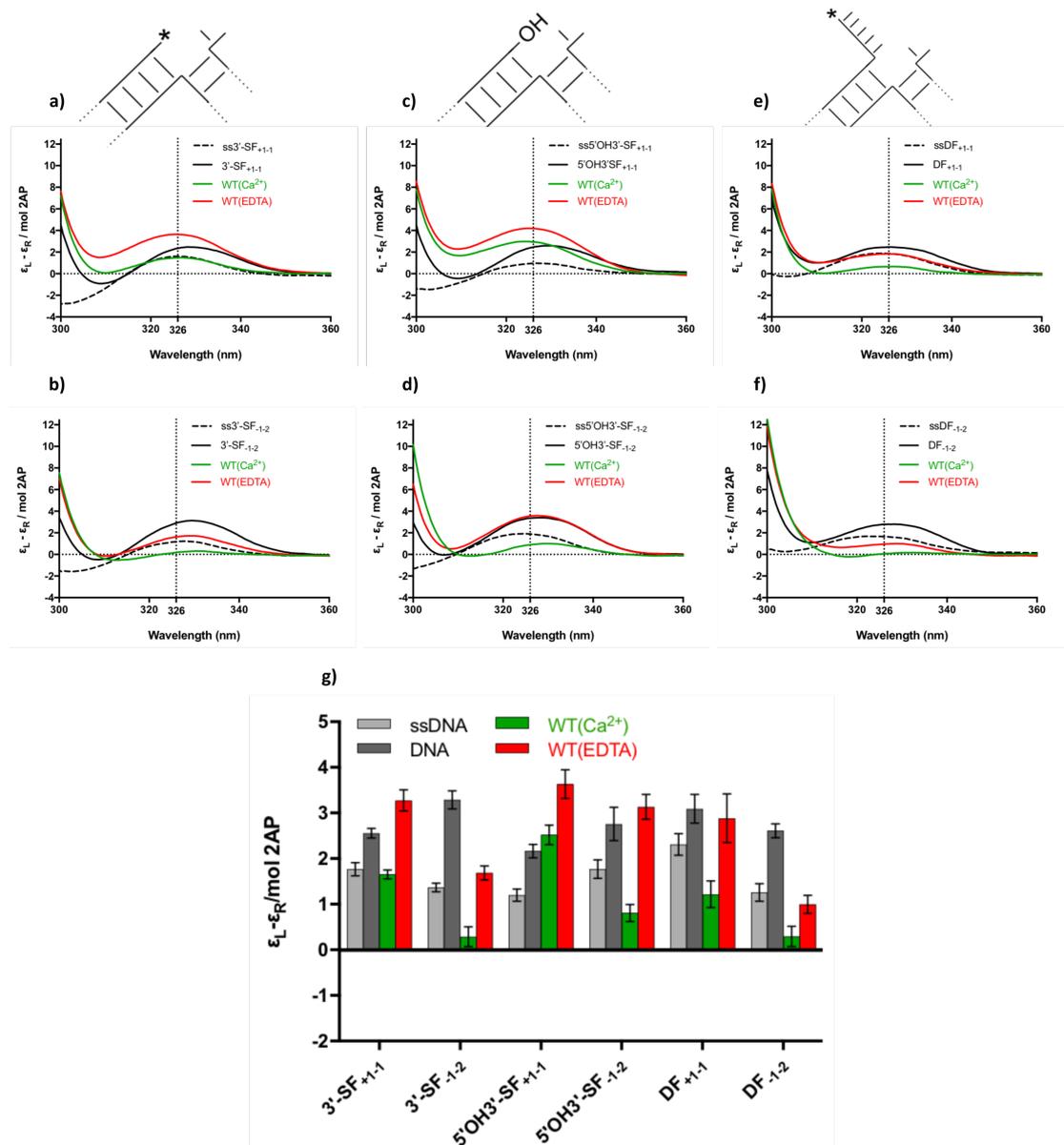
Figure 3.15. *ECCD spectra of WTmFEN-1 complexes in* **Ca²⁺** *and* **EDTA** *besides the corresponding* ***ss*** *and* ***ds*** *of each substrate.* **a)** *WT:*3'-SF$_{+1-1}$ *complex.* **b)** *WT:3'-SF$_{-1-2}$ complex.* **c)** *WT:5'OH3'-SF$_{+1-1}$ complex.* **d)** *WT:5'OH3'-SF$_{-1-2}$ complex.* **e)** *WT:DF$_{+1-1}$ complex.* **f)** *WT:DF$_{-1-2}$ complex.* **g)** *Ellipticity changes at 326 nm for WT complexes and their corresponding ss and ds-DNAs containing 2-AP dimer. All data sets are three experimental repeats and standard errors are shown. DNA constructs are shown schematically with a 5'-monophosphate was represented as * and a 5'-hydroxyl was represented as* **(OH)**.

Clearly, mFEN-1 requires its metal ion active site and a substrate where the +1 nucleotide is 5'-phosphorylated to produce ideal conformational changes to position the scissile phosphate for hydrolysis. Additionally, lack of a 5'-flap does not affect the conformational changes. Based on these results, it is possible to represent the similar conformational changes of both 3'-SF and DF substrates as the typical conformational changes that are required for mFEN-1 catalytic efficiency. Following the changes of ECCD traces of both substrates suggests that ideal conformational

89

changes reduce the degree of the electronic interaction between +1 and -1 nts also between -1 and -2 nts. In fact, this reduced exciton-coupling may result from different origins such as change of relative orientation of the +1 and -1 bases with different angles and/or distance the +1 and -1 nucleotides (unstacking) *(figure 3.16)*. Furthermore, the higher ECCD response of (2-AP)$_2$ in +1,-1 position than -1,-2 positions shows that the electronic interaction between -1 and -2 nucleotides are reduced further than +1 and -1 nucleotides. However, these results alone are not enough to accept or refute the previous suggestion about the DNU mechanism that was formulated based upon on the obtained unpaired -1 nt in product complex.[1,37] This mechanism suggested that this local DNA conformational change was restricted to being double nucleotide unpairing and may be combined with reorientation of +1 and -1 nts to position the phosphate bond into the active site.[1,2,37] However, combined hFEN-1 ECCD results, which will be discussed in *section 4*, with later results indicated that the scissile phosphate bond was positioned into the active site with fully base-paired DNA. These data suggested that Y40 and positive side chains of the helical arch untwist the dsDNA. This reorientation alters the base overlap and the base pairing strength of +1 and -1 nts and affect the ECCD signals.[93] Taken together, the DNU and untwisting mechanisms leads us to rethink the total mechanical events. It is possible to imagine that an interaction force between the enzyme and the DNA leads to untwisted DNA with full base pairing (but nucleobase unstacking) to position the target phosphate bond into the active site. Completion of the hydrolysis chemistry causes disappearance of the +1nt forcing structural changes that may cause -1 nt unpairing. Accordingly, the conformational changes are considered as reorientation of +1 and -1 nucleotides.
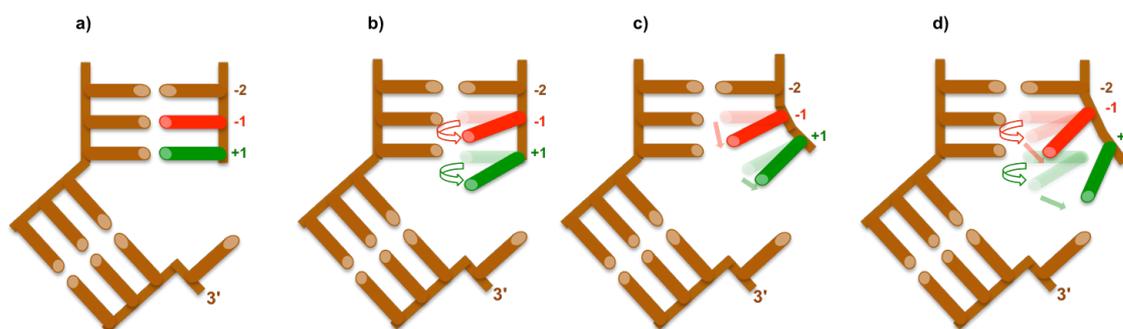
Figure 3.16. Cartoon to illustrate different DNA conformational changes that lead to changes in the electronic interaction between +1 and -1 nts and between -1 and -2 nts. a) Normal DNA. b) DNA with reoriented +1 and -1 nts with different angles without effect on the base pairing and stacking of both nts. c) DNA with distanced +1 and -1 nts causing unstacking and maybe DNU. d) DNA with reoriented and distanced +1 and -1 nts causing unstacking and DNU.

3.3.3. Effects of mFEN-1 mutations (A159V & E160D) on local conformational changes of reacting duplex upon their complexes with DF, 3'-SF or 5'OH3'-SF compared with the WT:

The effects of A159V and E160D mutations on the local conformational changes of exonucleolytic and DF substrates were investigated by comparison of their complexes with the corresponding WT complex to indicate any differences. Furthermore, both mutations were compared to each other to investigate if they differ in their effects. Overall, both mutations affected the ECCD response of 3'-SF, 5'OH3'-SF and DF complexes.

Compared with WT:3'-SF complex, A159V and E160D mutations affected the ECCD traces of their complexes with 3'-SF substrate. In Ca^{2+}, both mutant mFEN-1 complexes with 3'-SF$_{+1-1}$ substrate increased the ECCD signal analogously further than the comparable WT complex and the identical free substrate *(figure 3.17.a.)*. By adding EDTA, the calculated ECCD intensity at 326 nm in Ca^{2+} increased slightly to approach the WT signal *(figure 3.17.g.h)*. A159V and E160D mFEN-1 complexes with 3'-SF$_{-1-2}$ showed a diverse response that also differed from WT complex in the presence of the divalent metal ion. While A159V caused a deeper minimum than WT, E160D led to higher signal than WT at 326 nm *(figure 3.17.b.g)*. ECCD intensity at 326 nm of these 3'-SF$_{-1-2}$ mutant complexes increased to approach the corresponding WT complex and the single strand signals (ss3'-SF$_{-1-2}$) upon adding EDTA *(figure 3.17.h)*.

Similarly to mutant mFEN-1:3'-SF complexes, more variations also were detected with mutant enzyme:5'OH3'-SF complexes compared to the corresponding WT complex. Both mutations affected the ECCD signals of their complexes with 5'OH3'-SF similarly to their complexes with 3'-SF. The ECCD signal associated with 5'OH3'-SF$_{+1-1}$ complex with either A159V or E160D increased analogously, but further than WT complex and the identical free substrate in presence of Ca^{2+} *(figure 3.17.c.)*. Both of these complexes produced only a very slight increase in the ECCD intensity at 326 nm by adding EDTA to approach the corresponding WT complex signal *(figure 3.17.g.h)*. In addition, the signals that associated with A159V and E160D complexes of a (2-AP)$_2$ in -1 and -2 positions were lower and higher than WT:5'OH3'-SF$_{-1-2}$ complex in Ca^{2+} respectively *(figure 3.17.d)*. The calculated intensity at 326 nm increased slightly to similar extent of WT signal in EDTA *(figure 3.17.g.h)*.

Less marked effects of mutations on ECCD response were observed with complexes involving DF substrate. The ECCD traces of the WT:DF complexes were altered with the presence of either A159V or E160D mutation, but not with the same impact as seen with exonucleolytic complexes. Compared to WT:DF$_{+1-1}$ complex, E160D did not alter the ECCD intensity at 326 nm in Ca^{2+}

and EDTA significantly, however, A159V mutation increased the intensity in Ca^{2+} slightly, which was increased more by adding EDTA to approach the WT and E160D intensity on EDTA *(figure 3.17.e.g.h)*. With DF_{-1-2} substrate, both mutations reduced the ECCD signal, but more significantly with A159V than E160D *(figure 3.17.f)*. By adding EDTA, intensity of both mutations complexes increased to reach intensity of the corresponding WT complex *(figure 3.17.g.h)*.
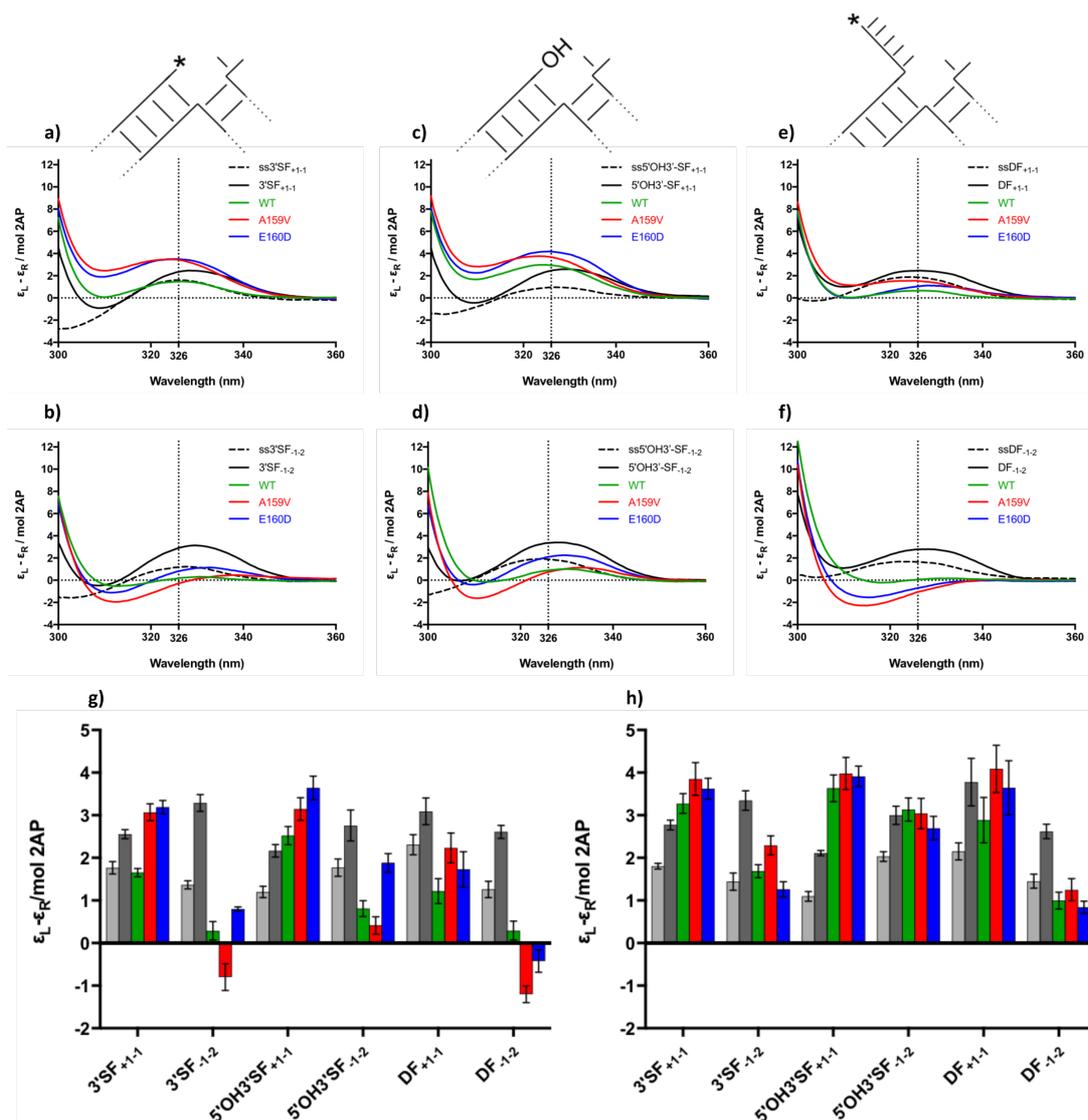
Figure 3.17. *ECCD spectra of **WT**, **A159V** and **E160D** mFEN-1 complexes with: **a)** 3'-SF$_{+1-1}$. **b)** WT:3'-SF$_{-1-2}$. **c)** 5'OH3'-SF$_{+1-1}$. **d)** 5'OH3'-SF$_{-1-2}$. **e)** DF$_{+1-1}$. **f)** DF$_{-1-2}$. Besides that the corresponding **ss** and **ds** of each substrate also represented. Ellipticity changes at 326 nm for **WT**, **A159V** and **E160D** mFEN-1 complexes and their corresponding ss and ds-DNAs containing 2AP dimer are shown **g)** in Ca^{2+} and **h)** in EDTA. All data sets are three experimental repeats and standard errors are shown. DNA constructs are shown schematically with a 5'-monophosphate was represented as * and a 5'-hydroxyl was represented as **(OH)**.*

According to the results of mutant mFEN-1 complexes with either exonucleolytic or DF substrate, some differences from the corresponding WTmFEN-1 complexes were observed. This suggests unfavourable effects of A159V and E160D mutations on the local conformational changes of reacting duplex of each substrate, reflecting positioning of the scissile phosphate-bond in the active site.

First, with all substrates (3'-SF, 5'OH3'-SF and DF), A159V mutation similarly affects mFEN-1 ability to produce the ideal local conformational changes. Compared to the ECCD response of the analogous WT complex, A159V effects on ECCD response may reflect differences in a partition between oriented and not oriented and/or imperfect orientation of the 2-AP nucleobases in +1 and -1 positions. However, the contrary effect of mutation on the interaction of $(2\text{-AP})_2$ in +1,-1 positions, which had higher response than corresponding WT complex, and -1,-2 positions, which had lower response than corresponding WT complex, may support the possibility of the imperfect reorientation of +1 and -1 nts. This imperfect reorientation decreases the electronic interaction between +1 and -1 also between -1 and -2, but higher and lower than required respectively. Second, the E160D mutation affects local conformations of 3'-SF and 5'OH3'-SF substrates analogously, whereas the effect on DF is slightly different. The higher ECCD response of E160D complexes with either 3'-SF or 5'OH3'-SF, compared to the corresponding WT complex, may reflect increase a partition of the not oriented forms and/or imperfect orientation of the 2-AP nucleobases. The DF complexes that appears equal and lower signals with $(2\text{-AP})_2$ in +1,-1 and -1,-2 positions respectively lead to suggest that the +1 and -1 nucleotides are oriented imperfectly. As a result, the electronic interaction between +1 and -1 nucleotides is decreased from the ideal seen with the WT complex; however, the electronic interaction between -1 and -2 nucleotides is decreased further than required.

Taken together, both mutations alter the conformational changes of the WT complexes of three of the substrates (3'-SF, 5'OH3'-SF and DF), however, they differ slightly in their effects *(table 3.4)*. A159VmFEN-1 twists all the substrates similarly and decreases the electronic interaction between +1 and -1 also between -1 and -2 less and more than corresponding WT complex respectively. E160DmFEN-1 twists both exo substrates similarly decreasing the electronic interaction between the three nucleotides, but still less than the corresponding WT complex. With DF substrate, E160D mutation only alters the -1 and -2 nts conformational changes by decreasing their electronic interaction more than the corresponding WT complex. As a result of these imperfect conformational changes, both mutations do not have ability to position the scissile phosphate bond between the two metal ions active site to the same extent as the WT protein. Howver, the effects are much more marked in the case of A159V, which is in line with the

greater impact of A159V on the rate of reaction of DF substrate (also expected of 3'-SF sub-strate) than E160D, which has been demonstrated kinetically in *section 3.2.1.* In addition, the lower efficiency of WThFEN-1 to cleave 5'OH3'-SF, which was shown previously,[5] is expected to be worst in presence of these mutations. As was expected in the kinetic study, these altered conformational changes may result from the higher steric hindrance of A159V and the direct contact between E160D and metal ion active site, which can influence substrate positioning and the distance between divalent ions respectively.

Substrate	A159V		E160D	
	+1,-1	-1,-2	+1,-1	-1,-2
3'-SF	+	-	+	+
5'OH3'-SF	+	-	+	+
DF	+	-	=	-

Table 3.4. Strength of the electronic interaction between the two adjacent 2-Aps of +1,-1 and -1,-2 positions of the mutated mFEN-1 complexes comparing to the corresponding WT complex of the three studied substrates. The used symbols mean that (+) higher interaction (-) less interaction (=) similar interaction to the corresponding WT complex. The data were summarised from figure 3.18.

3.3.4. Summary and conclusion of the local conformational changes of WT and mutant mFEN-1:

Base-pair opening of free DNA is the only effect that increases the freedom of movement of 2-AP bases and changes the electronic interaction strength between them. However, without un-pairing, enzyme binding introduces other effects on this electronic interaction that are depend-ent on its metal ions active site, 5'-phosphate monoester of exo substrate, and mutations. WTmFEN-1 induces specific substrate conformational changes in the presence of Ca^{2+} and 5'-phosphate monoester of +1 nt that position the scissile phosphate bond into the active site. These structural changes cause reorientation of the +1 and -1 nts differently by reducing the electronic interaction between -1 and -2 nts also between +1 and -1 nts by DNA twisting. WTmFEN-1 complex with either DF or 3'-SF substrate produces an ideal conformational change that produces rapid reaction in the presence of viable cofactor and prior to this reduces the electronic interaction between -1 and -2 nts more than +1 and -1 nts. However, without the 5'-phosphate of the 3'-SF the enzyme's ability to reorient +1 nt is decreased. This explains why exo-substrate without a 5'-phosphate decreases hFEN-1 catalytic efficiency.[5] In addition, A159V

94

and E160D mutant residues reduce the enzyme ability to produce ideal conformational changes. The contrary effect of mutations on +1 and -1 interactions in most cases supports the possibility of the imperfect reorientation of +1 and -1. However, A159V effects are stronger than E160D effects. The much more marked imperfect conformational change with A159V explains the catalysis view that A159V decreases cleavage ability of mFEN-1 more than E160D (*section 3.2)*.

3.4. Measuring substrate bending upon enzyme binding by Fluorescence Resonance Energy Transfer analysis (FRET):

Kinetic studies reported at the beginning of this chapter suggest that the FEN-1 mutations associated with cancer, A159V and E160D, may have modestly altered substrate binding properties. To test this a substrate binding assay that measures the ability of the enzyme to bend the two duplexes of substrate towards each other was set up exploiting fluorescence resonance energy transfer (FRET). FRET is nonradiative physical process that transfers energy from an excited donor fluorophore to another acceptor fluorophore.[94] The energy transfer efficiency (E) is a separation distance (R) dependent between donor and acceptor coupling, and is very sensitive to small changes in this distance *(equation 3.5)*. The R_0 (Foster distance) value, which is typically the distance of donor and acceptor pair at which the energy transfer efficiency is 50 %, depends on the overlap integral between the spectrum of the donor emission and the acceptor absorption, their molecular orientation, quantity of the donor fluorescence, the acceptor absorption coefficient, and the refractive index of the solution.[86,95] Therefore, labelling DNA strands with donor and acceptor fluorophores are used to measure the DNA conformational changes that occur upon protein binding by reporting change of the energy transfer efficiency (E).

A previous study showed that a protein and labelled-DNA binding can affect the donor and/or acceptor fluorescence intensity, which reflected DNA bending around protein surface.[96] Other later studies measured the FRET efficiency of a DF substrate attached to donor and acceptor molecules at different positions to indicate the substrate bending upon FEN-1 binding.[47,67,97-99]

$$[E] = \frac{1}{[1+(R/R_0)^6]} \qquad \textit{Equation 3.5}$$

In this study, the induced emission of the acceptor was used to measure the energy transfer efficiency (E), which was calculated using the (ratio)$_A$ method *(equation 3.6)*. The (ratio)$_A$ is determined from fluorescent signal of the donor-only labelled molecule (DOL) and doubly-labelled molecule (DAL) at two specific wavelength pairs *(equation 3.6)*, and N the normalization factor *(equation 3.7)*.[86] Fluorescence measurements were made by direct excitation of the donor (FAM) or acceptor (TAMRA) at 490 nm or 560 nm (2 nm slit width) and observing the emission

at 515-650 nm or 575-650 nm (5 nm slit width). Additionally, spectra of unlabelled (NL) DNA sample was deducted from the appropriate emission spectra of the labelled DNAs in order to be corrected from effects of buffer and enzyme conditions. Because of the overlap between TAM-RA excitation and FAM emission spectra, donor excitation (λ_{EX}^{D}) of double-labelled DNA (DAL) causes the acceptor emission (λ_{EM}^{A}) resulting FRET signal ($F_{DA}\left(\lambda_{EX}^{D},\ \lambda_{EM}^{A}\right)$). In short, this FRET signal is composed of three fluorescent spectra, which are acceptor emission due to FRET, acceptor emission from direct excitation, and donor emission from direct excitation. To achieve pure FRET with acceptor emission from FRET and direct excitation only, donor emission $F_{D}(\lambda_{EX}^{D},\lambda_{EM}^{D})$ that is measured by donor only–labelled DNA (DOL) and normalised to match donor peak of DAL molecule should be subtracted from the FRET signal of DAL molecule. As well, the resultant pure FRET after subtraction is normalised by the acceptor maximum emission by its direct excitation in DAL molecule $F_{DA}\left(\lambda_{EX}^{A},\lambda_{EM}^{A}\right)$. Where F_{DA} and F_{D} represent the fluorescent signal of the doubly-labelled molecule (DAL) and donor-only-labelled molecule (DOL) at the given wavelengths respectively.

$$(ratio)_A = \frac{F_{DA}\left(\lambda_{EX}^{D},\lambda_{EM}^{A}\right) - N.F_{D}\left(\lambda_{EX}^{D},\lambda_{EM}^{D}\right)}{F_{DA}\left(\lambda_{EX}^{A},\lambda_{EM}^{A}\right)} \qquad \textbf{\textit{Equation 3.6}}$$

$$N = \frac{F_{DA}\left(\lambda_{EX}^{D},\lambda_{EM}^{D}\right)}{F_{D}\left(\lambda_{EX}^{D},\lambda_{EM}^{D}\right)} \qquad \textbf{\textit{Equation 3.7}}$$

The determined $(ratio)_A$ then was used to calculate energy transfer efficiency (E). Functionally, E depends on $(ratio)_A$ and the molar absorption extinction coefficients of donor (ε^D) and acceptor (ε^A) at the given excitation and emission wavelengths *(equation 3.8)*. ε^D (490)/ ε^A(560) and ε^A (490)/ ε^A(560) were determined from absorbance spectra of DAL substrate and excitation spectra of AOL substrate respectively.

$$E = \frac{(ratio)_A}{\left(\frac{\varepsilon^D (490)}{\varepsilon^A (560)}\right) - \left(\frac{\varepsilon^A (490)}{\varepsilon^A (560)}\right)} \qquad \textbf{\textit{Equation 3.8}}$$

Data were fitted by nonlinear squares regression by KaleidaGraph version 4.5.2 using *equation 3.9* below:

$$E = E_{min} + \frac{(E_{max}-E_{min})}{2[S]}\left[([S]+[P]+K_{bend}) - \sqrt{([S]+[P]+K_{bend})^2 - 4[S][P]}\right] \qquad \textbf{\textit{Equation 3.9}}$$

Accordingly, the dissociation constant (K_{bend}) value is calculated using the measured E at a specific total enzyme concentration [P] and substrate concentration [S], minimum energy transfer efficiency of the free substrate in absence of any enzymes (E_{min}) and the maximum energy transfer efficiency of the saturated substrate (E_{max}).

For mFEN-1 FRET experiments, only DF and 3'-SF labelled with donor and acceptor were used *(figure 3.18)*. Using (ratio)$_A$ method, E were determined by measuring the enhanced acceptor fluorescence at 37 °C of 10 nM of NL, DOL and DAL of DNA substrates. The experiments were carried out with and without catalytically non-viable Ca^{2+} ions to allow substrate bending without reaction. The donor or acceptor was excited at 490 nm or 560 nm and the emission signals were collected from 515-650 nm or 575-650 nm. These emission signals were corrected for buffer and enzyme background signal by subtracting the signal from NL substrate signal. The overall substrates conformations were compared to duplex DNA.
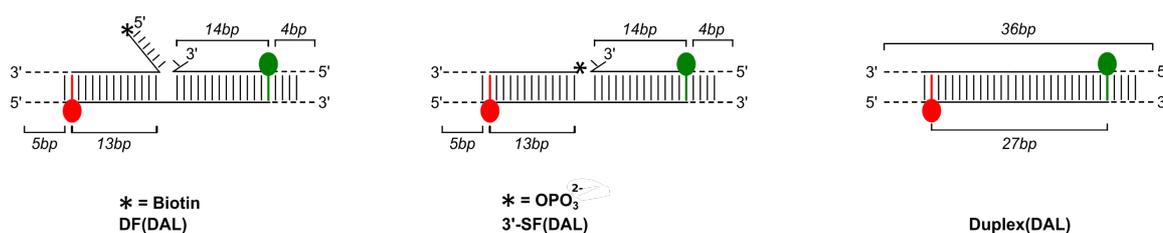
Figure 3.18. *Schematic of double flap (DF), 3'-single flap (3'-SF) and duplex DNA used in mFEN-1 FRET studies. Each **DAL** substrate contains **donor** (**fluorescein**) and **acceptor** (**TAMRA**). Non-labelled (NL), donor only labelled (DOL) and acceptor only labelled (AOL) substrates have versions of these constructs were also used.*

3.4.1. The 5'-flap is not required for DNA bending upon addition of WTmFEN-1:

The data observed with free DF and 3'-SF reported that the energy transfer efficiency (E) of the DF alone was similar ±Ca^{2+} (0.26-0.28), but was slightly higher than the E of the free 3'-SF, which also had similar E ±Ca^{2+} (0.19-0.20) *(figure 3.19.a)*. However, both of the free substrates had significantly greater E than the corresponding free duplex (0.1) *(figure 3.19.d)*. The calculated maximum distances (without protein) between donor and acceptor pair using *equation 3.5* were consistent with the above E details *(figure 3.19.e)*. The free DF, 3'-SF and DNA duplex substrates had similar r$_{max}$ ±Ca^{2+} and around 65 Å, 70 Å and 80 Å respectively. This indicates that the DF and 3'-SF substrates have an overall similar conformation that are more bent than a DNA duplex, even before adding protein. This is in agreement with the previous suggestion derived from single molecule experiments in which, a DF substrate was proposed to rapidly fluctuate between a bent and a linear extended conformation.[97] Furthermore, the slightly lower E of the free 3'-SF substrate indicates an overall bigger bending angle of the substrate than the free DF substrate, which might be expected as the number of flaps is reduced.

The gradual addition of WTmFEN-1 to doubly-labelled DF or 3'-SF substrate increased the energy transfer efficiency (E) significantly by increasing enzyme concentration until each substrate plateau was reached at saturating protein, regardless of whether metal ions were present or

not *(figure 3.19.a)*. When both substrates were saturated and fully bound to mFEN-1 (E at end-point), DF (DAL) substrate showed slightly lower E at the endpoints (≈5 %) than the corresponding 3'-SF ±Ca^{2+} *(figure 3.19.a.d)*. In addition, each substrate showed slightly higher E values that were reproducibly observed with Ca^{2+} ions present than the corresponding substrate signal in EDTA, which showed ≈10 % decrease *(figure 3.19.d)*. Surprisingly, these small differences of the E values at the end points were not enough to reflect any effect on the calculated minimum distance (at saturating protein) between donor and acceptor pair using *equation 3.5 (figure 3.19.e)*.

Both substrates (DF and 3'-SF) showed gradual decrees of distance to reach r_{min} values were estimated with ≈ 54 Å. The derived equilibrium dissociation constants K_{bend} were similar for both substrates in Ca^{2+}, however, it increased by a factor of two in EDTA for both substrates (DF: 8 ± 1 nM and 27 ± 2 nM / 3'-SF: 10 ± 1 nM and 21 ± 1 nM) *(figure 3.19.f) (table 3.5)*. The main cause of the endpoint variance at ±Ca^{2+} and between DF and 3'-SF is unknown. However, this variance does not affect DNA bending according to the similar r_{min} of both substrates ±Ca^{2+}. The K_{bend} values reflect that mFEN-1 has ability to bend DF and 3'-SF substrates similarly with and without metal ions presence *(table 3.5)*. This establishes that presence of the 5'-flap is not necessary for the global substrate bending. Relying on the previous indication of the importance of the divalent metal ions for 5'-flap threading[51] and position of the scissile phosphate bond into the active site,[37] these results mean that the DF and 3'-SF substrates bind and bend with similar affinity regardless either of these features have taken place. This is also consistent with kinetic *(section 3.2.1)* and ECCD *(section 3.3.3)* results of WTmFEN-1, which indicated that the divalent metal ions of WTmFEN-1 are necessary to induce specific conformational changes of the substrate to position the scissile phosphate bond into the active site and start the chemistry.
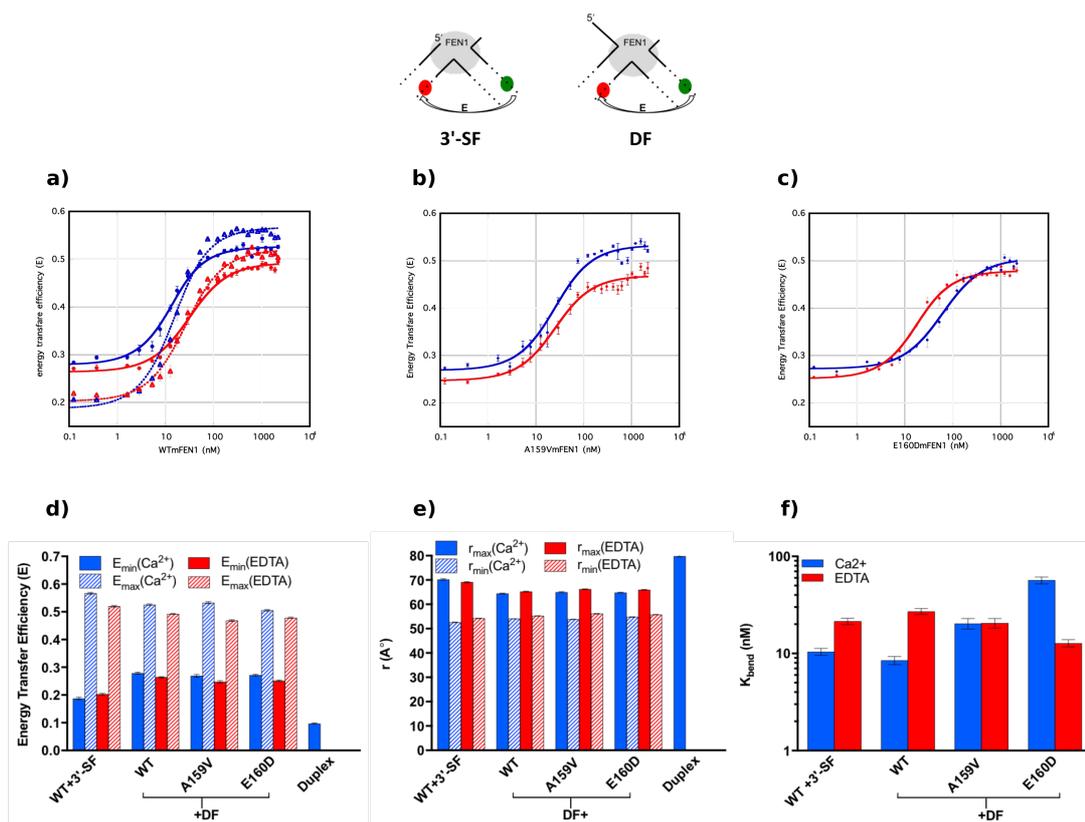
Figure 3.19. *FRET data in Ca^{2+} and EDTA. FRET binding curves of **a)** 3'-SF (dots) and DF (solid) to WTmFEN-1. **b)** DF to A159V. **c)** DF to E160D. The data points were fit to equation 3.9. **d)** The measured E_{min} and E_{max}. **e)** The calculated r_{max} and r_{min} using equation 3.5. **f)** The observed K_{bend} values. Each DAL substrate structure is shown and contains **donor** and **acceptor**.*

Substrate	mFEN-1	K_{bend} (nM) (Ca^{2+})	K_{bend} (nM) (EDTA)
3'-SF	WT	10 ± 0.8	21 ± 1.6
DF	WT	8 ± 0.8	27 ± 2
	A159V	20 ± 2	21 ± 2
	E160D	57 ± 5	13 ± 1

Table 3.5. *K_{bend} parameters for 3'-SF bound to WT and DF bound to WT, A159V and E160D mFEN-1 that were determined by FRET.*

3.4.3. Active site mutations (A159V & E160D) do not prevent bending but produce raised K_{bend}:

To investigate the requirements for DF substrate bending, A159V and E160D mFEN-1 were also tested. The mFEN-1 conserved residues A159 and E160 are located in α7 forming part of the mFEN-1 active site[5] from which interacts with the substrate positioning it to react. Both of the

mutated proteins increased FRET signals of DF (DAL) $\pm Ca^{2+}$ upon increasing their concentrations until the substrate plateau was reached at saturating protein *(figure 3.19.b.c)*. As with the WTmFEN-1, only subtle variations in the energy transfer efficiency (E) at end points were observed with and without divalent metal ions. A159V and E160D proteins showed 10 % and 5 % decrease of the E at end points in EDTA respectively compared to metal ions presence *(figure 3.19.b.c.d)*. Increasing mutated protein concentrations showed gradual decreasing of the DF (DAL) distance between donor and acceptor pair, which was calculated using *equation 3.5,* to reach similar $\pm Ca^{2+}$ r_{min} of WT ($\approx$ 55 Å) *(figure 3.19.e)*. In contrast, the calculated K_{bend} values showed differences from WT *(figure 3.19.f) (table 3.5)*. In Ca^{2+}, A159V and E160D showed 3-fold and 7-fold increase of K_{bend} values compared to WT, respectively. However, in EDTA, both mutations only decreased the K_{bend} values slightly compared to WT, as might be predicted as the substrate would not be expected to access the active site without divalent metal ions present *(table 3.5)*. Notably, all of the E and r_{min} changes demonstrate that DF (DAL) substrate adopts a bent conformational change with both mutated mouse proteins (A159V & E160D). According to the WT mutant mFEN-1 K_{bend} values, A159V mutation reduces the protein stability upon DNA binding and bending slightly, while E160D mutation decreased it more significantly. This demonstrates that A159V mutation may not be as critical as E160D mutation to DNA binding and bending. This may explain the low efficiency of these mutations in the kinetic study and proves that the lower catalytic efficiency of A159V is not because of worse bending *(section 3.2.2)*. In addition, one possibility is that failure to bend the substrate as efficiently as for wild type also impacts on the ECCD signal for the mutations *(section 3.3.4)*.

3.4.4. Summary and conclusion of the substrate bending upon WT and mutant mFEN-1 binding:

Collectively, FRET assays suggest that WTmFEN-1 has nanomolar affinity for DF and 3'-SF substrates, as summarized in *table 3.2*. Although differences in the bend angle were indicated with both substrates without protein compared to the free duplex, the measured dissociative constants ($\pm Ca^{2+}$) were very similar for both substrates and in accordance with the previous DF measurements. Accordingly, the absence of the 5'-flap does not affect WTmFEN-1 ability to bend 3'-SF substrate to reach similar r_{min} and E_{max} of DF substrate. Interestingly, despite the previous evidence of the importance and requirement of divalent metal ions bound in the active site for 5'-flap threading and scissile phosphate bond position, their absence does not affect WTmFEN-1 ability to produce the global substrate bending. Hence, it is clear that bending can occur without metal ions active site presence (in EDTA) or FEN-1 reaction features that need presence of the metal ions such as 5'-flap threading, or target phosphate bond position. The

importance of the A159 and E160 residues for DF substrate bending are inferred from their effect on FRET assay. The data of these mutated residues of A159V and E160D proteins suggest that while A159V mutation reduces mFEN-1 ability to bend DF substrate slightly (3-fold), E160D mutation has a much greater effect (7-fold). Thus, A159V mutation may not as critical as E160D mutation to DNA binding and bending.

3.5. Summary and conclusion of mFEN-1 (kinetic, ECCD and FRET):

The work described in this chapter has used different approaches to study mFEN-1 mechanism and the effect of active site mutations. Previous biochemical studies have described FEN-1 as a multifunctional enzyme with fen, exo and gen activities that preferred to interact with substrates that have a junction between two duplexes.[1,3,39] Despite the scarcity of many details about FEN-1 mechanism, many previous studies agree on the general features of the mechanism, which are substrate binding and bending,[2,5,47,97] 5'-flap threading,[5,50] and double nucleotide unpairing[3,44,64] or untwisting[93] to start chemistry in the active site. The existence of mutated FEN-1 proteins in cancer cells highlighted studying their effects on the enzyme efficiency. Two previous studies have reached conflicting results regarding the effects of mFEN-1 active site mutations. While one study indicated decreases of exo and gen activities of A159V and E160D mutated mFEN-1 with retention of fen activity,[18] a later study indicated reduced of fen activity of E160D mFEN-1.[19] Accordingly, this chapter used several approaches to first verify the validity of the enzyme possession of different activities, and second to confirm which of the previous studies around A159V and E160D active site mutations effects is more acceptable.

The kinetic study of WTmFEN-1 with different substrates proved that the enzyme interacts with different substrate structures, which have a junction between two duplexes and 1nt 3'-flap, similarly except the slight lower catalytic efficiency with 3'-SF substrate. In contradiction with the previous enzyme characterisation, WTmFEN-1 does not have different activities, but has the ability to recognise different substrates using the same chemistry. Both ECCD and FRET WTmFEN-1 results monitored the same structural changes and bending ability of DF and 3'-SF substrates upon the enzyme binding. Taken all together, substrate without 5'-flap can be bent and structurally changed ideally upon FEN-1 binding, but the chemistry can be affected leading to reduced enzyme catalytic efficiency.

The mutated residues proposed to be involved in mFEN-1 active site (A159V & E160D) were used in this study to investigate the effect of mFEN-1 active site mutations. Each mutation reduces the enzyme's ability to interact with all substrates similarly. The mutated A159V residue showed a greater reduction effect that is attributed to the mutated longer side chain, which

causes steric hindrance can influence substrate entry. The E160D mutation showed a smaller rate reduction related to the shorter aspartic acid side chain. Because of the direct contact of this residue with the metal ions in the active site, this mutation may increase the distance between them and decreases enzyme ability to hydrolyse the scissile phosphate bond. In addition, both mutations have hindered mFEN-1 ability to produce ideal conformational changes and bending angle that are required for enzyme interaction with all substrates. In contradiction with the previous studies of mFEN-1 mutations effect, which indicated decreasing of one or two activities of mFEN-1, these mutations affect the enzyme ability to organise all substrates similarly (DF, 3'-SF, and gap).

Interestingly, DF2'OH substrate decreases mFEN-1 mutations effects on chemistry, particularly A159V. The presence of the 2'OH effects the chemical environment producing a more reactive substrate that appears to overcome the steric hindrance and alterations to the distance between the two metal ions in the active site.

Chapter 4: Results and Discussion of hFEN-1:

As indicated before, FEN-1 biological function requires the enzyme to be able to recognise the correct DNA substrate and efficiently carry out a specific reaction, but to exclude, or drastically slow, the reaction of unsuitable DNAs. In turn, this relies on the interactions of enzyme and correct substrate both of which undergo conformational change during the recognition process. Thus both substrate structure and FEN-1 mutations can affect the enzyme efficiency. Substrate conformational changes involve bending of the DNA that identifies the presence of a DNA junction and allows the DNA-protein complex to form and a local DNA conformational change that permits the substrate to transfer to the active site. Studying the impact of DNA and protein structure on the ability of substrates to undergo these conformational changes will reveal the features of the protein that allow FEN-1 to discriminate between potential substrates.

A number of WT and different single, double and quadruple hFEN-1 mutations were chosen to study their effects on local substrate conformational changes and DNA bending *(figure 4.1)*. All these hFEN-1 mutations were expressed, purified, tested for folding and kinetically characterised by other members in the group (Dr. Nikesh Patel, Dr. Jack Exell and Dr. Mark Thompson). The effects of different mutations which will be studied on the maximal single turnover rates (k_{STmax}) are summarised in *table 4.1*.[93,100,101] Most of these mutations K93A, R100A, R104A, L130P, K132A, K93AR100A, R103AR129A, R103ER129E, R104AK132A, R104EK132E, and Quad E (R103ER104ER129EK132E) were within the hFEN-1 helical arch. Some of these mutations are of residues that are conserved across the 5' nuclease superfamily (K93A and R100A) and in the so-called gateway that surrounds the active site. The mutation L130P was designed to test whether the non-superfamily conserved helical cap structure was important for FEN-1 catalytic processes and substitutes a secondary structure disrupting for a leucine residue. Others (R104A, K132A, K93AR100A, R103AR129A, R103ER129E, R104AK132A, R104EK132E, and Quad E (R103ER104ER129EK132E)) were identified on the basis of the later D86N hFEN-1 threaded substrate structure (PDB code 5UM9) as candidates for electrostatic interactions between the substrate +1 phosphate diester and the helical cap. FEN-1-conserved Y40 was also included in the study, at this α-2 residues from the other half of the gateway. In addition, active site carboxylate, D181A was also included. Finally, L53A the 3'-flap pocket mutation was also involved in the study.
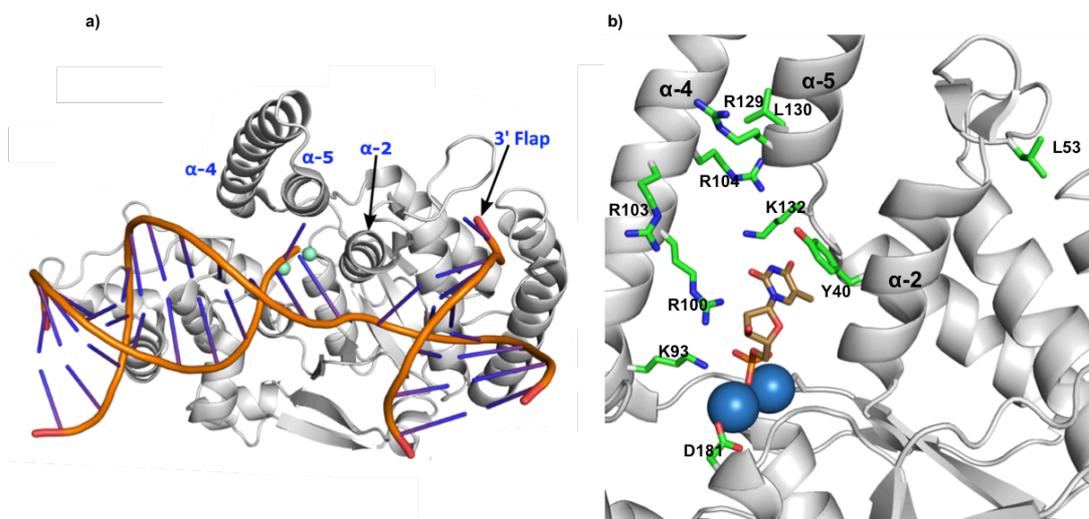
Figure 4.1. *Position of the studied mutated hFEN-1 residues.* ***a)*** *Schematic of DNU mechanism to position the scissile phosphodiester bond between +1 and -1 nts between the active site* **metal ions**. ***b)*** *Cartoon represents the active site of hFEN-1:product structure to show the -1 nt in contact with* **metal ions** *active site and the studied mutated residues from the helical gateway (base α2-α4), cap (top of α4 and α5) and 3'-flap pocket.*

hFEN-1	With DF rate decrease by---fold	With 3'-SF rate decrease by---fold
Y40A [2]	100	-
L53A [3]	1000	-
K93A [2]	2000	-
R100A [2]	7000	-
R104A [3]	40	-
L130P [1]	2000	-
K132A [3]	4	2
D181A [2]	10,000	-
K93AR100A [2]	6000	-
R103AR129A [3]	65	40
R103ER129E [3]	1500	-
R104AK132A [3]	220	40
R104EK132E [3]	4000	-
Quad E [3]	3000	-

Table 4.1. *Decrease of single turnover rates (K_{STmax} values) for cleavage of DF and 3'-SF substrates by the different studied mutations relative to WThFEN-1. The superscript numbers to indicate the person that performed the kinetic study: (1) Dr. Nikesh Patel, (2) Dr. Jack Exell or (3) Dr. Mark Thompson*

Accordingly, the arginine (R) and lysine (K) basic residues were mutated to alanine (A) or glutamic acid (E) to remove the attractive positive charge that effects the interaction with the DNA, the aliphatic leucine side chain (P) was mutated to proline (P) or alanine (A) to reduce the side chain length that has roles to maintain the protein secondary structure, tyrosine side chain was mutated to alanine to remove its aromatic ring that effect the interaction with the -1 nt of the DNA, and finally aspartic acid side chain (D) was mutated to alanine (A) to remove the attractive negative side chain that effect active site metal ions position. By comparison between the WT and mutated hFEN-1, the enzymes studied allowed the impact of helical arch mutations such as Y40A, K93A, R100A, K93AR100A and L130P, and active site mutation D181A on the local substrate structural changes to be studied (using ECCD). Using FRET technique, wild type and mutated hFEN-1 with helical arch mutations Y40A, K93A, R100A, R104A, L130P, K132A, K93AR100A, R103AR129A, R103ER129E, R104AK132A, R104EK132E and Quad E (R103ER104ER129EK132E), active site mutation D181A, and 3'-flap pocket mutation L53A were studied to indicate their effects on substrate bending upon protein binding.

4.1. Investigation of local conformational changes of substrate reacted duplex in hFEN-1 active site by low energy circular dichroism of 2-aminopurine (ECCD):

The ECCD experiments that were used to study mutated hFEN-1 proteins were based on the same theoretical basis as described before in *chapter 3 (section 3.3)* and used $(2\text{-AP})_2$ containing substrates. First, WThFEN-1 was studied with 3'-SF and 5'OH3'-SF substrates in the presence and absence of EDTA to investigate the importance of the active site metal ions and 5'-phosphate monoester that were indicated previously with mFEN-1 in *section 3.3.2*. Second, the ECCD technique was also used to illustrate effects of some helical arch mutated residues such as Y40A, K93A, R100A, L130P and K93AR100A, and the active site mutated residue D181A on the local DNA conformational change of 3'-SF and 5'OH3'-SF substrates upon protein binding compared to WThFEN-1. As summarized in *table 4.1*, previous studies have shown that under maximal single turnover conditions the mutations K93A, R100A, K93AR100A and L130P reduce the rate of the hFEN-1 reaction by factors of at least 2000,[51,100] and later studies showed that Y40 and D181A cause 10^2 and 10^4 rates decrease of hFEN-1 under single turnover conditions, respectively.[101] As a result, studying effects of such mutations can help to investigate the requirements and the importance of these local DNA conformational changes for an ideal hFEN-1 interaction.

4.1.1. Base-pair effects on local conformational changes of free ss and ds DNA:

The ECCD spectra of $(2\text{-AP})_2$ inserted into the ss DNAs and ds substrate constructs of 3'-SF and 5'OH3'-SF substrates were similar to those previously published[37] and that are described *section*

3.3.1. These results were obtained without enzyme to indicate base-pair effects and to be used as a basis for comparison with the subsequent experiments. The ECCD signals of the free DNAs (ss and ds) containing neighbouring 2-APs in Ca^{2+} did not change by adding EDTA *(figure 4.2.a.b)*. Furthermore, comparison of the ECCD intensity at 326 nm of three repeats of each free DNA (ss and ds) in Ca^{2+} and in EDTA *(figure 4.2.c)* also consistent with the typical ECCD traces *(figure 4.2.a.b)*. ssDNAs gave weaker ECCD signals and lower intensity at 326 nm combined with small blue shifts resulting in lower λ_{max} values compared to the free dsDNAs *(figure 4.2)*. Variance of the inserted adjacent 2-AP, whether in +1,-1 or -1,-2, did not affect either ECCD traces or the intensity at 326 nm of the single strand *(figure 4.2)*.

Accordingly, the base-pairing is the only effector on the adjacent 2-APs interaction in absence of the enzyme. While absence of the 5'-phosphate monoester did not affect the ECCD signals, base-pairing of the DNA increased the ECCD response significantly from ss to ds DNAs. Hence, the base-pair opening is the only effect that can increase freedom of movement of the adjacent 2-APs and reduce their interaction because of the weaker alignment of electronic transition dipoles in the ssDNA. This is in good agreement with the previous spectra of ssDNA having 2-APs[89] and the previous finding with mFEN-1 results *(section 3.3.1)*. In spite of this finding, the evidences from these results are not enough to identify the nature of this force in presence of enzyme, which could be reoriention and/or alteration of the distance of the adjacent bases with or without douple nucleotide unpairing.
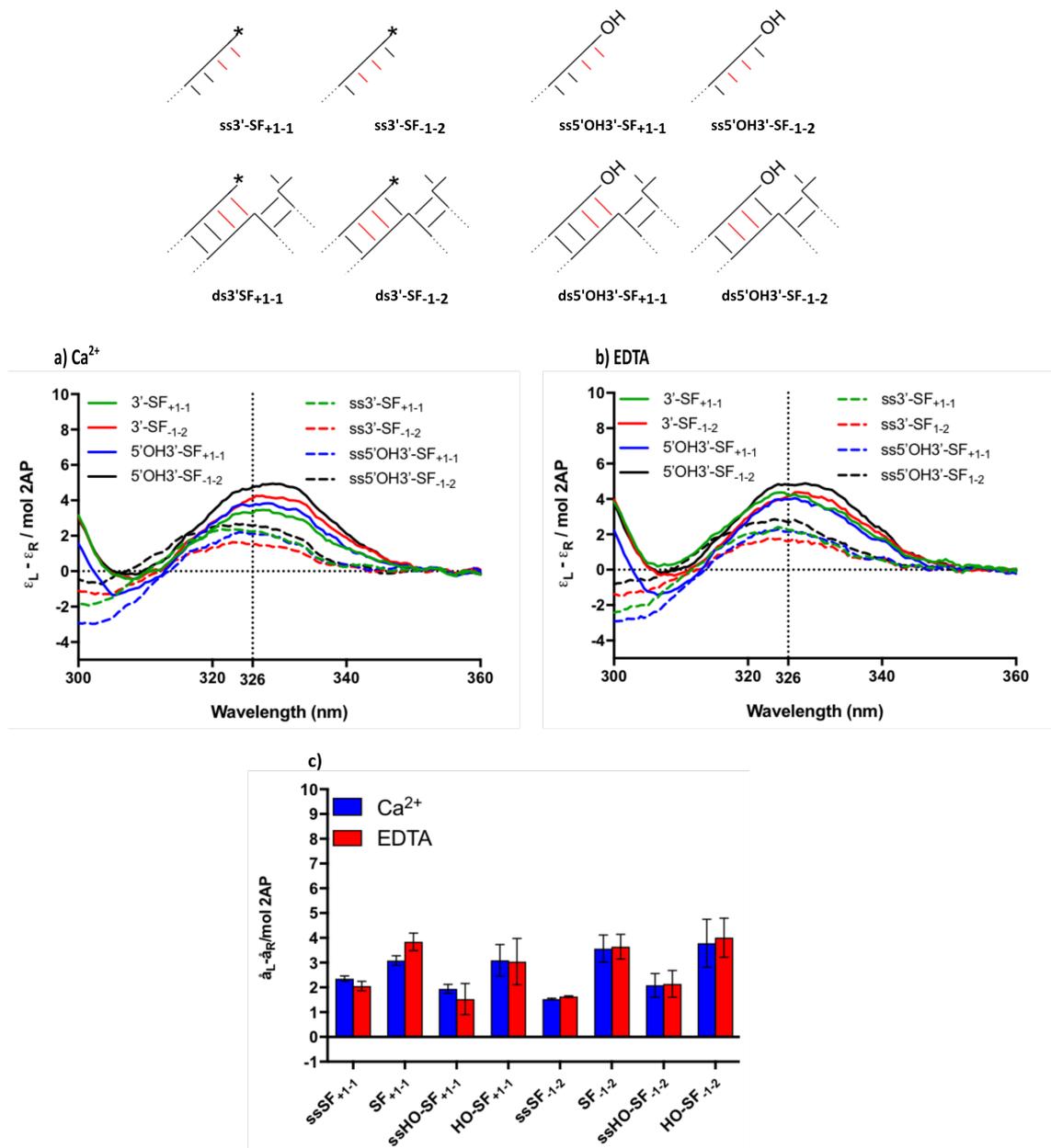
Figure 4.2. *ECCD spectra of free single strands (ss) and double strands (ds) of DNAs containing 2-AP dimer.* **a)** *and* **b)** *The ECCD traces of the unbound ds represented in solid lines, while their corresponding ss represented in dashed lines in Ca²⁺ and EDTA respectively.* **c)** *Ellipticity changes at 326 nm for ss and ds in* **Ca²⁺** *and* **EDTA**. *All data sets are three experimental repeats and standard errors are shown. DNA constructs are shown schematically. 2-AP was represented in* **red**, *a 5'-monophosphate was represented as* ***** *and a 5'-hydroxyl was represented as* **(OH)**.

4.1.2. Importance of hFEN-1 active site metal ions and the 5'-phosphate monoester for local conformational changes of the DNA reacting duplex upon WT protein binding:

An earlier report[5] indicates the importance of the 5'-phosphate of the +1 nt of the exo substrate kinetically for efficient hFEN-1 reactions. In addition, experiments with mFEN-1 demonstrate its importance for substrate binding and for producing ideal conformational changes upon mFEN-1 binding *(section 3.3.2)*. To test the universality of this earlier observation, ECCD results of WTh-FEN-1 with 3'-SF and 5'OH3'-SF substrates were used to probe the specific structural changes upon human protein binding. These studies revealed the importance of the active site metal ions and the 5'-phosphate for hFEN-1 to produce ideal conformational changes.

WThFEN-1 showed similar results to mFEN-1 complexes with 3'-SF and 5'OH3'-SF substrates. Both 3'-SF$_{-1-2}$ and 5'OH3'-SF$_{-1-2}$ substrates showed similar dramatic change in the ECCD response with a deep minimum ECCD signal at around 310 nm and low ECCD intensity at 326 nm in Ca^{2+} *(figure 4.3.c.g* and *figure 4.4.b.d)*. However, when the conformational change was studied with 2Aps at +1 and -1 position there were marked differences between substrates ±phosphate. At 326 nm with FEN-1- Ca^{2+}3'-SF$_{+1-1}$ produced an ECCD signal of 1.5 ± 0.4 (M AP)$^{-1}$.cm^{-1}, close to that of free ssDNA. In contrast, 5'OH3'-SF$_{+1-1}$ had much greater signal close to that of dsDNA when bound to the enzyme in Ca^{2+}. The ECCD signals and the intensity at 326 nm for the four substrates increased significantly by adding EDTA *(figure 4.3* and *figure 4.4)*.

As the 3'-SF substrate is one of the preferred substrates of hFEN-1 enzyme, its structural change upon addition of metal ions is presumed to be as an ideal conformational change. Based on the ECCD changes of the human WT complexes with 3'-SF$_{+1-1}$ and 3'-SF$_{-1-2}$ substrates, the ideal conformational changes in Ca^{2+} suggested decreasing the electronic interaction between +1 and -1 nucleotides also between -1 and -2 nucleotides. This may result from different origins such as reorientation and/or changes in the distance of the adjacent bases +1 and -1. In addition, the comparison between the +1,-1 and -1,-2 substrates showed that the electronic interaction is reduced between -1 and -2 nucleotides further than +1 and -1 nucleotides. This result of the ideal conformational changes is in consistent with the previous result that obtained with mFEN-1 *(section 3.3.2)* showing that it is an intrinsic property of FEN-1 proteins from different sources.
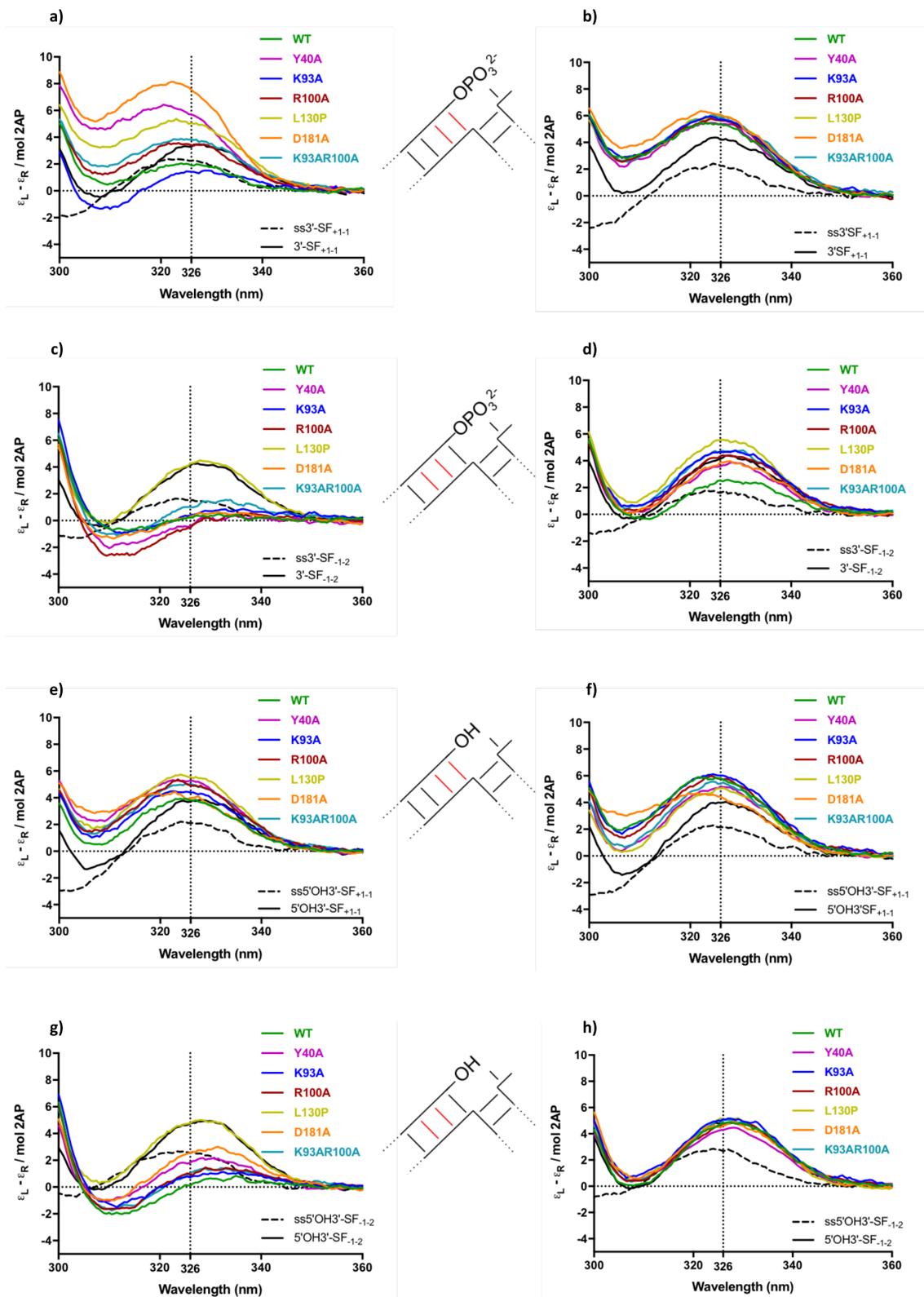
Figure 4.3. *ECCD spectra of WT and mutated hFEN-1 complexes in Ca^{2+} and EDTA beside the corresponding **ss** and **ds*** *of each substrate.* ***a)*** *hFEN-1:*3'-SF$_{+1-1}$ *complexes in Ca^{2+}.* ***b)*** *hFEN-1:*3'-SF$_{+1-1}$ *complexes in EDTA.* **c)** *hFEN-1:3'-SF$_{-1-2}$* *complexes in Ca^{2+}.* **d)** *hFEN-1:3'-SF$_{-1-2}$ complexes in EDTA.* **e)** *hFEN-1:5'OH3'-SF$_{+1-1}$ complexes in Ca^{2+}.* **f)** *hFEN-1:5'OH3'-SF$_{+1-1}$ complexes in EDTA.* **g)** *hFEN-1:5'OH3'-SF$_{-1-2}$ complexes in Ca^{2}.* **h)** *hFEN-1:5'OH3'-SF$_{-1-2}$ complexes in EDTA. DNA*

*constructs are shown schematically with 2-APs were represented in **red**, a 5'-monophosphate was represented as * and a 5'-hydroxyl was represented as **(OH)**.*
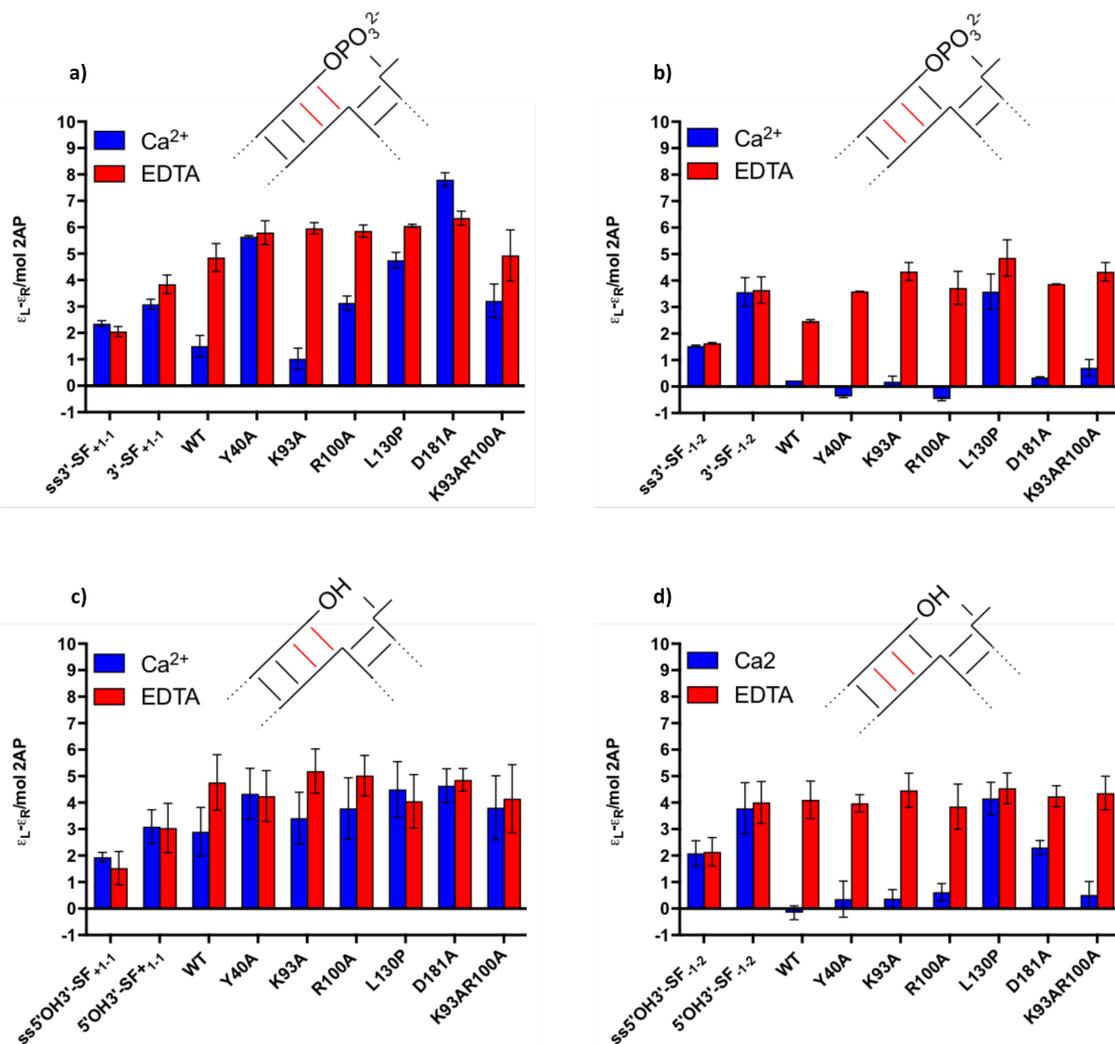
Figure 4.4. *ECCD intensity at 326 nm of free ssDNAs and dsDNAs, and hFEN-1 complexes with substrates containing 2-AP dimer in **Ca^{2+}** and **EDTA**. **a)** The ss, ds and complexes of 3'-SF$_{+1-1}$. **b)** The ss, ds and complexes of 3'-SF$_{-1-2}$. **c)** The ss, ds and complexes of 5'OH3'-SF$_{+1-1}$. **a)** The ss, ds and complexes of 5'OH3'-SF$_{-1-2}$. All data sets are three experimental repeats and standard errors are shown. DNA constructs are shown schematically. 2-AP was represented in **red**, a 5'-monophosphate was represented as * and a 5'-hydroxyl was represented as **(OH)**.*

According to these ECCD results, the two metal ions active site (Ca^{2+}) and the 5'-phosphate are also essential for the local structural change of the exo substrates upon WThFEN-1 binding. First, increased ECCD signals upon addition of EDTA confirms the importance of the active site metal ions to produce the local conformational changes of 3'-SF and 5'OH3'-SF substrates upon hFEN-1 binding, even if these changes are not ideal for the enzyme efficiency.[37] Second, the different behaviour of the WT:5'OH3'-SF$_{+1-1}$ complex compared to the corresponding complex of 3'-SF$_{+1-1}$ substrate in Ca^{2+} suggests that the 5'-phosphate of the exo substrate is required to pro-

duce an ideal structural changes upon hFEN-1 binding involving +1 and -1 nts, however, it is not required for -1 and -2 structural changes. In the case of the 5'OH3'-SF substrate, its loss of ability to direct the +1 nt correctly as required as in 3'-SF substrate suggests it to be closer to -1 nt than the ideal situation. This result is in line with and may explain the previous report about decrease of hFEN-1 reaction efficiency as a result of the lacking of the 5'-phosphate of the exo substrate.[5] Both results of the importance of the metal ions active site and the 5'-phosphate for the DNA conformational changes upon hFEN-1 binding are consistent with what were previously observed for DNA conformational changes upon mFEN-1 binding *(section 3.3.2)*. The importance of the metal ions active site is also in line with what was previously observed for DNA conformational changes study of DF_{-1-2} and P_{-1-2}.[37] As indicated before *(section 3.3.2) (figure 3.17)*, more details about the nature of this conformational change were derived by combining WT and mutated hFEN-1 results with other later results.[93]

4.1.3. Effects of hFEN-1 helical arch mutations and active site mutation on local conformational changes of reacting duplex upon their complexes with 3'-SF and 5'OH3'-SF compared with WT:

To investigate the requirements and the importance of the ideal DNA conformational changes, effects of helical arch mutations K93A, R100A, K93AR100A, L130P and Y40A, and active site mutation D181A were studied with 3'-SF and 5'OH3'-SF *(figure 4.5)*. The residues K93 and R100 are located at the base of α4 forming a part of the hFEN-1 helical gateway[5] from where they protrude into the hFEN-1 active site and are not predicted to be involved in substrate interactions until the DNA is positioned to react. L130 is a component of the helical cap (α5) so it is far from the active site, but the L130P mutation is presumed to affect the secondary structure of the cap.[100] Y40 is a part from the gateway (α2) and close to the active site seen to interact with the +1 nucleotide of the substrate complex, whereas it stacks on the -1 nucleotide in the product complex structure.[1,5] D181 is an active site residue with direct contact with the catalytic metal ions in hFEN-1 structure.[5] D181A mutation could alter the number of metal ions bound and/or their positions.
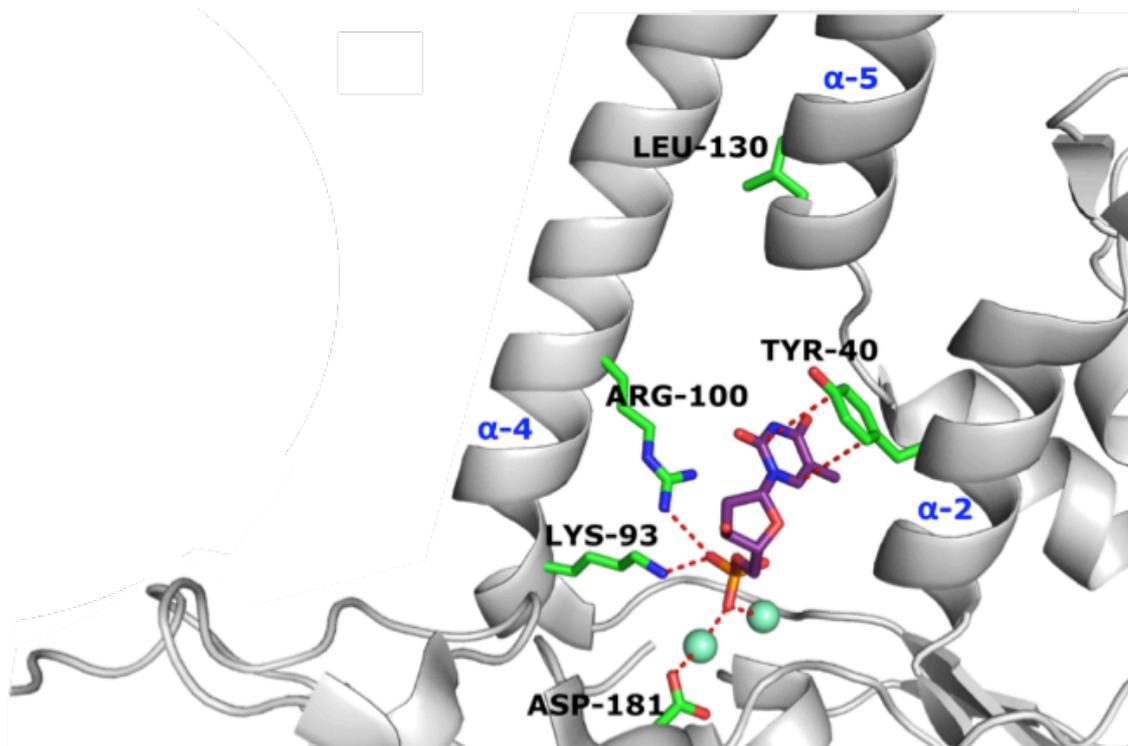
Figure 4.5. *Cartoon represents the active site of hFEN-1:product structure during the DNU mechanism to show the -1 nt in contact with metal ions active site and the studied mutated residues from the gateway (base α2-α4), cap (top of α4 and α5) and the active site. The oxygen and nitrogen atoms of the side chains were shown.*

3'-SF$_{-1-2}$ substrate was bound to WT and mutated hFEN-1 to investigate mutations effects on the substrate conformational changes. K93A and D181A were able to bring an analogous ECCD spectrum and so close ECCD intensity at 326 nm to WT protein with Ca^{2+} *(figure 4.3.c and figure 4.4.b)*. In fact, it is surprising for D181A brought about a similar ECCD response to WT. This suggests that this mutation does not affect the metal dependent ideal local conformational changes of hFEN-1, which is in contrary to the previous observation of no active site metal ions visible in an X-ray structure of D181A bound to 3'-SF substrate in the presence of Ca^{2+} where the DNA remained base paired and not within the active site.[5]

As seen previously with hFEN-1:DF$_{-1-2}$ complexes[37,100] and mFEN-1: DF$_{-1-2}$ complexes *(figure 3.16.f in section 3.3.3)*, spectra of 3'-SF$_{-1-2}$ produced by R100A and Y40A with Ca^{2+} contained an additional minimum at 310 nm *(figure 4.3.c)* and lower ECCD intensity at 326 nm than WT. In contrast, the double mutant (K93AR100A) showed higher ECCD intensity at 326 nm than WT *(figure 4.4.b)*. This suggests that these mutations (Y40A, R100A and double mutant) have an altered conformational change of the -1 and -2 nts from that produced by WT, D181A and K93A hFEN-1. According to the ECCD changes, R100A, Y40A and double mutant reduce the electronic interaction between the two adjacent 2-APs of the DNA compared to the corresponding free DNA, which reflects increasing distance and/or an alternative orientation of them. However, compared to the WThFEN-1 complex, while R100A and Y40A increase the distance more than re-

quired as in WT, K93AR100A increases it but still less than required as in WT. Thus,Tyr40 and Arg100 are required to correctly position the substrate for reaction. In contrast, the ECCD trace and the ECCD intensity at 326 nm with L130P were similar $\pm Ca^{2+}$ and close to the corresponding free dsDNA indicating that this mutation that alters the structure of the helical cap prevents the local conformational changes of 3'-SF$_{-1-2}$ completely.

When the same mutated hFEN-1s were employed with 3'-SF$_{+1-1}$ substrate, only K93A most closely resembled the behaviour of the WT protein. However, Y40A, R100A, L130P, D181A and K93AR100A did not significantly alter the signal with 3'-SF$_{+1-1}$ at 326 nm $\pm Ca^{2+}$. This reflects that these mutated helical arch residues produce imperfect or prevent the protein form bringing about changes of +1 and -1 nts compared to the WT protein. Thus, the basic residues of the arch, the intact structure of the cap and the proper positioning of metal ions by D181A are all required to produce the substrate conformation presumed to be required for reaction.

The conformational changes of the 5'OH3'-SF substrate was studied with WT and mutated hFEN-1 to indicate effects of the mutations and absence of the 5'-phosphate monoester on the local DNA conformational changes. According to the results in *section 4.1.2*, the absence of the 5'-phosphate of exo substrate affects the WThFEN-1 ability to bring about ideal local DNA conformational change involving the +1 and -1 nts in substrate DNAs unlike the conformational change involving the -1 and -2 nts, which are not affected at all. In addition to the similarity of the WThFEN-1 interaction with 3'-SF$_{-1-2}$ and 5'OH3'-SF$_{-1-2}$, all the mutated proteins had also the same response to 5'OH3'-SF$_{-1-2}$ as 3'-SF$_{-1-2}$ with the exception of D181A, where hFEN-1 in Ca^{2+} reduced the ECCD signal to a lesser extent *(figure 4.4.d)*. Hence, the absence of a 5'-phosphate does not prevent the WT and mutated hFEN-1 proteins bringing about a local DNA conformational change involving the -1 and -2 nts as in 3'-SF substrate (except D181A), but the orientation of 2-APs may differ from that adopted by the 5'-phosphorylated form *(figure 2.15)*. However, the absence of the 5'-phosphate monoester prevents D181AhFEN-1 enzyme ability to produce the ideal local conformational change within -1 and -2 nts as in 3'-SF substrate.

On the other hand, when the 5'-phosphate was removed from the 3'-SF substrate involving +1 and -1 of 2-APs (5'OH3'-SF$_{+1-1}$), ECCD signals of the hFEN-1s were significantly altered compared to 5'OH3'-SF$_{-1-2}$ complexes as they were for the WT protein. All showed a smaller decrease in ECCD signal at 326 nm in the presence of divalent ions relative to the same samples in EDTA *(figure 4.3.e.f and figure 4.4.c)*. Moreover, the maximum of the signal with hFEN-1:Ca^{2+} was blue-shifted relative to free corresponding dsDNA. This demonstrates the importance of the 5'-phosphate monoester of 3'-SF substrate to produce an ideal local conformational change involv-

ing the +1 and -1 nts. Earlier literature observed a 20-fold decrease in reaction efficiency of a 3'-SF substrate lacking the 5'-phoshate monoester measuring k_{cat}/K_M, but even larger impacts (~200-fold) on the rate of reaction at higher substrate concentration were measured recently. It seems likely that these rate decelerations can be assigned to the imperfect local conformational changes. Furthermore, the crystal structure of the substrate (without 5'-phoshate monoester) remained base-paired and did not move the active site despite the presence of active site metal ions. In combination, all these results suggest that the local DNA conformational change necessary for reaction requires a 5'-phosphate in the 3'-SF substrate or a 5'-phosphate diester in the DF substrate.[5]

4.1.4. Summary and conclusion of the local conformational changes of WT and mutant hFEN-1:

Taken together changes of the +1,-1 and -1-2 of 3'-SF substrate *(table 4.2)* indicate that when the hFEN-1 protein was altered to K93A, there was no effect on the structural change of the 3'-SF substrate in the presence of the divalent metal ions compared to its absence. Hence, K93 residue does not play a role in substrate positioning and the impact of its mutation to alanine seems to be entirely related to catalysis.[88]

In contrast, all other mutations affected the ideal conformational change. Although, D181A mutation produces an ideal conformational change within -1 and -2 as in WT, it prevents the ideal conformational change of the +1 and -1 nts. The ECCD changes suggest that the +1 nt may be stacked to -1 nt and moved analogously causing an increase in the ECCD signal. The lack of the perfect reorientation of the nucleobases in this instance may be related to metal ion positioning in the mutated protein (since D181 is directly coordinated to one of the active site M^{2+} ions).

On the other hand, Y40A, R100A and K93AR100A appear to play a role in reorientation of the +1 and -1 nts also -1 and -2 nts. Y40A and R100A reduced the ECCD signal, and the double mutant (K93AR100A) increased the signal with 3'-SF$_{-1,2}$ in presence of Ca^{2+} compared to WThFEN-1. This may reflect the single mutations effect to increase the distance between the -1 and -2 nts further than required as with WThFEN-1, while double mutation increased it but still lesser extent than with WThFEN-1. Though, these single and double mutations affect differently on +1 and -1 nts by stacking and orient them similarly to keep the distance between them as in dsDNA. The data presented here support the idea that the Y40 and R100 residues play an important role in optimal substrate positioning, and their mutations to alanine were found to reduce the rate of cleavage of DF substrate by a factor of 100 and 7000, respectively. Because both residues (Y40

and R100) contact the cleaved phosphate monoester in product structures, they may well position the scissile phosphate diester in active site positioned substrate complex *(figure 4.5)*.

In contrast, perturbation of the secondary structure of the helical cap (L130P) does not facilitate the local DNA conformational change when divalent metal ions are added, and prevent the inclusion of a +1 nt. This difficulty with the local DNA conformational change is consistent with the indicated 3000-fold decrease of the mutated enzyme ability to incise a 5'-flap substrate *(table 4.1)*. In theory this mutation could alter the positioning of all the other helical arch residues that have been studied here so this drastic impact could well be explained by a lack of a helical cap structure.

According to the +1,-1 and -1,-2 changes of 5'OH3'-SF substrate *(table 4.2)*, hFEN-1 protein reorients the -1 and -2 nts of the substrate as in 3'-SF substrate except with the mutation D181A. However, only small changes were observed with the 5'OH3'-SF$_{+1-1}$ ECCD signals and mutated enzymes compared to the WThFEN-1. Assuming that the position adopted by the substrate in the presence of hFEN-1:Ca^{2+} reflects the catalytically viable conformation, the 5'-phosphate monoester of 3'-SF substrate must form a key interaction required to assemble this state. This explains earlier work demonstrated that neutralization of the charge of this +1 5'-phosphate by conversion to methyl phosphate is detrimental to reaction.[102] In addition, this result is consistent with earlier work observed reducing of hFEN-1 efficiency to interact with exo substrate lacking 5'-phosphate monoester.[5]

S	2-Aps position	Y40A	K93A	R100A	L130P	D181A	K93AR100A
3'-SF	+1,-1	+	=	+	+	+	+
	-1,-2	-	=	-	+	=	+
5'OH3'-SF	+1,-1	=	=	=	=	=	=
	-1,-2	-	=	-	+	+	+

Table 4.2 *Strength of the electronic interaction between the two adjacent 2-Aps of +1,-1 and -1,-2 positions of the mutated hFEN-1 complexes comparing to the corresponding WT complex with DF and 3'-SF substrates. The used symbols mean that (+) higher interaction (-) less interaction (=) similar interaction to the corresponding WT complex. The data were summarised from figure 4.3.*

4.2. Measuring substrate bending upon enzyme binding by Fluorescence Resonance Energy Transfer (FRET):

FRET experiments were used to investigate the effects of substrate structure and mutations on substrate binding and bending brought about by hFEN-1. FRET experiments that were used to study the wild type and mutated hFEN-1 proteins were based on the same theoretical basis described before in *chapter 3 (section 3.4)*. First, WThFEN-1 was studied with the different substrates DF, $MMDF_{+1}$, 3'-SF, 5'-OH3'-SF, 5'-SF, nicked DNA and SA-DF substrates and compared to duplex DNA to investigate any roles for base pairing, 5' and 3' flaps and the presence of a 5'-phosphate monoester on bending the substrate upon enzyme binding *(figure 4.6)*.
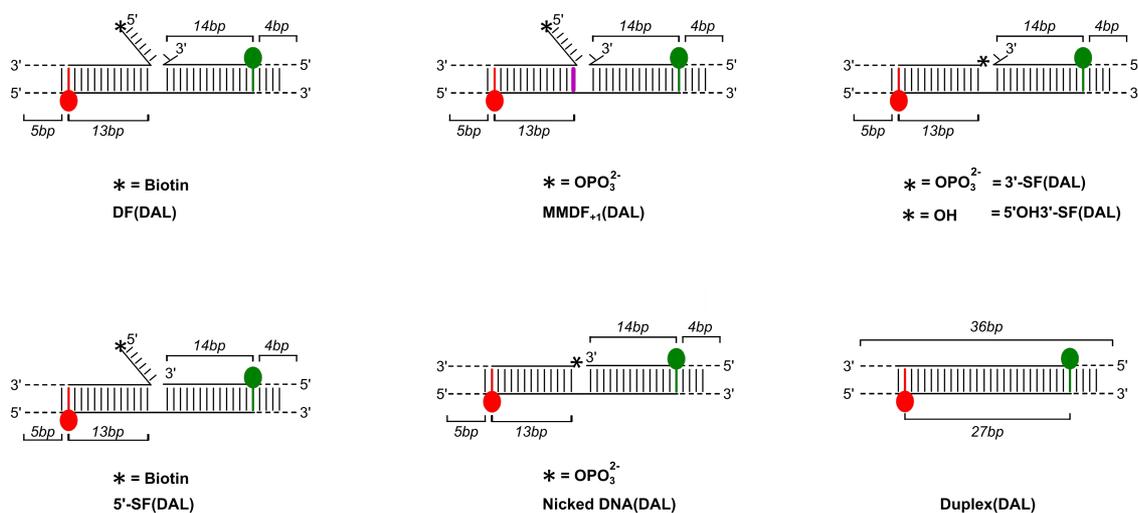
Figure 4.6. *Schematic of double flap (DF), mismatch double flap+1 ($MMDF_{+1}$), 3'-single flap (3'-SF), 5'OH3'-single flap (5'OH3'-SF), 5'-single flap (5'-SF), nicked DNA and duplex DNA used in hFEN-1 FRET studies. Each **DAL** substrate contains **donor** (**fluorescein**) and **acceptor** (**TAMRA**). Non labelled (NL), donor only labelled (DOL) and acceptor only labelled (AOL) substrates have versions of these constructs were also used.*

Second, mutated hFEN-1s were studied with DF, 3'-SF or both substrates to indicate their effects on substrate binding and bending *(figure 4.6)*. The different mutated helical arch residues Y40A, K93A, R100A, R103A, R103E, R104A, R104E, R129A, R129E, L130P, K132A, and K132E, the active site mutated residue D181A and the mutated 3'-flap pocket residue L53A were studied as a single, double or quadruple hFEN-1 mutations using FRET *(figure 4.7)*. According to the new D86NhFEN-1 threaded DF substrate crystal structure, the basic residues such as arginine (R) and lysine (K) interact with the DNA to position the phosphate backbone by specific electrostatic interactions leading to rotation dsDNA into the active site with fully basepairing intact. Active site carboxylate residues such as aspartic acid (D) with its negative charge affects position of the active site metal ion, and the tyrosine residue (Y) with its aromatic ring effects the stacking force

with either the +1 or -1 nt, which it is stacked upon in substrate and product structures respectively. Differently, the leucine residue (L) with its aliphatic side chain affects the protein structure. While the L130 conversion to proline is predicted to dramatically affects the secondary structure of the helical cap, L53 alters the secondary structure of the 3'-flap pocket and its ability to accommodate the 3'-flap.
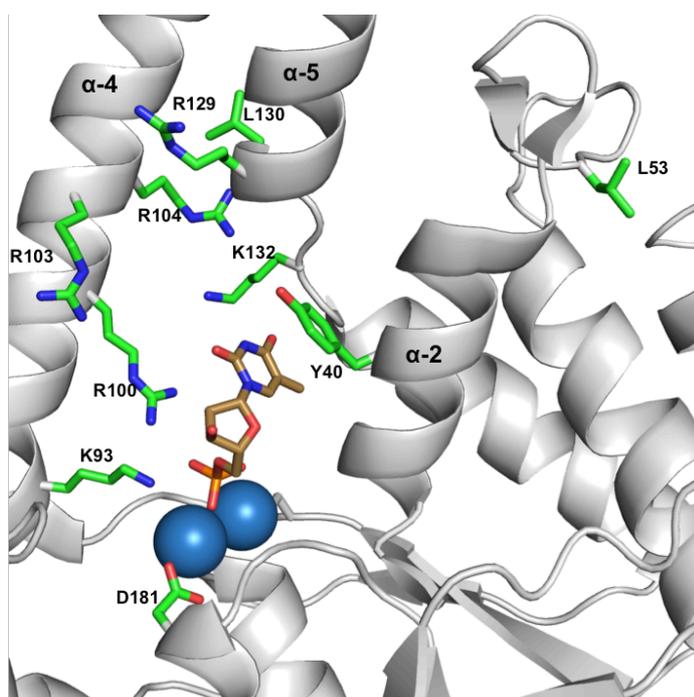
Figure 4.7. *Cartoon represents the active site of hFEN-1:product structure to show the -1 nt in contact with* **metal ions** *active site and the studied mutated residues from the helical gateway (base α2-α4), cap (top of α4 and α5), active site and 3'-flap pocket. The* **oxygen** *and* **nitrogen** *atoms of the side chains were shown.*

4.2.1. Watson-Crick base pair, 3'-flap and 5'-phosphate monoester are required for DNA bending upon WThFEN-1 binding:

Several different DNAs with the potential to interact with FEN-1 were studied to elucidate the important features for interaction between DNA and protein *(figure 4.6)*. All of the E_{min} values of the protein-free substrates (DF, $MM_{+1}DF$, 3'-SF, 5'OH3'-SF, 5'-SF, nicked DNA, and SA-DF) were similar $\pm Ca^{2+}$ and greater than duplex, except the E_{min} value of the nicked DNA was lower with ~2-fold than duplex *(figure 4.9 and figure 4.10.a)*. The calculated maximum distances (without protein) between donor and acceptor pair using *equation 3.5* were in line with the above E_{min} comparisons *(figure 4.10.b)*. All of the free substrates showed smaller distance (less than 67 Å) than the DNA duplex (80 Å), except the nicked DNA that showed bigger distance (91 Å). This suggests that all of the substrates except nicked DNA have overall conformations that are more

bent than duplex DNA, even before addition of protein. The exception was nicked DNA that has a FRET value that was less than DNA duplex. Indeed, the nicked DNA result is a bit unexpected and since it implies that the dyes become further apart than in duplex DNA. One possible explanation is that the duplexes of the nicked DNA rotate with respect to one another as well as bending and this rotation may actually increase the net distance between the dyes, however, the DNA duplex cannot do this *(figure 4.8)*. The exo substrates 3'-SF and 5'-SF showed slightly lower E_{min} ±Ca^{2+} than DF substrate *(figure 4.9.a.c.e and figure 4.10.a)*, which suggests that the absence of one flap may reduce the overall bending slightly (has bigger net bending angle). In contrast, SA-DF had higher E_{min} in Ca^{2+} than DF substrate *(figure 4.9.a and figure 4.10.a)*, which suggests that more bent overall conformation than uncomplexed DNA, likely due to the presence of a bulky streptavidin (SA) homotetramer conjugated to the 5'-terminus.
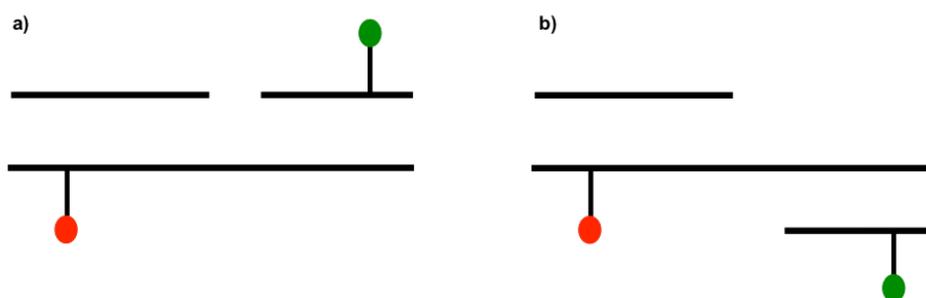
Figure 4.8. *Cartoon representation of the proposed rotation of the nicked DNA duplexes that can effect the interaction between the two dyes* **donor** *and* **acceptor**. ***a)*** *Non-rotated nicked DNA.* ***b)*** *Rotated nicked DNA so the dyes become further.*

Energy transfer efficiency (E) of the donor and acceptor labelled (DAL) substrates ±Ca^{2+} increased gradually upon increasing WThFEN-1 concentration arriving to a saturated situation where addition of more enzyme produced no further change *(figure 4.9)*. For unknown reasons, the 3'-SF and 5'OH3'-SF substrates showed slightly higher E_{max} (around 10 %) than DF substrate ±Ca^{2+}. However, $MM_{+1}DF$ which as a mismatch at the duplex junction 5'-terminus and 5'-SF substrates showed 20 % decrease and nick DNA showed 60 % decrease of the E_{max} values compared to DF substrate ±Ca^{2+}. The small differences between E_{max} values of the different substrates reflected the effects on the calculated r_{min} (minimum distance at saturating protein) between donor and acceptor pair using *equation 3.5 (figure 4.10.b)*. While 3'-SF and 5'OH3'-SF substrates had lesser r_{min} the other substrates had larger r_{min}. Each substrate showed slightly higher E_{max} values that were reproducibly observed with Ca^{2+} ions present than the corresponding substrate signal in EDTA *(figure 4.10.a)*. The main cause of the endpoint variance at ±Ca^{2+} is unknown.

The derived K_{bend} values in Ca^{2+} were generally similar for DF and 3'-SF (no 5'-flap) substrates. However, compared to the DF substrate, $MM_{+1}DF$ (mismatch), SA-DF (5'-streptavidin) and nicked DNA increased K_{bend} values by around 5-fold, the 5'-SF (no 3'-flap) by around 10-fold and the 5'OH3'-SF by around 20-fold. In EDTA, all substrates changed the K_{bend} values by 2-factor. Except for DF, $MM_{+1}DF$ and 3'-SF, interestingly the other substrates stabilized the WThFEN-1:DNA complex in the presence of EDTA *(figure 4.10.c) (table 4.3)*. As the substrate cannot position in the active site when metal ions are absent, these differences $\pm Ca^{2+}$ suggest that some of the alterations are much more destabilising to this metal ion positioned state than when the DNA is not positioned for reaction (EDTA) state.

Hence, WThFEN-1 has ability to bend all substrates with and without metal ions present. In this case, substrates are bent upon WThFEN-1 binding regardless either of whether the features that need metal ions presence such as 5'-flap threading[51] and scissile phosphate bond position[37] have taken place. In addition, while the 5'-flap presence or its accommodation underneath the helical cap are not required for global substrate bending, which is in line with the previous result of DF and 3'-SF bending upon mFEN-1 binding in *section 3.1.4*, a Watson-Crick base pair of +1 nt, a 3'-flap and a 5'-phosphate monoester or diester at the terminus of the reacting duplex are critical for efficient substrate bending. Furthermore, the absence of both the flaps together (nicked DNA), reduces the enzyme ability to bend the substrate significantly.
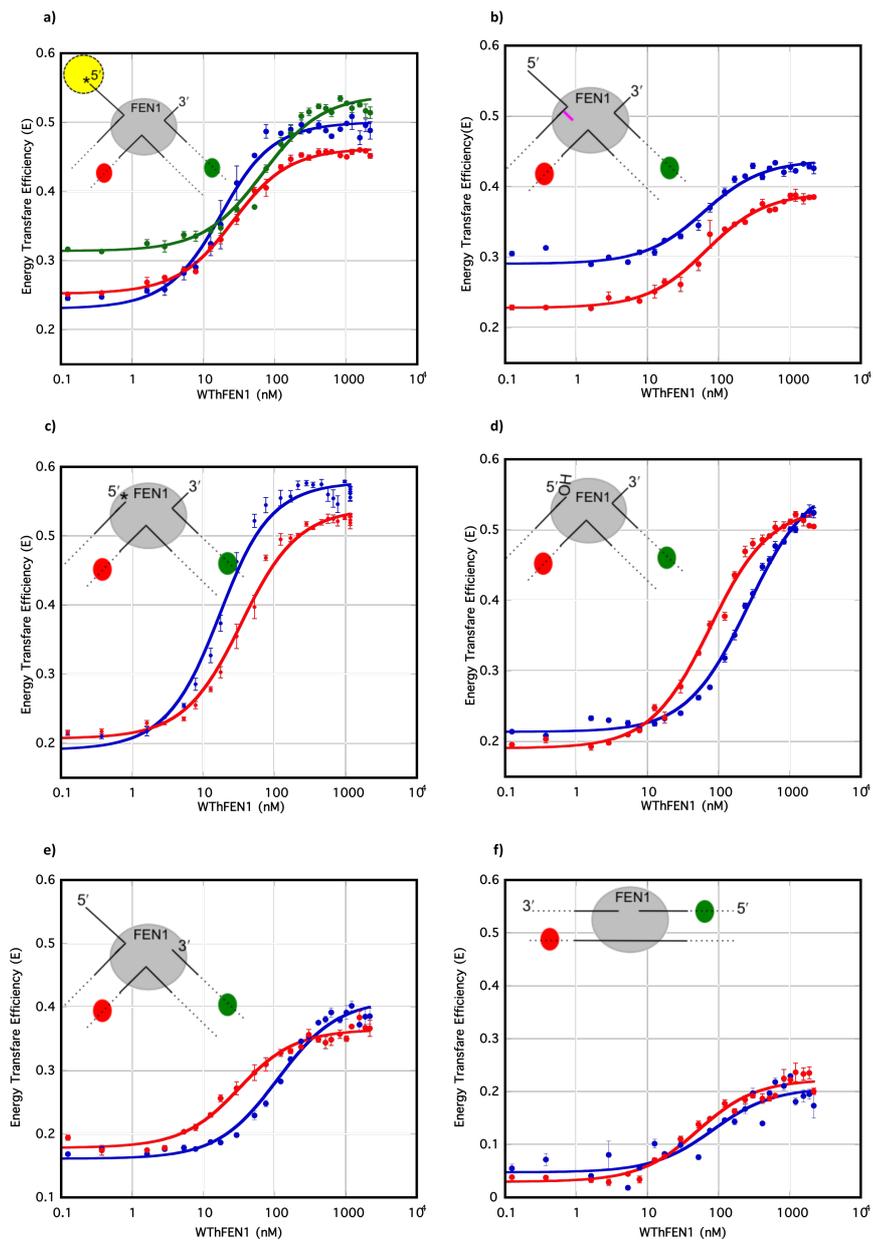
Figure 4.9. FRET curves of WThFEN-1 in Ca^{2+} and **EDTA** upon binding to **a)** DF and **SA-DF in Ca^{2+}**, **b)** $MM_{+1}DF$, **c)** 3'-SF, **d)** 5'OH3'-SF, **e)** 5'-SF, and **f)** Nick DNA. The data points were fit to equation 3.9. Each DAL substrate structure is shown and contains **donor** and **acceptor**. The bulky streptavidin showed in yellow ball connected to the 5'-flap of the DF substrate.
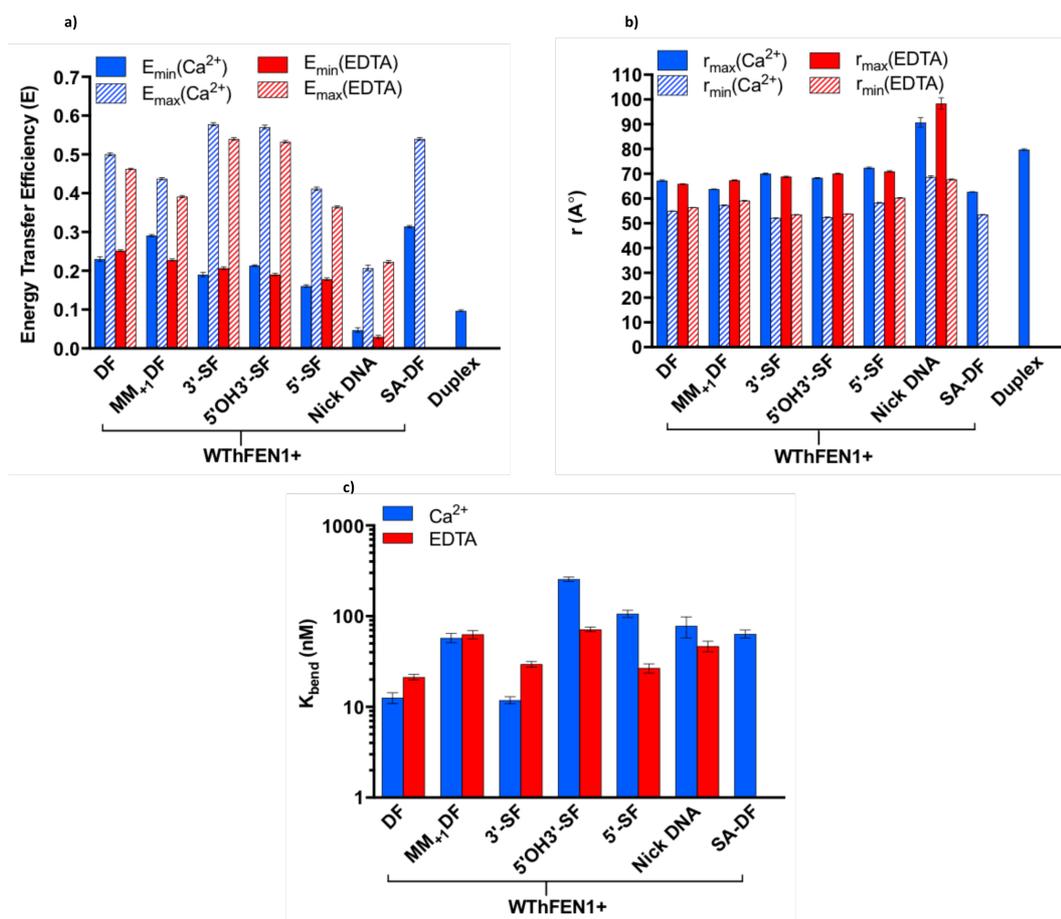
4.10. *FRET data of WThFEN-1 binding different substrates and free DNA duplex in* **Ca^{2+}** *and* **EDTA.** *a) The measured* E_{min} *and* E_{max}. *b) The calculated* r_{max} *and* r_{min} *using equation 3.5.* *c) The observed* K_{bend} *values.*

Substrate	DF	MM$_{+1}$DF	3'-SF	5'OH3'-SF	5'-SF	Nick DNA	SA-DF
K_{bend} (nM) (Ca^{2+})	13 ± 1	58 ± 7	12 ± 1	256 ± 13	107 ± 10	78 ± 20	64 ± 6
K_{bend} (nM) (EDTA)	21 ± 1	63 ± 7	30 ± 2	72 ± 4	27 ± 3	47 ± 6	-

Table 4.3. K_{bend} *parameters for WThFEN-1 bound different substrates that were determined by FRET.*

4.2.2. Effects of helical arch mutations and active site mutation on DF and 3'-SF bending:

To investigate the requirements and importance of DF and 3'-SF substrates bending upon hFEN-1 binding, helical arch and active site mutations were tested using FRET. Besides the helical arch mutations Y40A, K93A, R100A, L130P, and K93AR100A and active site mutation D181A that were studied using ECCD to indicate their effects on the local substrate conformational change *(section 4.1.3)*, other helical arch mutations such as R103AR129, R103ER129E, R104AK132A,

121

R104EK132E and R103ER104ER129EK132E (Quad E) were studied using FRET to indicate their effects on DF substrate bending. With 3'-SF substrate only R104A, K132A, R103AR129 and R104AK132A mutations were studied to indicate their effects. These helical arch residues were mutated at positions postulated to guide the phosphodiester backbone and stabilize the inverted ssDNA orientation when the flap is threaded through the helical arch and positioned to react *(figure 4.7)*.

DF (DAL) and 3'-SF (DAL) substrates adopted a bent conformation when bound to the mutated proteins as seen by an increase in FRET signal upon addition of hFEN-1 *(figure 4.11)*. Like the WT enzyme, differences between E_{max} values $\pm Ca^{2+}$ were observed with all proteins with titrations in Ca^{2+} buffer producing a higher value, except K93A and D181A where the end points remained constant $\pm Ca^{2+}$ *(figure 4.11 and figure 4.12.a.d)*. In addition, increasing enzyme concentration decreased the distance between donor and acceptor pair of both substrates, which was calculated using *equation 3.5*, to be similar $\pm Ca^{2+}$ r_{min} of WT, except R104EK132E and Quad E mutations (with DF substrate) that showed slightly larger r_{max} and r_{min} values *(figure 4.12.b.e)*. As with WThFEN-1, only subtle variations in K_{bend} were observed with and without divalent metal ions (2-fold at most) for these two substrates *(figure 4.12.c.f) (table 4.4)*. The exception was Y40A (with DF) and R104A (with3'-SF) where mutation stabilized the hFEN-1:DNA complex in the presence of EDTA. Only small changes in K_{bend} were observed relative to the WT protein $\pm Ca^{2+}$ (less than 3-fold at most), except for the repulsive R104EK132E and Quad E mutations (with DF) that showed 5- and 17-fold increase, respectively. This suggests that all mutated proteins do not substantially affect binding and bending of the DF and 3'-SF substrates, thus only the R104EK132E and Quad E mutated proteins are critical to DF substrate binding and bending and may be by repelling the ss5'-flap. To be in agreement with the WThFEN-1 FRET results *(section 4.2.1)* and the previous result, which suggested that 5'-flap threading is not required for substrate bending,[51] the flap may be directed by these residues but without threading. According to all of these FRET results, deficient substrate positioning, not poor binding, is the major contributing factor to diminished protein activity to interact with DF or 3'-SF substrates under single turnover conditions *(table 4.1)*, except R104EK132E and Quad E proteins that caused imperfect DF substrate bending. However, comparing the relative increases of the K_{bend} values of both R104EK132E and Quad E proteins to their severe impact on activity under single turnover conditions with 4000- and 18,000-fold decreased reaction rates respectively shows that the enzymes were folded and capable of substrate binding. This large rate decrease is remarkable for mutated residues not acting in catalysis and distance from the active site.
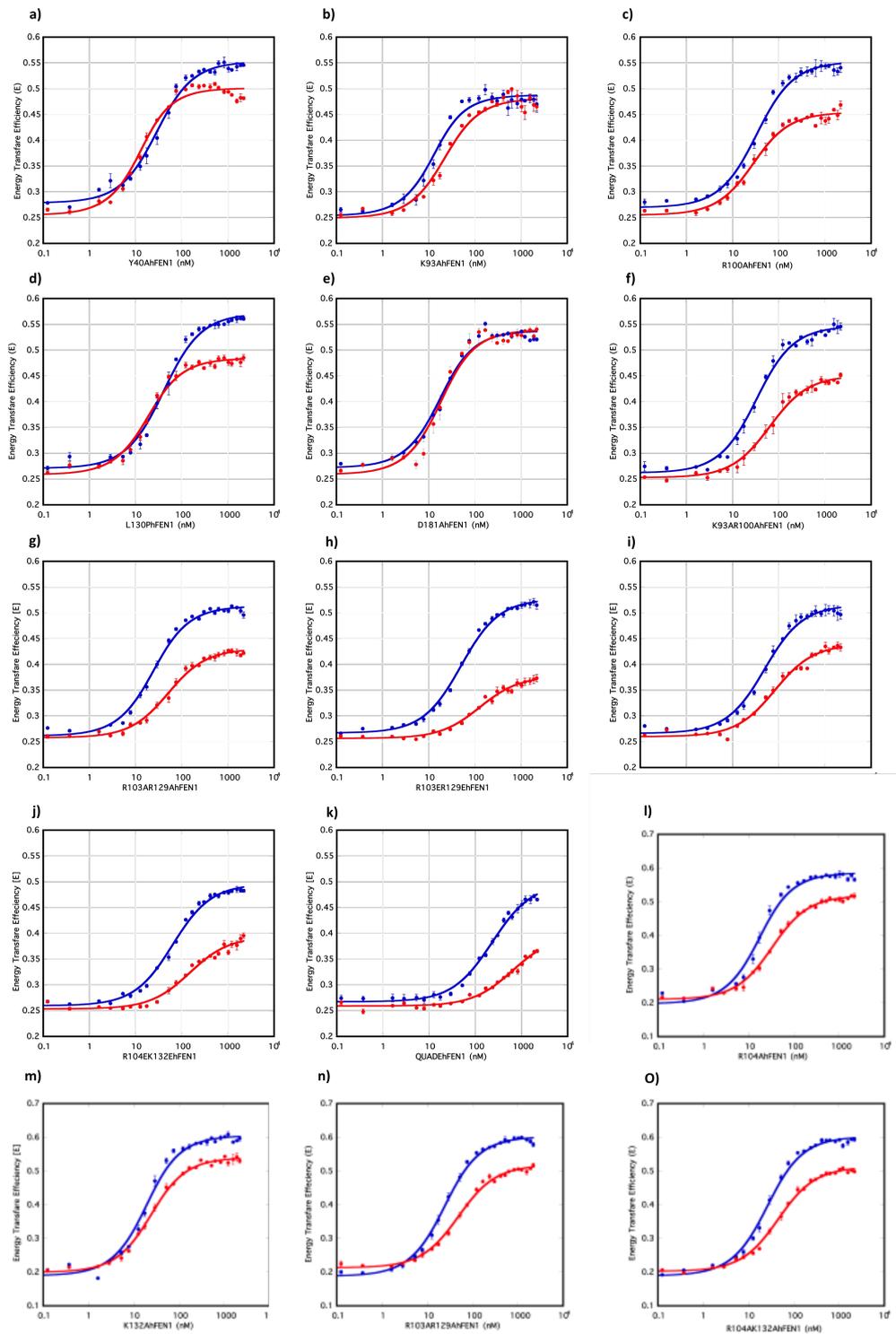
Figure 4.11. FRET curves of DF and 3'-SF substrates in **Ca^{2+}** and **EDTA** upon binding to mutant hFEN-1. **a)** Y40A with DF. **b)** K93A with DF. **c)** R100A with DF. **d)** L130P with DF. **e)** D181A with DF. **f)** K93AR100A with DF. **g)** R103AR129A with DF. **h)** R103ER129E with DF. **i)** R104AK132A with DF. **j)** R104EK132E with DF. **k)** Quad E (R103ER104ER129EK132E) with DF. L) R104A with 3'-SF. m) K132A with 3'-SF. **n)** R103AR129A with 3'-SF. **o)** R104AK132A with 3'-SF. The data points were fit to equation 3.9.
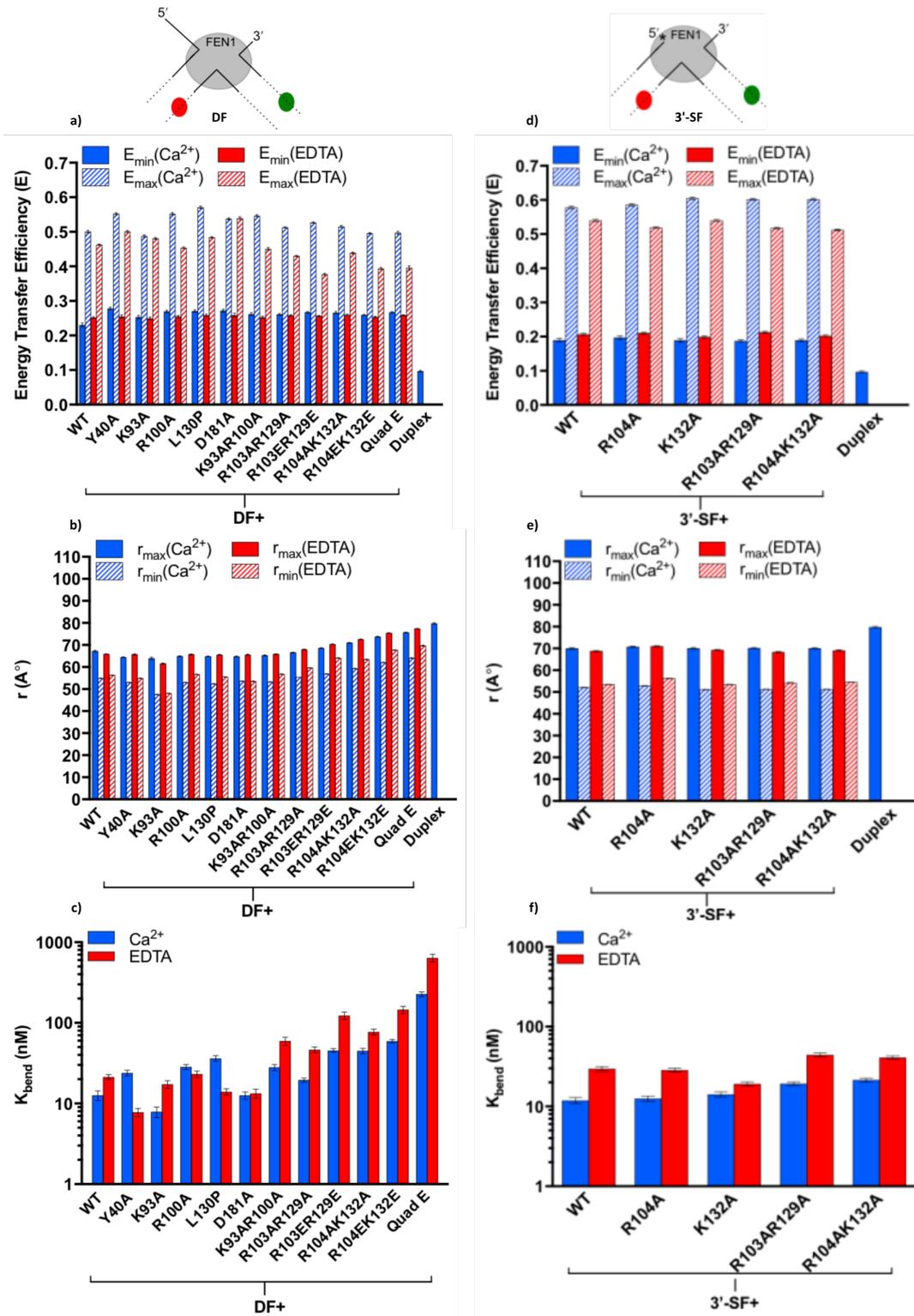
Figure 4.12. *FRET data of WT and mutant hFEN-1 binding to DF and 3'-SF substrates compared to free DNA duplex in* **Ca²⁺** *and* **EDTA**. ***a)*** *The measured E_{min} and E_{max} of DF.* ***b)*** *The calculated r_{max} and r_{min} using equation 3.5 of DF.* ***c)*** *The observed K_{bend} values of DF.* ***d)*** *The measured E_{min} and E_{max} of 3'-SF.* ***e)*** *The calculated r_{max} and r_{min} using equation 3.5 of 3'-SF.* ***f)*** *The observed K_{bend} values of 3'-SF. DF (DAL) and 3'-SF (DAL) structures are shown and contain* **donor** *and* **acceptor**.

hFEN-1	K_{bend} (nM) of DF		K_{bend} (nM) of 3'-SF	
	Ca^{2+}	EDTA	Ca^{2+}	EDTA
WT	13 ± 1	21 ± 1	12 ± 1	30 ± 2
Y40A	24 ± 2	8 ± 1	-	-
K93A	8 ± 1	17 ± 2	-	-
R100A	28 ± 2	23 ± 2	-	-
R104A	-	-	24 ± 2	8 ± 1
L130P	36 ± 3	14 ± 1	-	-
K132A	-	-	8 ± 1	17 ± 2
D181A	13 ± 1	13 ± 2	-	-
K93AR100A	28 ± 2	59 ± 7	-	-
R103AR129A	20 ± 1	46 ± 4	28 ± 2	23 ± 2
R103ER129E	45 ± 2	123 ± 13	-	-
R104AK132A	45 ± 4	77 ± 6	36 ± 3	14 ± 1
R104EK132E	59 ± 3	146 ± 15	-	-
Quad E	227 ± 16	638 ± 70	-	-

Table 4.4. K_{bend} *parameters for WT and mutant hFEN-1 bind DF and 3'-SF substrates that were determined by FRET.*

4.2.3. Presence and positioning of the 1nt 3'-flap are required for substrate bending:

An L53A 3'-flap pocket mutation was also studied with DF substrate to investigate any roles for binding of the 1nt 3'-flap during the substrate binding and bending. The results were compared to the previous FRET results of the WThFEN-1 with DF and 5'-SF substrates, which were discussed before in *section 4.2.1*.

DF (DAL) substrate adopted a bent conformation when bound to L53AhFEN-1 protein as shown by an increase in FRET signal upon addition of protein *(figure 4.11.a)*. Unlike the WT enzyme, where differences between E_{max} values ±Ca^{2+} were observed with DF and 5'-SF substrates with titrations in Ca^{2+} buffer producing a higher value, end points of L53A remained constant ±Ca^{2+} *(figure 4.11.a.b)*. Increasing enzyme concentration decreased the distance between donor and acceptor pair of both substrates, which was calculated using *equation 3.5*, to be similar ±Ca^{2+} r_{min}, however, the r_{min} of the DF:L53A complex was slightly less *(figure 4.11.c)*. While, WThFEN-1 showed only subtle variations in ±Ca^{2+} K_{bend} values with both substrates, L53A (with DF) stabilized its complex in EDTA *(figure 4.11.d) (table 4.5)*. Relative to WT:DF complex, WT:5'-SF and L53A:DF complexes showed 8- and 6-fold increase of the K_{bend}, respectively. This suggests that

absence of the 1nt 3'-flap or its imperfect contact with the enzyme caused by 3'-flap pocket mutations such as L53A mutation are important to substrate binding and bending. This may in part explain decreases in rate of the L53A protein catalysing reaction of the DF substrate, although it is clear that other factors are at play to produce rate decreases under enzyme saturating conditions single turnover conditions *(table 4.1)*.
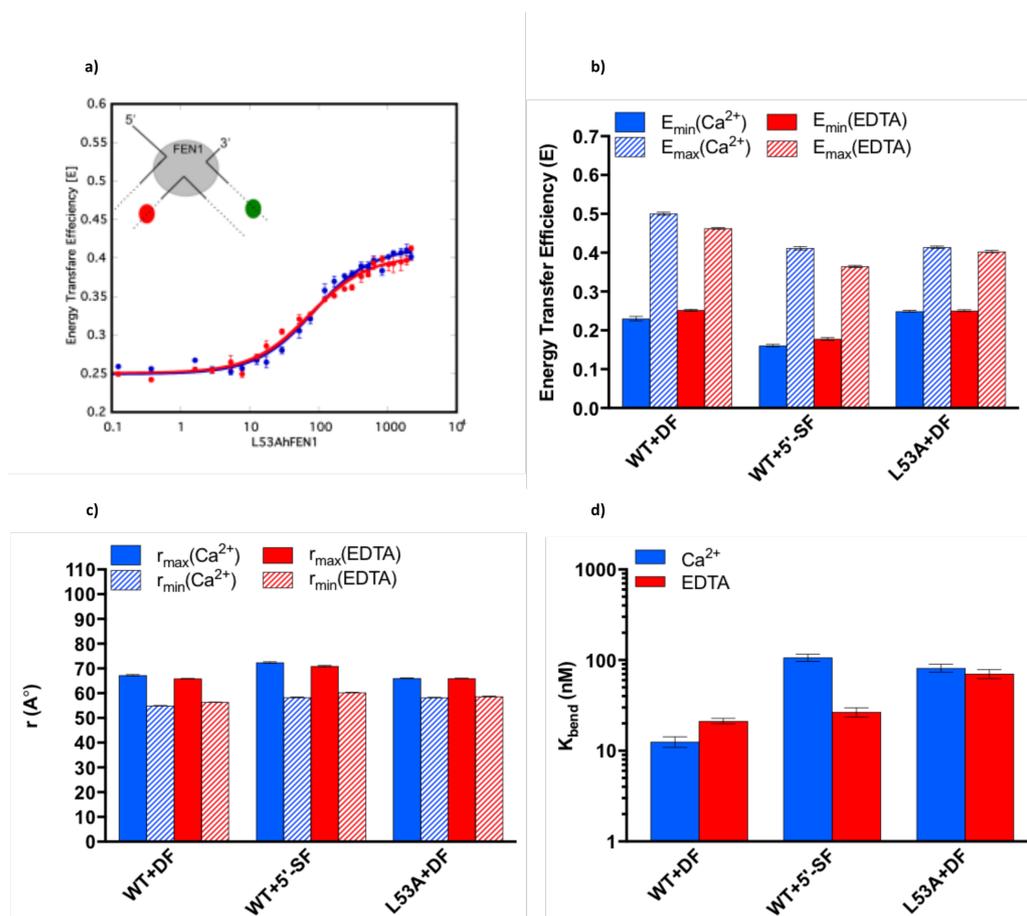
Figure 4.11. *FRET data of WT binding to DF and 5'-SF substrates and L53A hFEN-1 binding to DF substrate in* **Ca^{2+}** *and* **EDTA**. ***a)*** *FRET curve of DF upon binding to L53AhFEN-1.* ***b)*** *The measured E_{min} and E_{max}.* ***c)*** *The calculated r_{max} and r_{min} using equation 3.5.* ***d)*** *The observed K_{bend} values. DF (DAL) structure is shown and contain* **donor** *and* **acceptor**.

Substrate	K_{bend} (nM) (Ca^{2+})	K_{bend} (nM) (EDTA)
WT with DF	13 ± 1	21 ± 1
WT with 5'-SF	107 ± 10	27 ± 3
L53A with DF	82 ± 8	71 ±

Table 4.5. *K_{bend} parameters for WThFEN-1 and L53A bind DF and 5'-SF substrates that were determined by FRET.*

4.2.4. Summary and conclusion of the substrate bending upon WT and mutant hFEN-1:

To conclude, the substrates are bent upon hFEN-1 binding $\pm Ca^{2+}$, which means that substrate bending does not require the features that need metal ions present such as 5'-flap accommodation[51] and scissile phosphate bond position on active site metal ions[37] to take place. WThFEN-1 bends DF and 3'-SF with similar efficiency, however, this bending efficiency reduces without DNA base pairing at the terminus of the reacting duplex, a 5'-phosphate monoester, a 3'-flap or both flaps. The roles of both flaps are also confirmed by mutated enzymes results. First, the non-requirement of the 5'-flap is confirmed by the helical arch mutations results. Most of the mutated helical arch residues do not reduce enzyme ability to bend DF and 3'-SF substrates and this in agreement with the WThFEN-1 results about the non-requirement of the 5'-flap threading for substrate binding. However, reducing of DF substrate bending by some mutated cap residues such as R104EK132E and Quad E may suggests that alteration of these cap residues are detrimental to substrate binding and bending by repelling the ss5'-flap. To be in agreement with the previous result, this interaction does not mean threading but only interaction to direct the substrate. Second, the importance of the 3'-flap is confirmed by the 3'-flap pocket mutation results. Lacking of this flap or its imperfect contact with the enzyme reduces the enzyme ability to bend the substrate efficiently.

4.3. Summary and conclusion of hFEN-1 (ECCD and FRET):

To conclude, as the local DNA conformational changes and substrate bending are important steps required before FEN-1 catalysis, any imperfect effect on one of these can lead to deficient enzyme efficiency. According to ECCD and FRET results, both processes require different conditions. While the local DNA conformational change requires presence of the divalent metal ions, the substrates are bent upon hFEN-1 binding $\pm Ca^{2+}$, which means that substrate bending does not require the features that need metal ions presence such as 5'-flap accommodation[51] and scissile phosphate bond position[37] to take place. Consistent with ECCD and FRET WThFEN-1 results with 3'-SF and 5'OH3'-SF substrates, both processes require the presence of the 5'-phosphate monoester to be achieved ideally. Further requirements for efficient substrate bending include DNA base pairing and 3'-flap or the presence of both flaps.

Most of the studied helical arch mutations such as Y40A, R100A, D181A and K93AR100A reduce hFEN-1's ability to produce an ideal local DNA conformational change with 3'-SF and 5'OH3'-SF substrates. K93A does not impact the local DNA structural change, while L130P prevents this substrate rearrangement totally. Importantly, all of these mutated helical arch residues (Y40A, R100A, L130P, D181A and K93AR100A) do not reduce enzyme ability to bend DF and 3'-SF sub-

strates and therefore local conformational change deficiencies observed in ECCD are not the consequence of inability to bind substrate. Accordingly, deficient substrate positioning, but not poor bending, is likely to be the main cause of reduced hFEN-1 activity under single turnover conditions when interacting with DF substrate except K93A where the cause may relate to the chemistry *(table 4.1)*.

The other helical arch mutations that were only studied by FRET (R104A, K132A, R103AR129A, R103ER129E, R104AK132A, R104EK132E and Quad E) also do not affect the substrate bending substantially except R104EK132E and Quad E mutations that reduces enzyme ability to bend DF substrate but not 3'-SF substrate, which may suggest that these cap residues are primarily required for ss5'-flap steering. In this case, poor bending is not the factor that leads to reduction of the enzyme's efficiency in single turnover assays interacting with DF or 3'-SF substrates. Although R104EK132E and Quad E mutations reduce bending ability of hFEN-1, there is a huge effect on rate reaction as shown in *table 4.1* leading to the suggestion that more effects such as imperfect conformational changes or chemistry are at play.

Chapter 5: Conclusion:

Overall, this study unravels the interrelationships between events in the FEN-1 catalytic cycle. Studies of the mFEN-1 enzyme were undertaken to find out the molecular basis underlying the observation that mutations of FEN-1 result in cancers in mice. Combined studies of the mFEN-1 enzyme show that it appears to recognise different DNA structures within one active site. Although the 5'-product of reaction differs with each substrate, it seems likely that the same chemistry is used to hydrolyse the target the equivalent phosphodiester bond to release a product that corresponds to hydrolysis 1nt into downstream duplex of the 5'-terminus. The enzyme recognises different structures of DNA that have the common features of a junction between two duplexes with 1nt 3'-flap. Although, the enzyme has ability to recognise the 3'-SF substrate that lacks a 5'-flap, its catalytic efficiency is slightly enhanced with the double flap substrates such as DF, gap and DF2'OH. Indeed, in the absence of the 5'-flap the enzyme ability to bind and bend the substrate and position the substrate perfectly are retained. Therefore, the 5'-flap is not a pre-requisite for efficient FEN-1 reaction. In contrast, it was demonstrated that the combined absence of the 5'-flap and the 5'-terminal phosphate significantly reduces the enzyme's ability to bend and position the substrate. This explains how the absence of the 5'-phosphate further decreases the rate of reaction of an exonucleolytic FEN-1 substrate and suggests a key role for interactions with this part of the substrate in DNA binding and correct positioning for reaction. Furthermore, it was also shown that although active site metal ions are required for enzyme chemistry and for substrate positioning for reaction, they are not required for initial interaction with the substrate including bending.

Significant impacts on the rate of reactions were caused by mutated enzymes that have been implicated in cancer. The active site mutations A159V and E160D of mFEN-1 reduce the enzyme's ability to catalyse the reactions of all of the substrates tested (DF, gap and DF2'OH). This implies that it is unlikely that genome instability that results in cancer is the result of altering or abolishing the ability of FEN-1 to act on one particular substrate. However, the mutations themselves have different outcomes with A159V producing greater reductions in the rate of reaction than E160D. Using a DF2'OH substrate, where the leaving group nucleotide is ribose mimicking the substrate that would be encountered during RNA primer removal, the impact of both mutations is decreased particularly with A159V. Although both mutations produce defective enzyme chemistry, and its ability to produce ideal conformational changes and substrate bending, they affect differently depending on the position and the nature of the mutated residue.

To extend studies of the important requirements of both enzyme and substrate for efficient FEN-1 reaction, hFEN-1 was also studied. Similar to mFEN-1, the substrate structure and some mutated residues of hFEN-1 affect the enzyme's ability to bend and position the substrate perfectly that is necessary for catalysis. The presence of the 5'-phosphate monoester of the 3'-SF substrate is important for hFEN-1 to bend and position the substrate perfectly confirming the previous results with mFEN-1. Furthermore, the absence of the 3'-flap alone (5'-SF substrate) reduces hFEN-1's ability to bend the substrate perfectly, and similar reduction occurs when absence of the 3'-flap combines with absence of the 5'-flap in the product of the reaction nicked DNA. Additionally effects of substrate structure, many of hFEN-1 mutations affect the enzyme's ability to bend or position the substrate ideally beside their effect on the enzyme chemistry that was proven in previous studies. The active site mutation D181A and the helical arch mutations Y40A, K93A, R100A, L130P and K93AR100A hinder the enzyme's capacity to produce an ideal local DNA conformational change required to position the specific target phosphate bond into the active site. The exceptions to this were K93A, which did not impact the local DNA structural change, and L130P, which prevented this substrate rearrangement totally. Preventing the perfect local DNA conformational changes because of the human D181A active site mutation is in agreement with mFEN-1 active site mutations A159V and E160D effects. On the other hand, all of these mutated residues had little effect on the DF substrate binding and bending.

The impact of other helical arch mutations R104A, K132A, R103AR129A, R103ER129E, R104AK132A, R104EK132E and Quad E and 3'-flap pocket mutation L53A on substrate binding and bending were also studied. Each of these mutations was already known to alter the rate of the hFEN-1 reaction but to differing extents. With the exception of R104EK132E, Quad E and L53A mutations that all reduced the enzyme ability to bend the DF substrate perfectly, other helical arch mutations had small impacts on substrate dissociation constants.

During the course of these our collaboration with John Tainer's lab resulted in a crystal structure of D86N hFEN-1 bound to a threaded substrate.[93] In this structure the DNA substrate appeared to be positioned to react, with the scissile phosphodiester bond in contact with active site metal ions. The 5'-flap DNA was threaded through the helical arch, with the +1 phosphate of the flap surrounded by basic residues, notably Arg 104 and Lys 132 whereas Arg 100 is positioned to stabilise both the positioning of the scissile phosphodiester bond and the +1 phosphodiester. The functional significance of these observations is demonstrated by the results described here such as the lack of ability of R100A to bring about DNA conformational change. Interestingly, although removal of one of these basic residues alone does not drastically alter the ability of FEN-1 to bind and bend the substrates, clearly removal of the +1 5'-phosphate does. Similarly,

creating a repulsive force for the +1 5'-phosphate with R104EK132E and Quad E was detrimental to enzyme-substrate stability. In addition, it is shown here that removal of the +1 5'-phosphate prevents proper positioning of the DNA substrate as might be predicted on the basis of this structure.

As the DNA had moved towards its active site position it twisted towards α5 of the helical cap considerably shortening the distance between Arg 129 and other basic residues and the template DNA of the substrate. Here again removal of Arg 129 by itself did not appear to impact on substrate dissociation constant dramatically, but it would be of interest to test the impact of mutation of this residue on the ability to position substrate for reaction by ECCD in the future. Furthermore, the results obtained here with the helical cap destructive mutation α5 L130P, where the positioning of all basic residues could well be altered, do suggest a role for α5 in substrate positioning for reaction. Finally the +1 nucleobase at end of the DNA duplex was stacked upon Tyr 40, which had flipped towards the active site. This is in accord with the result shown here, where Y40A cannot bring about a perfect local DNA conformational change.

Chapter 6: References:

(1)   MacNeill, S.: *The Eukaryotic Replisome: A Guide to Protein Structure and Function.*; Springer, 2012. pp. 301-326.

(2)   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*, 74-84.

(3)   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*, 433-437.

(4)   Lieber, M. R. The FEN-1 family of structure-specific nucleases in eukaryotic DNA replication, recombination and repair. *BioEssays* **1997**, *19*, 233-240.

(5)   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.; Cooper, P. K.; Grasby, J. A.; Tainer, J. A. Human Flap Endonuclease Structures, DNA Double-Base Flipping, and a Unified Understanding of the FEN1 Superfamily. *Cell (Cambridge, MA, U. S.)* **2011**, *145*, 198-211.

(6)   Alberts, B.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P.; Editors: *Molecular Biology of The Cell, Fourth Edition*; Garland Science, 2002.

(7)   Griffiths, A. J. F.: *An Introduction to Genetic Analysis, 6th Edition*; Freeman, 1996.

(8)   Watson, J. D.; Crick, F. H. C. Molecular structure of nucleic acids. A structure for deoxyribose nucleic acid. *Nature* **1953**, *171*, 737-738.

(9)   von Hippel, P. H. From "simple" DNA-protein interactions to the macromolecular machines of gene expression. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 79-105.

(10)   Nakamura, T.; Zhao, Y.; Yamagata, Y.; Hua, Y.-j.; Yang, W. Watching DNA polymerase η make a phosphodiester bond. *Nature (London, U. K.)* **2012**, *487*, 196-201.

(11)   Parker, D. J.; Xiao, Y.; Aguilar, J. M.; Silverman, S. K. DNA Catalysis of a Normally Disfavored RNA Hydrolysis Reaction. *J. Am. Chem. Soc.* **2013**, *135*, 8472-8475.

(12)   Yang, W. Nucleases: diversity of structure, function and mechanism. *Q Rev Biophys* **2011**, *44*, 1-93.

(13)   Rice, P. A. C., Carl C.: *Protein-Nucleic Acid Interactions : Structural Biology*; Royal Society of Chemistry, 2008. pp. 333-366.

(14)   Jencks, W. P. General acid-base catalysis of complex reactions in water. *Chem. Rev.* **1972**, *72*, 705-718.

(15)   More O'Ferrall, R. A. Relations between E2 and ElcB mechanisms of β-elimination. *J. Chem. Soc. B* **1970**, 274-277.

(16)   Westheimer, F. H. Pseudo-rotation in the hydrolysis of phosphate esters. *Accounts Chem. Res.* **1968**, *1*, 70-78.

(17)   Kirby, A. J.; Jencks, W. P. Reactivity of nucleophilic reagents toward the p-nitrophenyl phosphate dianion. *J. Am. Chem. Soc.* **1965**, *87*, 3209-3216.

(18)   Zheng, L.; Dai, H.; Zhou, M.; Li, M.; Singh, P.; Qiu, J.; Tsark, W.; Huang, Q.; Kernstine, K.; Zhang, X.; Lin, D.; Shen, B. Fen1 mutations result in autoimmunity, chronic inflammation and cancers. *Nat. Med.* **2007**, *13*, 812-819.

(19)   Larsen, E.; Kleppa, L.; Meza, T. J.; Meza-Zepeda, L. A.; Rada, C.; Castellanos, C. G.; Lien, G. F.; Nesse, G. J.; Neuberger, M. S.; Laerdahl, J. K.; Doughty, R. W.; Klungland, A. Early-Onset Lymphoma and Extensive Embryonic Apoptosis in Two Domain-Specific Fen1 Mice Mutants. *Cancer Res.* **2008**, *68*, 4571-4579.

(20)   Alberts, B. DNA replication and recombination. *Nature* **2003**, *421*, 431-435.

(21)   O'Donnell, M.; Langston, L.; Stillman, B. Principles and concepts of DNA replication in bacteria, archaea, and eukarya. *Cold Spring Harb Perspect Biol* **2013**, *5*.

(22)   Burgers, P. M. J. Polymerase dynamics at the eukaryotic DNA replication fork. *J. Biol. Chem.* **2009**, *284*, 4041-4045.

(23)   Okazaki, R.; Okazaki, T.; Sakabe, K.; Sugimoto, K. Mechanism of DNA replication possible discontinuity of DNA chain growth. *Jpn J Med Sci Biol* **1967**, *20*, 255-260.

(24)   Ogawa, T.; Okazaki, T. Discontinuous DNA replication. *Annu. Rev. Biochem.* **1980**, *49*, 421-457.

(25)   MacNeill, S. A. DNA replication: Partners in the Okazaki two-step. *Curr. Biol.* **2001**, *11*, R842-R844.

(26)   Chen, C.; Editor: *New Research Directions in DNA Repair*; InTech, 2013.

(27)   Schermerhorn, K. M.; Delaney, S. A chemical and kinetic perspective on base excision repair of DNA. *Acc. Chem. Res.* **2014**, *47*, 1238-1246.

(28)   Dianov, G.; Lindahl, T. Reconstitution of the DNA base excision-repair pathway. *Curr. Biol.* **1994**, *4*, 1069-1076.

(29)   Orans, J.; McSweeney, E. A.; Iyer, R. R.; Hast, M. A.; Hellinga, H. W.; Modrich, P.; Beese, L. S. Structures of Human Exonuclease 1 DNA Complexes Suggest a Unified Mechanism for Nuclease Family. *Cell (Cambridge, MA, U. S.)* **2011**, *145*, 212-223.

(30)   Lee, S.-H.; Klugel, M. F.; Biertumpfel, C.; Princz, L. N.; Pfander, B.; Habermann, B. Human Holliday junction resolvase GEN1 uses a chromodomain for efficient DNA recognition and cleavage. *Elife* **2015**, *4*.

(31)   Koch, S. C.; Simon, N.; Ebert, C.; Carell, T. Molecular mechanisms of xeroderma pigmentosum (XP) proteins. *Q Rev Biophys* **2016**, *49*, e5.

(32)   Keijzers, G.; Liu, D.; Rasmussen, L. J. Exonuclease 1 and its versatile roles in DNA repair. *Crit. Rev. Biochem. Mol. Biol.* **2016**, *51*, 440-451.

(33)   Liu, Y.; Freeman, A. D. J.; Declais, 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*, 2565-2575.

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

(35)   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*, 10762-10775.

(36)   Hosfield, D. J.; Frank, G.; Weng, Y.; Tainer, J. A.; Shen, B. Newly discovered archaebacterial flap endonucleases show a structure-specific mechanism for DNA substrate binding and catalysis resembling human flap endonuclease-1. *J. Biol. Chem.* **1998**, *273*, 27154-27161.

(37)   David Finger, L.; 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*, 9839-9847.

(38)   Shen, B.; Qiu, J.; Hosfield, D.; Tainer, J. A. Flap endonuclease homologs in archaebacteria exist as independent proteins. *Trends Biochem. Sci.* **1998**, *23*, 171-173.

(39)   Nazarkina, Z. K.; Lavrik, O. I.; Khodyreva, S. N. Flap endonuclease-1 and its role in the processes of DNA metabolism in eucaryotic cells. *Mol Biol* **2008**, *42*, 405-421.

(40)   Brosh, R. M., Jr.; Von Kobbe, C.; Sommers, J. A.; Karmakar, P.; Opresko, P. L.; Piotrowski, J.; Dianova, I.; Dianov, G. L.; Bohr, V. A. Werner syndrome protein interacts with human flap endonuclease 1 and stimulates its cleavage activity. *EMBO J.* **2001**, *20*, 5791-5801.

(41)   Karanja, K. K.; Livingston, D. M. C-terminal flap endonuclease (rad27) mutations: lethal interactions with a DNA ligase I. mutation (cdc9-p) and suppression by proliferating cell nuclear antigen (POL30) in Saccharomyces cerevisiae. *Genetics* **2009**, *183*, 63-78.

(42)   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.

(43)   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**, *85*, 1101-1112.

(44)   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*, 31713-31724.

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

(46)   Harrington, J. J.; Lieber, M. R. The characterization of a mammalian DNA structure-specific endonuclease. *EMBO J.* **1994**, *13*, 1235-1246.

(47)   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 (Cambridge, MA, U. S.)* **2004**, *116*, 39-50.

(48)   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*, 717-729.

(49)   Negritto, M. C.; Qiu, J.; Ratay, D. O.; Shen, B.; Bailis, A. M. Novel function of Rad27 (FEN-1) in restricting short-sequence recombination. *Mol. Cell. Biol.* **2001**, *21*, 2349-2358.

(50)   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*, 30377-30383.

(51)   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*, 4507-4519.

(52)   Harrington, J. J.; Lieber, M. R. DNA structural elements required for FEN-1 binding. *J. Biol. Chem.* **1995**, *270*, 4503-4508.

(53)   Lyamichev, V.; Brow, M. A. D.; 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.

(54)   Kaiser, M. W.; Lyamicheva, N.; Ma, W.; Miller, C.; Neri, B.; Fors, L.; Lyamichev, V. I. A comparison of eubacterial and archaeal structure-specific 5'-exonucleases. *J. Biol. Chem.* **1999**, *274*, 21387-21394.

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

(56)   Storici, F.; Henneke, G.; Ferrari, E.; Gordenin, D. A.; Huebscher, U.; Resnick, M. A. The flexible loop of human FEN1 endonuclease is required for flap cleavage during DNA replication and repair. *EMBO J.* **2002**, *21*, 5930-5942.

(57)   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*, 8542-8547.

(58)   Finger, L. D.; Blanchard, M. S.; Theimer, C. A.; Sengerova, 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*, 22184-22194.

(59)   Friedrich-Heineken, E.; Huebscher, U. The Fen1 extrahelical 3'-flap pocket is conserved from archaea to human and regulates DNA substrate specificity. *Nucleic Acids Res.* **2004**, *32*, 2520-2528.

(60)   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*, 1772-1784.

(61)   Deshpande, R. A.; Shankar, V. Ribonucleases from T2 family. *Crit. Rev. Microbiol.* **2002**, *28*, 79-122.

(62)   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*, 65-72.

(63)   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*, 450-456.

(64)   Tomlinson, C. G.; Syson, K.; Sengerova, B.; Atack, J. M.; Sayers, J. R.; Swanson, L.; Tainer, J. A.; Williams, N. H.; Grasby, J. A. Neutralizing Mutations of Carboxylates That Bind Metal 2 in T5 Flap Endonuclease Result in an Enzyme That Still Requires Two Metal Ions. *J. Biol. Chem.* **2011**, *286*, 30878-30887.

(65)   Xu, Y.; Potapova, O.; Leschziner, A. E.; Grindley, N. D. F.; Joyce, C. M. Contacts between the 5' nuclease of DNA polymerase I and its DNA substrate. *J. Biol. Chem.* **2001**, *276*, 30167-30177.

(66)   Williams, R.; Sengerova, 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*, 34-48.

(67)   Exell, J. Spectroscopic evidence for catalyticqlly-required FEN1-mediated DNA conformational change; a novel strategy for FEN1 inhibition. The University of Sheffield, 2015.

(68)   Cassano, A. G.; Anderson, V. E.; Harris, M. E. Understanding the transition states of phosphodiester bond cleavage: Insights from heavy atom isotope effects. *Biopolymers* **2004**, *73*, 110-129.

(69)   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*, 25-33.

(70)   Pingoud, V.; Wende, W.; Friedhoff, P.; Reuter, M.; Alves, J.; Jeltsch, A.; Mones, L.; Fuxreiter, M.; Pingoud, A. On the Divalent Metal Ion Dependence of DNA Cleavage by Restriction Endonucleases of the EcoRI Family. *J. Mol. Biol.* **2009**, *393*, 140-160.

(71)   Bartkova, J.; Horejsi, Z.; Koed, K.; Kraemer, A.; Tort, F.; Zieger, K.; Guldberg, P.; Sehested, M.; Nesland, J. M.; Lukas, C.; Orntoft, T.; Lukas, J.; Bartek, J. DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. *Nature (London, U. K.)* **2005**, *434*, 864-870.

(72)   Hoeijmakers, J. H. J. Genome maintenance mechanisms for preventing cancer. *Nature (London, U. K.)* **2001**, *411*, 366-374.

(73)   Vainio, H.; Boffetta, P. Mechanisms of the combined effect of asbestos and smoking in the etiology of lung cancer. *Scand. J. Work, Environ. Health* **1994**, *20*, 235-242.

(74)   Loeb, L. A. A mutator phenotype in cancer. *Cancer Res.* **2001**, *61*, 3230-3239.

(75)   Kyriacou, C. P. An introduction to genetic analysis (7th edition), edited by A. J. F. Griffiths, J. H. Miller, D. T. Suzuki, R. C. LLewontin and W. M. Gelbart. *Trends Genet.* **2000**, *16*, 368.

(76)   Kucherlapati, M.; Yang, K.; Kuraguchi, M.; Zhao, J.; Lia, M.; Heyer, J.; Kane, M. F.; Fan, K.; Russell, R.; Brown, A. M. C.; Kneitz, B.; Edelmann, W.; Kolodner, R. D.; Lipkin, M.; Kucherlapati, R. Haploinsufficiency of Flap endonuclease (Fen1) leads to rapid tumor progression. *Proc Natl Acad Sci U S A* **2002**, *99*, 9924-9929.

(77)   Reva, B.; Antipin, Y.; Sander, C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res.* **2011**, *39*, e118.

(78)   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*, 3341-3348.

(79)   Kim, K.; Biade, S.; Matsumoto, Y. Involvement of flap endonuclease 1 in base excision DNA repair. *J. Biol. Chem.* **1998**, *273*, 8842-8848.

(80)   Lengauer, C.; Kinzler, K. W.; Vogelstein, B. Genetic instabilities in human cancers. *Nature (London)* **1998**, *396*, 643-649.

(81)   Sato, M.; Girard, L.; Sekine, I.; Sunaga, N.; Ramirez, R. D.; Kamibayashi, C.; Minna, J. D. Increased expression and no mutation of the Flap endonuclease (FEN1) gene in human lung cancer. *Oncogene* **2003**, *22*, 7243-7246.

(82)   Abdel-Fatah, T. M. A.; Russell, R.; Albarakati, N.; Maloney, D. J.; Dorjsuren, D.; Rueda, O. M.; Moseley, P.; Mohan, V.; Sun, H.; Abbotts, R.; Mukherjee, A.; Agarwal, D.; Illuzzi, J. L.; Jadhav, A.; Simeonov, A.; Ball, G.; Chan, S.; Caldas, C.; Ellis, I. O.; Wilson, D. M.; Madhusudan, S. Genomic and protein expression analysis reveals flap endonuclease 1 (FEN1) as a key biomarker in breast and ovarian cancer. *Mol. Oncol.* **2014**, *8*, 1326-1338.

(83)   Chung, L.; Onyango, D.; Guo, Z.; Jia, P.; Dai, H.; Liu, S.; Zhou, M.; Lin, W.; Pang, I.; Li, H.; Yuan, Y. C.; Huang, Q.; Zheng, L.; Lopes, J.; Nicolas, A.; Chai, W.; Raz, D.; Reckamp, K. L.; Shen, B. The FEN1 E359K germline mutation disrupts the FEN1-WRN interaction and FEN1 GEN activity, causing aneuploidy-associated cancers. *Oncogene* **2015**, *34*, 902-911.

(84)   Sun, H.; He, L.; Wu, H.; Pan, F.; Wu, X.; Zhao, J.; Hu, Z.; Sekhar, C.; Li, H.; Zheng, L.; Chen, H.; Shen, B. H.; Guo, Z. The FEN1 L209P mutation interferes with long-patch base excision repair and induces cellular transformation. *Oncogene* **2017**, *36*, 194-207.

(85)   Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **1976**, *72*, 248-254.

(86)   Clegg, R. M. Fluorescence resonance energy transfer and nucleic acids. *Methods Enzymol* **1992**, *211*, 353-388.

(87)   Wu, Z.; Lin, Y.; Xu, H.; Dai, H.; Zhou, M.; Tsao, S.; Zheng, L.; Shen, B. High risk of benzo[α]pyrene-induced lung cancer in E160D FEN1 mutant mice. *Mutat. Res., Fundam. Mol. Mech. Mutagen.* **2012**, *731*, 85-91.

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

(89)   Johnson, N. P.; Baase, W. A.; Von Hippel, P. H. Low-energy circular dichroism of 2-aminopurine dinucleotide as a probe of local conformation of DNA and RNA. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 3426-3431.

(90)   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. U. S. A.* **1986**, *83*, 5434-5438.

(91)   Eritja, R.; Horowitz, D. M.; Walker, P. A.; Ziehler-Martin, J. P.; 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.

(92)   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*, 025028/025021-025028/025016.

(93)   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.; Sarker, A. H.; Her, M. Z.; Rashid, F.; Hamdan, S. M.; Mirkin, S. M.; Grasby, J. A.; Tainer, J. A. Phosphate steering by Flap Endonuclease 1 promotes 5'-flap specificity and incision to prevent genome instability. *Nat. Commun.* **2017**, *8*, 15855.

(94)   Stryer, L.; Haugland, R. P. Energy transfer: a spectroscopic ruler. *Proc Natl Acad Sci U S A* **1967**, *58*, 719-726.

(95)   Fairclough, R. H.; Cantor, C. R. The use of singlet-singlet energy transfer to study macromolecular assemblies. *Methods Enzymol* **1978**, *48*, 347-379.

(96)   Steinmetzer, K.; Behlke, J.; Brantl, S.; Lorenz, M. CopR binds and bends its target DNA: A footprinting and fluorescence resonance energy transfer study. *Nucleic Acids Res.* **2002**, *30*, 2052-2060.

(97)   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*, 1857-1872.

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

(99)   Rashid, F.; Harris, P. D.; Zaher, M. S.; Sobhy, M. A.; Joudeh, L. I.; Piwonski, H.; Habuchi, S.; Hamdan, S. M.; Yan, C.; Ivanov, I.; Yan, C.; Ivanov, I.; Tsutakawa, S. E.; Tainer, J. A.; Tainer, J. A. Single-molecule FRET unveils induced-fit mechanism for substrate selectivity in flap endonuclease 1. *Elife* **2017**, *6*.

(100)   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*, 34239-34248.

(101)   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*, 8258-8268.

(102)   Allawi, H. T.; Kaiser, M. W.; Onufriev, A. V.; Ma, W.-P.; Brogaard, A. E.; Case, D. A.; Neri, B. P.; Lyamichev, V. I. Modeling of Flap Endonuclease Interactions with DNA Substrate. *J. Mol. Biol.* **2003**, *328*, 537-554.

7.1. Alignment of sequence-1: [mFEN-1_mRNA.xdna] with sequence-2:

7.1.1. [E160DmFEN1_mRNA.xdna] Similarity: 961/961 (100%) (Parallel):

The sequence reaction was performed by the University of Sheffield Medical School. The muta-tion position is highlighted (red). Nucleotides that could not be properly determined are denot-ed with a *.

```
      M   G   I   H   G   L   A   K   L   I
Seq_1  1     ------------------------------atgggaattcacggccttgccaaactaatt   30
                                          ||||||||||||||||||||||||||||||
Seq_2  1     tTTTGTTtaACTTtAAGaAggAGATATACCATGGGAATTCaCGGCCttgCCAAACTAATT   60
      F   C   L   T   L   R   R   R   Y   T   M   G   I   H   G   L   A   K   L   I

      A   D   V   A   P   S   A   I   R   E   N   D   I   K   S   Y   F   G   R   K
Seq_1  31    gctgatgtggcccccagtgccatccgtgagaatgacatcaagagctactttggtcgcaaa   90
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  61    GCTGATGTGGCCCCCagtgCCATCcgtGAGAATGACATCAAGAGCTACTTTGGTCGCAAA   120
      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   G   G
Seq_1  91    gtggccatcgatgcctccatgagcatctaccagttcctgattgctgttcgtcagggtggg   150
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  121   GTGGCCATCGATGCCTCCATGAGCATCTACCAGTTCCTGATTGCTGTTCGTCAGGGTGGG   180
      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   Y   R
Seq_1  151   gatgtgctgcagaacgaggagggtgagaccaccagccacctgatgggcatgttctaccgt   210
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  181   GATGTGCTGCAGAACGAGGAGGGTGAGACCACCAGCCACCTGATGGGCATGTTCTACCGT   240
      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   P   P
Seq_1  211   accatccgcatgatggagaatggcatcaagcctgtgtacgtctttgatggcaaaccacca   270
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  241   ACCATCCGCATGATGGAGAATGGCATCAAGCCTGTGTACGTCTTTGATGGCAAACCACCA   300
      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   K   Q
Seq_1  271   cagctgaagtcaggcgagctggccaagcgcagtgagaggcgcgccgaggctgagaagcaa   330
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  301   CAGCTGAAGTCAGGCGAGCTGGCCAAGCGCAGTGAGAGGCGCGCCGAGGCTGAGAAGCAA   360
      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   E   A   G   M   E   E   E   V   E   K   F   T   K   R   L
Seq_1  331   ctgcagcaggctcaggaggctgggatggaggaggaggtggagaagttcaccaagaggctc   390
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  361   CTGCAGCAGGCTCAGGAGGCTGGGATGGAGGAGGAGGTGGAGAAGTTCACCAAGAGGCTC   420
      L   Q   Q   A   Q   E   A   G   M   E   E   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
Seq_1  391   gtgaaggtcaccaagcaacacaatgatgagtgcaaacacctgctgagcctcatgggcatc   450
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  421   GTGAAGGTCACCAAGCAACACAATGATGAGTGCAAACACCTGCTGAGCCTCATGGGCATC   480
      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   A   K   A   G
Seq_1  451   ccttaccttgatgcacccagcgaggcagaggccagctgtgctgccctggcaaaggctggc   510
             ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Seq_2  481   CCTTACCTTGATGCACCCAGCGAGGCAGATGCCAGCTGTGCTGCCCTGGCAAAGGCTGGC   540
      P   Y   L   D   A   P   S   E   A   D   A   S   C   A   A   L   A   K   A   G
      K   V   Y   A   A   A   T   E   D   M   D   C   L   T   F   G   S   P   V   L
```

138

```
Seq_1  511   aaagtctatgctgcggccacggaggacatggactgcctcactttttggcagccccgtgcta  570
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  541   AAAGTCTATGCTGCGGCCACGGAGGACATGGACTGCCTCACTTTTGGCAGCCCCGTGCTA  600
             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   L   S
Seq_1  571   atgcgacacttaactgccagtgaggccaagaagctgcccatccaagagttccatctgagc  630
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  601   ATGCGACACTTAACTGCCAGTGAGGCCAAGAAGCTGCCCATCCAAGAGTTCCATCTGAGC  660
             M   R   H   L   T   A   S   E   A   K   K   L   P   I   Q   E   F   H   L   S


             R   V   L   Q   E   L   G   L   N   Q   E   Q   F   V   D   L   C   I   L   L
Seq_1  631   cgcgtcctgcaggagctgggtctgaaccaggagcagtttgtggatctgtgcatcctgctg  690
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  661   CGCGTCCTGCAGGAGCTGGGTCTGAACCAGGAGCAGTTTGTGGATCTGTGCATCCTGCTG  720
             R   V   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   L   I
Seq_1  691   ggtagcgactactgcgagagcatccgtggcattgggcccaagcgggctgtggatctcatc  750
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  721   GGTAGCGACTACTGCGAGAGCATCCGTGGCATTGGGCCCAAGCGGGCTGTGGATCTCATC  780
             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   S   K   Y   P   V
Seq_1  751   cagaaacataagagcatcgaggagatcgtgaggcggctggaccccagcaagtaccccgtt  810
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  781   CAGAAACATAAGAGCATCGAGGAGATCGTGAGGCGGCTGGACcCCAGCAAGTACCCCGTT  840
             Q   K   H   K   S   I   E   E   I   V   R   R   L   D   P   S   K   Y   P   V


             P   E   N   W   L   H   K   E   A   Q   Q   L   F   L   E   P   E   V   L   D
Seq_1  811   ccagagaactggctccacaaggaagcccagcagctcttcctggagccagaagtactggac  870
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  841   CCAGAGAACTGGCTCCACAAGGAAGCCCAGCAGCTctTCCTGGAGCCAGAAGTACTGGAC  900
             P   E   N   W   L   H   K   E   A   Q   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   V   K   F   M
Seq_1  871   ccagagtctgtggagctgaagtggagcgagcccaaatgaagaagagttggtcaaatttatg  930
             |||||||||||||||||||||| ||||||||||||||||||| ||||||| ||
Seq_2  901   CCAGAGTCTGTGGAGCtgaagtGGa-cGAGCCAAATGAAGAagagTtGg-caaaTTt-tg  957
             P   E   S   V   E   L   K   W   T       S   Q   M   K   K   S   W       Q   I   L


             C   G   E   K   Q   F   S   E   E   R   I   R   S   G   V   K   R   L   S   K
Seq_1  931   tgtggtgaaaagcagttttctgaagagcgaattcgcagtggggtcaagcggctgagtaag  990
             |||||||||||||||||| |||||||||||
Seq_2  958   TgtggtgaAAAGcagTTTT-TgAAGAGCGaat----------------------------  988
             C   G   E   K   Q   F   L   K   S   E   X


             S   R   Q   G   S   T   Q   G   R   L   D   D   F   F   K   V   T   G   S   L
Seq_1  991   agccgccagggcagcacccagggacgcctcgatgatttcttcaaggtgacaggctcactc  1050
Seq_2  989   ------------------------------------------------------------  988


             S   S   A   K   R   K   E   P   E   P   K   G   P   A   K   K   K   A   K   T
Seq_1  1051  tcctcagctaagcgcaaggagccagaacccaaggggcctgctaagaagaaagcaaagact  1110
Seq_2  989   ------------------------------------------------------------  988


             G   G   A   G   K   F   R   R   G   K   *
Seq_1  1111  ggggggagcggggaagttccgaaggggaaaataa   1143
Seq_2  989   --------------------------------   988
                                             X
```

7.1.2. [E160DmFEN1-mRNA.xdna] similarity: 1057/1061 (99.62%)] (anti-parallel):

The sequence reaction was performed by the University of Sheffield Medical School. The mutation position is high-

lighted (red). Nucleotides that could not be properly determined are denoted with a *.

```
       M  G  I  H  G  L  A  K  L  I  A  D  V  A  P  S  A  I  R  E
Seq_1   1     atgggaattcacggccttgccaaactaattgctgatgtggcccccagtgccatccgtgag   60

Seq_2 1135    ------------------------------------------------------------   1136


       N  D  I  K  S  Y  F  G  R  K  V  A  I  D  A  S  M  S  I  Y
Seq_1  61     aatgacatcaagagctactttggtcgcaaagtggccatcgatgcctccatgagcatctac   120
                              ||  |     ||||  |||||||||||||||||||||||  |
Seq_2 1135    ------------------tTTGGtccAaa-tggcCAtCgAtgcctccATgAGCAt-t--   1099
                               L  V  Q  M     A  I  D  A  S  M  S  I


       Q  F  L  I  A  V  R  Q  G  G  D  V  L  Q  N  E  E  G  E  T
Seq_1 121     cagttcctgattgctgttcgtcaggtggggatgtgctgcagaacgaggagggtgagacc   180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2 1098    CAGTTCCTGATTGCTGTTCGTCAgGGTGGGGATGtGCtgcaGAACGAGGAGGGTGAGAcc   1039
              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  Y  R  T  I  R  M  M  E  N  G  I  K
Seq_1 181     accagccacctgatgggcatgttctaccgtaccatccgcatgatggagaatggcatcaag   240
               |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Seq_2 1038    --CAGCCACCTGATGGGCatGTTCt-cCGTACCATCCGCATGATGGAGAATGGCATCAAG   982
                Q  P  P  D  G  H  V  L     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
Seq_1 241     cctgtgtacgtctttgatggcaaaccaccacagctgaagtcaggcgagctggccaagcgc   300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2 981     CCTGTGTACGTCTTTGATGGCAAACCACCACAGCTGAAGTCAGGCGAGCTGgcCAAGCGC   922
              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  E  A  G  M  E
Seq_1 301     agtgagaggcgcgccgaggctgagaagcaactgcagcaggctcaggaggctgggatggag   360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2 921     AGTGAGAGGCGCGCCGAGGCTGAGAAGCAACTGCAGCAGGCTCAGGAGGCTGGGATGGAG   862
              S  E  R  R  A  E  A  E  K  Q  L  Q  Q  A  Q  E  A  G  M  E


       E  E  V  E  K  F  T  K  R  L  V  K  V  T  K  Q  H  N  D  E
Seq_1 361     gaggaggtggagaagttcaccaagaggctcgtgaaggtcaccaagcaacacaatgatgag   420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2 861     GAGGAGGTGGAGAAGTTCACCAAGAGGCTCGTGAAGGTCACCAAGCAACACAATGATGAG   802
              E  E  V  E  K  F  T  K  R  L  V  K  V  T  K  Q  H  N  D  E


       C  K  H  L  L  S  L  M  G  I  P  Y  L  D  A  P  S  E  A  E
Seq_1 421     tgcaaacacctgctgagcctcatgggcatcccttaccttgatgcacccagcgaggcagag   480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2 801     TGCAaACACCTGCTGAGCCTCATGGGCATCCCTTACCTTGATGCACCCAGCGAGGCAGAT   742
              C  K  H  L  L  S  L  M  G  I  P  Y  L  D  A  P  S  E  A  D


       A  S  C  A  A  L  A  K  A  G  K  V  Y  A  A  A  T  E  D  M
Seq_1 481     gccagctgtgctgccctggcaaaggctggcaaagtctatgctgcggccacggaggacatg   540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2 741     GCCAGCTGTGCTGCCCTGGCAAAGGCTGGCAAAGTCTATGCTGCGGCCACGGAGGACATG   682
              A  S  C  A  A  L  A  K  A  G  K  V  Y  A  A  A  T  E  D  M


       D  C  L  T  F  G  S  P  V  L  M  R  H  L  T  A  S  E  A  K
Seq_1 541     gactgcctcacttttggcagccccgtgctaatgcgacacttaactgccagtgaggccaag   600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2 681     GACTGCCTCACTTTTGGCAGCCCCGTGCTAATGCGACACTTAACTGCCAGTGAGGCCAAG   622
              D  C  L  T  F  G  S  P  V  L  M  R  H  L  T  A  S  E  A  K


       K  L  P  I  Q  E  F  H  L  S  R  V  L  Q  E  L  G  L  N  Q
Seq_1 601     aagctgcccatccaagagttccatctgagccgcgtcctgcaggagctgggtctgaaccag   660
```

```
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  621   AAGCTGCCCATCCAAGAGTTCCATCTGAGCCGCGTCCTGCAGGAGCTGGGTCTGAACCAG   562
             K  L  P  I  Q  E  F  H  L  S  R  V  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
Seq_1  661   gagcagtttgtggatctgtgcatcctgctgggtagcgactactgcgagagcatccgtggc   720
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  561   GAGCAGTTTGTGGATCTGTGCATCCTGCTGGGTAGCGACTACTGCGAGAGCATCCGTGGC   502
             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  V
Seq_1  721   attgggcccaagcgggctgtggatctcatccagaaacataagagcatcgaggagatcgtg   780
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  501   ATTGGGCCCAAGCGGGCTGTGGATCTCATCCAGAAACATAAGAGCATCGAGGAGATCGTG   442
             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  S  K  Y  P  V  P  E  N  W  L  H  K  E  A  Q
Seq_1  781   aggcggctggaccccagcaagtaccccgttccagagaactggctccacaaggaagcccag   840
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  441   AGGCGGCTGGACCCCAGCAAGTACCCCGTTCCAGAGAACTGGCTCCACAAGGAAGCCCAG   382
             R  R  L  D  P  S  K  Y  P  V  P  E  N  W  L  H  K  E  A  Q


             Q  L  F  L  E  P  E  V  L  D  P  E  S  V  E  L  K  W  S  E
Seq_1  841   cagctcttcctggagccagaagtactggacccagagtctgtggagctgaagtggagcgag   900
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  381   CAGCTCTTCCTGGAGCCAGAAGTACTGGACCCAGAGTCTGTGGAGCTGAAGTGGAGCGAG   322
             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  V  K  F  M  C  G  E  K  Q  F  S  E  E  R
Seq_1  901   ccaaatgaagaagagttggtcaaatttatgtgtggtgaaaagcagttttctgaagagcga   960
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  321   CCAAATGAAGAAGAGTTGGTCAAATTTATGTGTGGTGAAAAGCAGTTTTCTGAAGAGCGA   262
             P  N  E  E  E  L  V  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  L
Seq_1  961   attcgcagtggggtcaagcggctgagtaagagccgccagggcagcacccagggacgcctc   1020
             |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Seq_2  261   ATTCGCAGTGGGGTCAAGCGGCTGAGTAAGAGCCGCCAGGGTAGCACCCAGGGACGCCTC   202
             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  P
Seq_1  1021  gatgatttcttcaaggtgacaggctcactctcctcagctaagcgcaaggagccagaaccc   1080
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  201   GATGATTTCTTCAAGGTGACAGGCTCACTCTCCTCAGCTAAGCGCAAGGAGCCAGAACCC   142
                D  D  F  F  K  V  T  G  S  L  S  S  A  K  R  K  E  P  E  P


             K  G  P  A  K  K  K  A  K  T  G  G  A  G  K  F  R  R  G  K
Seq_1  1081  aaggggcctgctaagaagaaagcaaagactggggggagcggggaagttccgaagggggaaa   1140
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  141   AAGGGGCCTGCTAAGAAGAAAGCAAAGACTGGGGGGAGCGGGGAAGTTCCGAAGGGGAAAA   82
             K  G  P  A  K  K  K  A  K  T  G  G  A  G  K  F  R  R  G  K


                                            *
Seq_1  1141  ------------------taa----------------------------------------   1143
                                            |||
Seq_2  81    CATCATCATCATCATCACtAaGcTTGCGGCCGCACTCGAGCACCACCACCACCACCActG   22
             H  H  H  H  H  H  *  A  C  G  R  T  R  A  P  P  P  P  L


                                    X  X
Seq_1  1144  --------------------   1143

Seq_2  21    AGATCCgGGCtgtAacaAgcg   1
             R  S  G  L  *  Q  A
```

7.1.3. [WTmFEN1.seq] similarity: 1108/1125 (98.49 %) (Parallel):

The sequence reaction was performed by the University of Sheffield Medical School. Nucleotides that could not be properly determined are denoted with a *.

```
                                                                M  G  I  H  G
Seq_1    1    --------------------------------------------atgggaattcacggc   15
                                                          |||||||||||||||
Seq_2    1    aGaatTCCTcTagaTaTTTTGTTtaCTTTAaGAaGGAgaTATACCATGGGAATTCACGGC   60
              R  I  P  L  D  I  L  F  T  L  R  R  R  Y  T  M  G  I  H  G

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

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

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

              G  M  F  Y  R  T  I  R  M  M  E  N  G  I  K  P  V  Y  V  F
Seq_1    196  ggcatgttctaccgtaccatccgcatgatggagaatggcatcaagcctgtgtacgtcttt   255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2    241  GGCATGTTCTACCGTACCATCCGCATGATGGAGAATGGCATCAAGCCTGTGTACGTCTTT   300
              G  M  F  Y  R  T  I  R  M  M  E  N  G  I  K  P  V  Y  V  F

              D  G  K  P  P  Q  L  K  S  G  E  L  A  K  R  S  E  R  R  A
Seq_1    256  gatggcaaaccaccacagctgaagtcaggcgagctggccaagcgcagtgagaggcgcgcc   315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2    301  GATGGCAAACCACCACAGCTGAAGTCAGGCGAGCTGGCCAAGCGCAGTGAGAGGCGCGCC   360
              D  G  K  P  P  Q  L  K  S  G  E  L  A  K  R  S  E  R  R  A

              E  A  E  K  Q  L  Q  Q  A  Q  E  A  G  M  E  E  V  E  K
Seq_1    316  gaggctgagaagcaactgcagcaggctcaggaggctgggatggaggaggaggtggagaag   375
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2    361  GAGGCTGAGAAGCAACTGCAGCAGGCTCAGGAGGCTGGGATGGAGGAGGAGGTGGAGAAG   420
              E  A  E  K  Q  L  Q  Q  A  Q  E  A  G  M  E  E  V  E  K

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

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

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

              G  S  P  V  L  M  R  H  L  T  A  S  E  A  K  K  L  P  I  Q
Seq_1    556  ggcagccccgtgctaatgcgacacttaactgccagtgaggccaagaagctgcccatccaa   615
```

```
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2   601  GGCAGCCCCGTGCTAATGCGACACTTAACTGCCAGTGAGGCCAAGAAGCTGCCCATCCAA   660
              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   V   L   Q   E   L   G   L   N   Q   E   Q   F   V   D
Seq_1   616  gagttccatctgagccgcgtcctgcaggagctgggtctgaaccaggagcagtttgtggat   675
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2   661  GAGTTCCATCTGAGCCGCGTCCTGCAGGAGCTGGGTCTGAACCAGGAGCAGTTTGTGGAT   720
              E   F   H   L   S   R   V   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
Seq_1   676  ctgtgcatcctgctgggtagcgactactgcgagagcatccgtggcattgggcccaagcgg   735
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2   721  CTGTGCATCCTGCTGGGTAGCGACTACTGCGAGAGCATCCGTGGCATTGGGCCCAAGCGG   780
              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   V   R   R   L   D   P
Seq_1   736  gctgtggatctcatccagaaacataagagcatcgaggagatcgtgaggcggctggacccc   795
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2   781  GCTGTGGATCTCATCCAGAAACATAAGAGCATCGAGGAGATCGTGAGGCGGCTGGACCCC   840
              A   V   D   L   I   Q   K   H   K   S   I   E   E   I   V   R   R   L   D   P

              S   K   Y   P   V   P   E   N   W   L   H   K   E   A   Q   Q   L   F   L   E
Seq_1   796  agcaagtaccccgttccagagaactggctccacaaggaagcccagcagctcttcctggag   855
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2   841  AGCAAGTACcCCGTTCCAGAGAACTGGCTCCACAAGGAAGCCCAGCAGCTCTTCCTGGAG   900
              S   K   Y   P   V   P   E   N   W   L   H   K   E   A   Q   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
Seq_1   856  ccagaagtactggacccagagtctgtggagctgaagtggagcgagccaaatgaagaagag   915
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2   901  CCAGAAGTACTGGACCCAGAGTCTGTGGAGCTGAAGTGGAGCGAGCCAAATGAaGAAGa-   959
              P   E   V   L   D   P   E   S   V   E   L   K   W   S   E   P   N   E   E   D

              L   V   K   F   M   C   G   E   K   Q   F   S   E   E   R   I   R   S   G   V
Seq_1   916  ttggtcaaatttatgtgtggtgaaaagcagttttctgaagagcgaattcgcagtggggtc   975
             |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Seq_2   960  tTGGTCAAATTTATGTGTGGTGAAA-GCAGTTTTCTGAAGAGCGAATTCgCAGTGGGGTC   1018
               W   S   N   L   C   V   V   K   A   V   F   *   R   A   N   S   Q   W   G   Q

              K   R   L   S   K   S   R   Q   G   S   T   Q   G   R   L   D   D   F   F   K
Seq_1   976  aagcggctgagtaagagccgccagggcagcacccagggacgcctcgatgatttcttcaag   1035
             |||||||||||||||||||||||| |   |||||||||||||||||||||||||||||||
Seq_2   1019 AAGCGGCTGagtAagaGCCGCCAGGGTA--ACCcAGGgACGCCTCGATGATTTCTTCAAG   1076
               A   A   E   *   E   P   P   G   *       P   R   D   A   S   M   I   S   S   R

              V   T   G   S   L   S   S   A   K   R   K   E   P   E   P   K   G   P   A   K
Seq_1   1036 gtgacaggctcactctcctcagctaagcgcaaggagccagaacccaaggggcctgctaag   1095
             |||          |||||||||||||||||||||||||||||    ||||||||||||||||
Seq_2   1077 Gtg------tcactctccTCAGCTAAGcgcAAGGAGCCa--accCAAGGGgCCtGCTAa-   1127
               C               H   S   P   Q   L   S   A   R   S   Q       P   K   G   P   A   K

              K   K   A   K   T   G   G   A   G   K   F   R   R   G   K   *
Seq_1   1096 aagaaagcaaagactggggagcggggaagttccgaaggggaaaataa   1143
                          ||| |       |||         || |   | |  |||||||||
Seq_2   1128 -------aaaaaAaaCTGGGgaaCCGggAaattCcaAGGGgaAa----   1164
                K   K   T   G   E   P   G   N   S   K   G   K   X   X
```

7.1.4. [WTmFEN1.seq] similarity: 1039/1049 (99.05 %) (anti-parallel):

The sequence reaction was performed by the University of Sheffield Medical School. Nucleotides that could not be properly determined are denoted with a *.

```
           M  G  I  H  G  L  A  K  L  I  A  D  V  A  P  S  A  I  R  E
Seq_1   1  atgggaattcacggccttgccaaactaattgctgatgtggcccccagtgccatccgtgag   60

Seq_2 1119 ------------------------------------------------------------ 1120


           N  D  I  K  S  Y  F  G  R  K  V  A  I  D  A  S  M  S  I  Y
Seq_1  61  aatgacatcaagagctactttggtcgcaaagtggccatcgatgcctccatgagcatctac  120
                                             |  |        ||||||||||   |||
Seq_2 1119 --------------------------------gGCCATCgatgctCCATgAGct--tAC  1095
                                            A  I  D  A  P  *  A     Y

           Q  F  L  I  A  V  R  Q  G  G  D  V  L  Q  N  E  E  G  E  T
Seq_1 121  cagttcctgattgctgttcgtcagggtggggatgtgctgcagaacgaggagggtgagacc  180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2 1094 CAGTTcCTGATTGCTGTTcgtcAGGGTGGGGATgtGCTGCAGAACGAGGaGGGTGAGACc 1035
           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  Y  R  T  I  R  M  M  E  N  G  I  K
Seq_1 181  accagccacctgatgggcatgttctaccgtaccatccgcatgatggagaatggcatcaag  240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2 1034 aCCAGCCaCCTGATGGGCATGTTCTACCGtaCCATCCGCATGATGGAGAATGgcATCAAG  975
           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  P  P  Q  L  K  S  G  E  L  A  K  R
Seq_1 241  cctgtgtacgtctttgatggcaaaccaccacagctgaagtcaggcgagctggccaagcgc  300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2 974  CCTGTGTACGTCTTTGATGGCAAACCACCACAGCTGAAGTCAGGCGAGCTGGCCAAGCGC  915
           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  E  A  G  M  E
Seq_1 301  agtgagaggcgcgccgaggctgagaagcaactgcagcaggctcaggaggctgggatggag  360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2 914  AGTGAGAGGCGCGCCGAGGCTGAGAAGCAACTGCAGCAGGCTCAGGAGGCTGGGATGGAG  855
           S  E  R  R  A  E  A  E  K  Q  L  Q  Q  A  Q  E  A  G  M  E

           E  E  V  E  K  F  T  K  R  L  V  K  V  T  K  Q  H  N  D  E
Seq_1 361  gaggaggtggagaagttcaccaagaggctcgtgaaggtcaccaagcaacacaatgatgag  420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2 854  GAGGAGGTGGAGAAGTTCACCAAGAGGCTCGTGAAGGTCACCAAGCAACACAATGATGAG  795
           E  E  V  E  K  F  T  K  R  L  V  K  V  T  K  Q  H  N  D  E

           C  K  H  L  L  S  L  M  G  I  P  Y  L  D  A  P  S  E  A  E
Seq_1 421  tgcaaacacctgctgagcctcatgggcatcccttaccttgatgcacccagcgaggcagag  480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2 794  TGCAAACACCTGCTGAGCCTCATGGGCATCCCTTACCTTGATGCACCCAGCGAGGCAGAG  735
           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  A  K  A  G  K  V  Y  A  A  A  T  E  D  M
Seq_1 481  gccagctgtgctgccctggcaaaggctggcaaagtctatgctgcggccacggaggacatg  540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2 734  GCCAGCTGTGCTGCcCTGGCAAAGGCTGGCAAAGTCTATGCTGCGGCCACGGAGGACATG  675
           A  S  C  A  A  L  A  K  A  G  K  V  Y  A  A  A  T  E  D  M

           D  C  L  T  F  G  S  P  V  L  M  R  H  L  T  A  S  E  A  K
Seq_1 541  gactgcctcacttttggcagccccgtgctaatgcgacacttaactgccagtgaggccaag  600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2 674  GACTGCCTCACTTTTGGCAGCCCCGTGCTAATGCGACACTTAACTGCCAGTGAGGCCAAG  615
           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  V  L  Q  E  L  G  L  N  Q
Seq_1 601  aagctgcccatccaagagttccatctgagccgcgtcctgcaggagctgggtctgaaccag  660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Seq_2    614   AAGCTGCCCATCCAAGAGTTCCATCTGAGCCGCGTCCTGCAGGAGCTGGGTCTGAACCAG   555
               K   L   P   I   Q   E   F   H   L   S   R   V   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
Seq_1    661   gagcagtttgtggatctgtgcatcctgctgggtagcgactactgcgagagcatccgtggc   720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2    554   GAGCAGTTTGTGGATCTGTGCATCCTGCTGGGTAGCGACTACTGCGAGAGCATCCGTGGC   495
               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   V
Seq_1    721   attgggcccaagcgggctgtggatctcatccagaaacataagagcatcgaggagatcgtg   780
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2    494   ATTGGGCCCAAGCGGGCTGTGGATCTCATCCAGAAACATAAGAGCATCGAGGAGATCGTG   435
               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   S   K   Y   P   V   P   E   N   W   L   H   K   E   A   Q
Seq_1    781   aggcggctggaccccagcaagtaccccgttccagagaactggctccacaaggaagcccag   840
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2    434   AGGCGGCTGGACCCCAGCAAGTACCCCGTTCCAGAGAACTGGCTCCACAAGGAAGCCCAG   375
               R   R   L   D   P   S   K   Y   P   V   P   E   N   W   L   H   K   E   A   Q


               Q   L   F   L   E   P   E   V   L   D   P   E   S   V   E   L   K   W   S   E
Seq_1    841   cagctcttcctggagccagaagtactggacccagagtctgtggagctgaagtggagcgag   900
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2    374   CAGCTCTTCCTGGAGCCAGAAGTACTGGACCCAGAGTCTGTGGAGCTGAAGTGGAGCGAG   315
               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   V   K   F   M   C   G   E   K   Q   F   S   E   E   R
Seq_1    901   ccaaatgaagaagagttggtcaaatttatgtgtggtgaaaagcagttttctgaagagcga   960
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2    314   CCAAATGAAGAAGAGTTGGTCAAATTTATGTGTGGTGAAAAGCAGTTTTCTGAAGAGCGA   255
               P   N   E   E   E   L   V   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   L
Seq_1    961   attcgcagtggggtcaagcggctgagtaagagccgccagggcagcacccagggacgcctc   1020
               |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Seq_2    254   ATTCGCAGTGGGGTCAAGCGGCTGAGTAAGAGCCGCCAGGGTAGCACCCAGGGACGCCTC   195
               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   P
Seq_1    1021  gatgatttcttcaaggtgacaggctcactctcctcagctaagcgcaaggagccagaaccc   1080
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2    194   GATGATTTCTTCAAGGTGACAGGCTCACTCTCCTCAGCTAAGCGCAAGGAGCCAGAACCC   135
               D   D   F   F   K   V   T   G   S   L   S   S   A   K   R   K   E   P   E   P


               K   G   P   A   K   K   K   A   K   T   G   G   A   G   K   F   R   R   G   K
Seq_1    1081  aaggggcctgctaagaagaaagcaaagactgggggagcggggaagttccgaagggggaaa   1140
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2    134   AAGGGGCCTGCTAAGAAGAAAGCAAAGACTGGGGGAGCGGGGAAGTTCCGAAGGGGGAAA   75
               K   G   P   A   K   K   K   A   K   T   G   G   A   G   K   F   R   R   G   K


                                              *
Seq_1    1141  ------------------taa----------------------------------------   1143
                                              |||
Seq_2    74    CATCATCATCATCATCACTAAGCTTGCGGCCGCACTCGAGCACCACCACCACCACCACTg   15
               H   H   H   H   H   H   *   A   C   G   R   T   R   A   P   P   P   P   P   L


                          X   X
Seq_1    1144  --------------  1143

Seq_2    14    AgaTCCGGCtgctc  1
               R   S   G   C   X
```

7.1.5. [A159VmFEN1.seq] similarity: 1115/1132 (98.50 %) (Parallel):

The sequence reaction was performed by the University of Sheffield Medical School. The mutation position is high-

lighted (red). Nucleotides that could not be properly determined are denoted with a *.

```
                                                     M   G   I   H   G   L
Seq_1    1    -------------------------------------------atgggaattcacggcc   16
                                                         ||||||||||||||||
Seq_2    1    GgaattCCTTaaATaTTTTGTTtACTTtAaGAAGGAGaTATACCATGGGAATTCACGGCC   60
                N   S   L   N   I   L   F   T   L   R   R   R   Y   T   M   G   I   H   G   L

              A   K   L   I   A   D   V   A   P   S   A   I   R   E   N   D   I   K   S   Y
Seq_1   17    ttgccaaactaattgctgatgtggcccccagtgccatccgtgagaatgacatcaagagct   76
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2   61    TTGCCAAACTAATTGCTGATGTGGCCCCCAGTGCCATCCgtgAGAATGACATCAAGAGCT  120
              A   K   L   I   A   D   V   A   P   S   A   I   R   E   N   D   I   K   S   Y

              F   G   R   K   V   A   I   D   A   S   M   S   I   Y   Q   F   L   I   A   V
Seq_1   77    actttggtcgcaaagtggccatcgatgcctccatgagcatctaccagttcctgattgctg  136
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  121    ACTTTGGTCGCAAAGTGGCCATCGATGCCTCCATGAGCATCTACCAGTTCCTGATTGCTG  180
              F   G   R   K   V   A   I   D   A   S   M   S   I   Y   Q   F   L   I   A   V

              R   Q   G   G   D   V   L   Q   N   E   E   G   E   T   T   S   H   L   M   G
Seq_1  137    ttcgtcagggtggggatgtgctgcagaacgaggagggtgagaccaccagccacctgatgg  196
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  181    TTCGTCAGGGTGGGGATGTGCTGCAGAACGAGGAGGGTGAGACCACCAGCCACCTGATGG  240
              R   Q   G   G   D   V   L   Q   N   E   E   G   E   T   T   S   H   L   M   G

              M   F   Y   R   T   I   R   M   M   E   N   G   I   K   P   V   Y   V   F   D
Seq_1  197    gcatgttctaccgtaccatccgcatgatggagaatggcatcaagcctgtgtacgtctttg  256
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  241    GCATGTTCTACCGTACCATCCGCATGATGGAGAATGGCATCAAGCCTGTGTACGTCTTTG  300
              M   F   Y   R   T   I   R   M   M   E   N   G   I   K   P   V   Y   V   F   D

              G   K   P   P   Q   L   K   S   G   E   L   A   K   R   S   E   R   R   A   E
Seq_1  257    atggcaaaccaccacagctgaagtcaggcgagctggccaagcgcagtgagaggcgcgccg  316
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  301    ATGGCAAACCACCACAGCTGAAGTCAGGCGAGCTGGCCAAGCGCAGTGAGAGGCGCGCCG  360
              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   E   A   G   M   E   E   E   V   E   K   F
Seq_1  317    aggctgagaagcaactgcagcaggctcaggaggctgggatggaggaggaggtggagaagt  376
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  361    AGGCTGAGAAGCAACTGCAGCAGGCTCAGGAGGCTGGGATGGAGGAGGAGGTGGAGAAGT  420
              A   E   K   Q   L   Q   Q   A   Q   E   A   G   M   E   E   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
Seq_1  377    tcaccaagaggctcgtgaaggtcaccaagcaacacaatgatgagtgcaaacacctgctga  436
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  421    TCACCAAGAGGCTCGTGAAGGTCACCAAGCAACACAATGATGAGTGCAAACACCTGCTGA  480
              T   K   R   L   V   K   V   T   K   Q   H   N   D   E   C   K   H   L   L   S

              L   M   G   I   P   Y   L   D   A   P   S   E   **A**   E   A   S   C   A   A   L
Seq_1  437    gcctcatgggcatcccttaccttgatgcacccagcgaggcagaggccagctgtgctgccc  496
              |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Seq_2  481    GCCTCATGGGCATCCCTTACCTTGATGCACCCAGCGAGGTAGAGGCCAGCTGTGCTGCCC  540
              L   M   G   I   P   Y   L   D   A   P   S   E   **V**   E   A   S   C   A   A   L

              A   K   A   G   K   V   Y   A   A   A   T   E   D   M   D   C   L   T   F   G
Seq_1  497    tggcaaaggctggcaaagtctatgctgcggccacggaggacatggactgcctcacttttg  556
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  541    TGGCAAAGGCTGGCAAAGTCTATGCTGCGGCCACGGAGGACATGGACTGCCTCACTTTTG  600
              A   K   A   G   K   V   Y   A   A   A   T   E   D   M   D   C   L   T   F   G

              S   P   V   L   M   R   H   L   T   A   S   E   A   K   K   L   P   I   Q   E
Seq_1  557    gcagccccgtgctaatgcgacacttaactgccagtgaggccaagaagctgcccatccaag  616
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Seq_2   601  GCAGCCCCGTGCTAATGCGACACTTAACTGCCAGTGAGGCCAAGAAGCTGCCCATCCAAG   660
               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  V  L  Q  E  L  G  L  N  Q  E  Q  F  V  D  L
Seq_1   617  agttccatctgagccgcgtcctgcaggagctgggtctgaaccaggagcagtttgtggatc   676
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2   661  AGTTCCATCTGAGCCGCGTCCTGCAGGAGCTGGGTCTGAACCAGGAGCAGTTTGTGGATC   720
               F  H  L  S  R  V  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
Seq_1   677  tgtgcatcctgctgggtagcgactactgcgagagcatccgtggcattgggcccaagcggg   736
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2   721  TGTGCATCCTGCTGGGTAGCGACTACTGCGAGAGCATCCGTGGCATTGGGCCCAAGCGGG   780
               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  V  R  R  L  D  P  S
Seq_1   737  ctgtggatctcatccagaaacataagagcatcgaggagatcgtgaggcggctggacccca   796
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2   781  CTGTGGATCTCATCCAGAAACATAAGAGCATCGAGGAGATCGTGAGGCGGCTGGACCCCA   840
               V  D  L  I  Q  K  H  K  S  I  E  E  I  V  R  R  L  D  P  S


               K  Y  P  V  P  E  N  W  L  H  K  E  A  Q  Q  L  F  L  E  P
Seq_1   797  gcaagtacccgttccagagaactggctccacaaggaagcccagcagctcttcctggagc   856
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2   841  GCAAGTACCCGTTCCAGAGAACTGGCTCCACAAGGAAGCCCAGCAGCTCTTCCTGGAGC   900
               K  Y  P  V  P  E  N  W  L  H  K  E  A  Q  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
Seq_1   857  cagaagtactggacccagagtctgtggagctgaagtggagcgagccaaatgaagaagagt   916
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2   901  CAGAAGTACTGGACCCAGAGTCTGTGGAGCTGAAGTGGAGCGAGCCAAATGAAGAAGAGT   960
               E  V  L  D  P  E  S  V  E  L  K  W  S  E  P  N  E  E  E  L


               V  K  F  M  C  G  E  K  Q  F  S  E  E  R  I  R  S  G  V  K
Seq_1   917  tggtcaaatttatgtgtggtgaaaagcagttttctgaagagcgaattcgcagtggggtca   976
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2   961  TGGTCAAATTTATGTGTGGTGAAAAGCAGTTTTCTGAAGAGCGAATTcgcAGTGGGGTCA   1020
               V  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  L  D  D  F  F  K  V
Seq_1   977  agcggctgagtaagagccgccagggcagcacccagggacgcctcgatgatttcttcaagg   1036
             |||||||||   |  |||||||||||  ||||||||||||||||||||||||||||||||
Seq_2   1021 AGCGGCTGga-aTAaagccGCCAGGgTAGCACCCAGGgACGCCTCGATGaTTTCTTCAAGg   1079
               R  L  N  K  A  A  R  V  A  P  R  D  A  S  M  I  S  S  R


               T  G  S  L  S  S  A  K  R  K  E  P  E  P  K  G  P  A  K  K
Seq_1   1037 tgacaggctcactctcctcagctaagcgcaaggagccagaacccaagggggcctgctaaga   1096
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2   1080 tgACAGGCTcactctCCTcAGCTAagcgcAAGGAgCCAgAACCcAAGGGGCCtGCTAa--   1137
               *  Q  A  H  S  P  Q  L  S  A  R  S  Q  N  P  R  G  L  L  K


               K  A  K  T  G  G  A  G  K  F  R  R  G  K  *
Seq_1   1097 agaaagcaaagactggggggagcggggaagttccgaagggggaaaataa  1143
                 ||| |  ||||||     ||| |  | | |||||||||
Seq_2   1138 ------aaaacAAaggGGggaacgGGGGAAAttccaaGGGGAAA----  1174
                 N  K  G  G  N  G  E  I  P  R  G  X    X
```

7.1.6. [A159VmFEN1.seq] similarity: 1028/1028 (100 %) (Anti-parallel):

The sequence reaction was performed by the University of Sheffield Medical School. The mutation position is high-lighted (red). Nucleotides that could not be properly determined are denoted with a *.

```
          M  G  I  H  G  L  A  K  L  I  A  D  V  A  P  S  A  I  R  E
Seq_1  1  atgggaattcacggccttgccaaactaattgctgatgtggcccccagtgccatccgtgag  60
                     |||
Seq_2 1105 ------att---------------------------------------------------  1103
                      F

          N  D  I  K  S  Y  F  G  R  K  V  A  I  D  A  S  M  S  I  Y
Seq_1 61  aatgacatcaagagctactttggtcgcaaagtggccatcgatgcctccatgagcatctac  120
                                                                        |
Seq_2 1102 -----------------------------------------------------------c  1102

          Q  F  L  I  A  V  R  Q  G  G  D  V  L  Q  N  E  E  G  E  T
Seq_1 121 cagttcctgattgctgttcgtcagggtggggatgtgctgcagaacgaggagggtgagacc  180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2 1101 CAGtTcCTGATTGCTGTTcgTCAgGGTgGGGATGTGCTGCAGAACGAGGAGGGTGAGACC  1042
          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  Y  R  T  I  R  M  M  E  N  G  I  K
Seq_1 181 accagccacctgatgggcatgttctaccgtaccatccgcatgatggagaatggcatcaag  240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2 1041 acCAGCcacCTGATGGGCATGTTCTACCGTACCATCCGCATGATGGAGAATGGCATCAAG  982
          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  P  P  Q  L  K  S  G  E  L  A  K  R
Seq_1 241 cctgtgtacgtctttgatggcaaaccaccacagctgaagtcaggcgagctggccaagcgc  300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2 981 CCTGTGTACGTCTTTGATGGCAAACCACCACAGCTGAAGTCAGGCGAGCTGGCCAAGCGC  922
          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  E  A  G  M  E
Seq_1 301 agtgagaggcgcgccgaggctgagaagcaactgcagcaggctcaggaggctgggatggag  360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2 921 AGTGAGAGGCGCGCCGAGGCTGAGAAGCAACTGCAGCAGGCTCAGGAGGCTGGGATGGAG  862
          S  E  R  R  A  E  A  E  K  Q  L  Q  Q  A  Q  E  A  G  M  E

          E  E  V  E  K  F  T  K  R  L  V  K  V  T  K  Q  H  N  D  E
Seq_1 361 gaggaggtggagaagttcaccaagaggctcgtgaaggtcaccaagcaacacaatgatgag  420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2 861 GAGGAGGTGGAGAAGTTCACCAAGAGGCTCGTGAAGGTCACCAAGCAACACAATGATGAG  802
          E  E  V  E  K  F  T  K  R  L  V  K  V  T  K  Q  H  N  D  E

          C  K  H  L  L  S  L  M  G  I  P  Y  L  D  A  P  S  E  A  E
Seq_1 421 tgcaaacacctgctgagcctcatgggcatcccttaccttgatgcacccagcgaggcagag  480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Seq_2 801 TGCAAACACCTGCTGAGCCTCATGGGCATCCCTTACCTTGATGCACCCAGCGAGGTAGAG  742
          C  K  H  L  L  S  L  M  G  I  P  Y  L  D  A  P  S  E  V  E

          A  S  C  A  A  L  A  K  A  G  K  V  Y  A  A  A  T  E  D  M
Seq_1 481 gccagctgtgctgccctggcaaaggctggcaaagtctatgctgcggccacggaggacatg  540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2 741 GCCAGCTGTGCTGCCCTGGCAAAGGCTGGCAAAGTCTATGCTGCGGCCACGGAGGACATG  682
          A  S  C  A  A  L  A  K  A  G  K  V  Y  A  A  A  T  E  D  M

          D  C  L  T  F  G  S  P  V  L  M  R  H  L  T  A  S  E  A  K
Seq_1 541 gactgcctcacttttggcagccccgtgctaatgcgacacttaactgccagtgaggccaag  600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2 681 GACTGCCTCACTTTTGGCAGCCCCGTGCTAATGCGACACTTAACTGCCAGTGAGGCCAAG  622
          D  C  L  T  F  G  S  P  V  L  M  R  H  L  T  A  S  E  A  K

          K  L  P  I  Q  E  F  H  L  S  R  V  L  Q  E  L  G  L  N  Q
Seq_1 601 aagctgcccatccaagagttccatctgagccgcgtcctgcaggagctgggtctgaaccag  660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Seq_2   621  AAGCTGCCCATCCAAGAGTTCCATCTGAGCCGCGTCCTGCAGGAGCTGGGTCTGAACCAG  562
             K   L   P   I   Q   E   F   H   L   S   R   V   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
Seq_1   661  gagcagtttgtggatctgtgcatcctgctgggtagcgactactgcgagagcatccgtggc  720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2   561  GAGCAGTTTGTGGATCTGTGCATCCTGCTGGGTAGCGACTACTGCGAGAGCATCCGTGGC  502
             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   V
Seq_1   721  attgggcccaagcgggctgtggatctcatccagaaacataagagcatcgaggagatcgtg  780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2   501  ATTGGGCCCAAGCGGGCTGTGGATCTCATCCAGAAACATAAGAGCATCGAGGAGATCGTG  442
             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   S   K   Y   P   V   P   E   N   W   L   H   K   E   A   Q
Seq_1   781  aggcggctggaccccagcaagtaccccgttccagagaactggctccacaaggaagcccag  840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2   441  AGGCGGCTGGACCCCAGCAAGTACCCCGTTCCAGAGAACTGGCTCCACAAGGAAGCCCAG  382
             R   R   L   D   P   S   K   Y   P   V   P   E   N   W   L   H   K   E   A   Q

             Q   L   F   L   E   P   E   V   L   D   P   E   S   V   E   L   K   W   S   E
Seq_1   841  cagctcttcctggagccagaagtactggacccagagtctgtggagctgaagtggagcgag  900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2   381  CAGCTCTTCCTGGAGCCAGAAGTACTGGACCCAGAGTCTGTGGAGCTGAAGTGGAGCGAG  322
             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   V   K   F   M   C   G   E   K   Q   F   S   E   E   R
Seq_1   901  ccaaatgaagaagagttggtcaaatttatgtgtggtgaaaagcagttttctgaagagcga  960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2   321  CCAAATGAAGAAGAGTTGGTCAAATTTATGTGTGGTGAAAAGCAGTTTTCTGAAGAGCGA  262
             P   N   E   E   E   L   V   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   L
Seq_1   961  attcgcagtggggtcaagcggctgagtaagagccgccagggcagcacccagggacgcctc  1020
             |||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Seq_2   261  ATTCGCAGTGGGGTCAAGCGGCTGAGTAAGAGCCGCCAGGGTAGCACCCAGGGACGCCTC  202
             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   P
Seq_1  1021  gatgatttcttcaaggtgacaggctcactctcctcagctaagcgcaaggagccagaaccc  1080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2   201  GATGATTTCTTCAAGGTGACAGGCTCACTCTCCTCAGCTAAGCGCAAGGAGCCAGAACCC  142
             D   D   F   F   K   V   T   G   S   L   S   S   A   K   R   K   E   P   E   P

             K   G   P   A   K   K   K   A   K   T   G   G   A   G   K   F   R   R   G   K
Seq_1  1081  aaggggcctgctaagaagaaagcaaagactggggggagcggggaagttccgaagggggaaa  1140
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2   141  AAGGGGCCTGCTAAGAAGAAAGCAAAGACTGGGGGGAGCGGGGAAGTTCCGAAGGGGAAAA  82
             K   G   P   A   K   K   K   A   K   T   G   G   A   G   K   F   R   R   G   K

                                              *
Seq_1  1141  ------------------taa----------------------------------------  1143
                                              |||
Seq_2    81  CATCATCATCATCATCACTAAGCTTGCGGCCGCACTCGAGCACCACCACCACCACCACtG  22
             H   H   H   H   H   H   *   A   C   G   R   T   R   A   P   P   P   P   P   L

                              X   X
Seq_1  1144  --------------------  1143

Seq_2    21  AGATCCGGCtgctAaCAaagc  1
             R   S   G   C   *   Q   S
```

7.2. Reaction condition for multiple turnover (MT) assays of WT, A159V and E160D mFEN-1:

[S] (nM)	[S]Stock (μM)	[S] Stock (nM)	[S] (ul)	FB (ul)	MM (ul)	10X[E] (nM)	10X[E] (ul)	Final Enzyme (pM)	EDTA (ul)	[S] in tube (nM)
5000	50	50000	5	0	40	1.5	4.5	150	250	90
2500	50	50000	2.5	2.5	40	1.5	4.5	150	200	56.25
1000	50	50000	1	4	40	0.5	4.5	50	200	22.5
750	50	50000	0.75	4.25	40	0.3	4.5	30	150	22.5
500	50	50000	0.5	4.5	40	0.25	4.5	25	150	15
250	5	5000	2.5	2.5	40	0.1	4.5	10	100	11.25
100	5	5000	4	16	160	0.06	18	6	100	18
75	5	5000	3	17	160	0.04	18	4	100	13.5
50	0.5	500	20	0	160	0.04	18	4	100	9
25	0.5	500	10	10	160	0.02	18	2	50	9
10	0.5	500	4	16	160	0.02	18	2	50	3.6
5	0.5	500	2	18	160	0.02	18	2	50	1.8

Table 7.1. *Reaction conditions for MT reaction of WTmFEN-1 with DF.*

[S] (nM)	[S] Stock (μM)	[S] Stock (nM)	[S] (ul)	FB (ul)	MM (ul)	10X[E] (nM)	10X[E] (ul)	Final Enzyme (pM)	EDTA (ul)	[S] in tube (nM)
5000	50	50000	5	0	40	8	4.5	800	250	90
2500	50	50000	2.5	2.5	40	6	4.5	600	200	56.25
1000	50	50000	1	4	40	2	4.5	200	200	22.5
750	50	50000	0.75	4.25	40	1.5	4.5	150	150	22.5
500	50	50000	0.5	4.5	40	1.5	4.5	150	150	15
250	5	5000	2.5	2.5	40	0.5	4.5	50	100	11.25
100	5	5000	4	16	160	0.3	18	30	100	18
75	5	5000	3	17	160	0.3	18	30	100	13.5
50	0.5	500	20	0	160	0.3	18	30	100	9
25	0.5	500	10	10	160	0.2	18	20	50	9
10	0.5	500	4	16	160	0.2	18	20	50	3.6
5	0.5	500	2	18	160	0.2	18	20	50	1.8

Table 7.2. *Reaction conditions for MT reaction of WTmFEN-1 with 3'-SF.*

[S] (nM)	[S] Stock (µM)	[S] Stock (nM)	[S] (ul)	1xFB (ul)	MM (ul)	10X[E] (nM)	10X[E] (ul)	Final Enzyme (pM)	EDTA(ul)	[S] in tube (nM)
5000	50	50000	5	0	40	3	4.5	300	250	90
2500	50	50000	2.5	2.5	40	2	4.5	200	200	56.25
1000	50	50000	1	4	40	1	4.5	100	200	22.5
750	50	50000	0.75	4.25	40	0.5	4.5	50	150	22.5
500	50	50000	0.5	4.5	40	0.5	4.5	50	150	15
250	5	5000	2.5	2.5	40	0.3	4.5	30	100	11.25
100	5	5000	4	16	160	0.1	18	10	100	18
75	5	5000	3	17	160	0.1	18	10	100	13.5
50	0.5	500	20	0	160	0.1	18	10	100	9
25	0.5	500	10	10	160	0.06	18	6	50	9
10	0.5	500	4	16	160	0.03	18	3	50	3.6
5	0.5	500	2	18	160	0.03	18	3	50	1.8

Table 7.3. *Reaction conditions for MT reaction of WTmFEN-1 with gap.*

[S] (nM)	[S]Stock (µM)	[S] Stock (nM)	[S] (ul)	1xFB (ul)	MM (ul)	10X[E] (nM)	10X[E] (ul)	Final [E] (pM)	EDTA (ul)	[S] in tube (nM)
5000	50	50000	5	0	40	5	4.5	500	250	90
2500	50	50000	2.5	2.5	40	2	4.5	200	200	56.25
1000	50	50000	1	4	40	1	4.5	100	200	22.5
750	50	50000	0.75	4.25	40	1	4.5	100	150	22.5
500	50	50000	0.5	4.5	40	0.5	4.5	50	150	15
250	5	5000	2.5	2.5	40	0.25	4.5	25	100	11.25
100	5	5000	4	16	160	0.1	18	10	100	18
75	5	5000	3	17	160	0.1	18	10	100	13.5
50	0.5	500	20	0	160	0.06	18	6	100	9
25	0.5	500	10	10	160	0.04	18	4	50	9
10	0.5	500	4	16	160	0.03	18	3	50	3.6
5	0.5	500	2	18	160	0.03	18	3	50	1.8

Table 7.4. *Reaction conditions for MT reaction of WTmFEN-1 with DF2'OH.*

[S] (nM)	[S] Stock (μM)	[S] Stock (nM)	[S] (ul)	1XFB (ul)	MM (ul)	10X[E] (nM)	10X[E] (ul)	Final Enzyme (pM)	EDTA (ul)	[S] in tube (nM)
5000	50	50000	5	0	40	400	4.5	50000	250	90
2500	50	50000	2.5	2.5	40	250	4.5	50000	200	56.25
1000	50	50000	1	4	40	100	4.5	20000	200	22.5
750	50	50000	0.75	4.25	40	75	4.5	20000	150	22.5
500	50	50000	0.5	4.5	40	50	4.5	20000	150	15
250	5	5000	2.5	2.5	40	25	4.5	10000	100	11.25
100	5	5000	4	16	160	20	18	2000	100	18
75	5	5000	3	17	160	15	18	2000	100	13.5
50	0.5	500	20	0	160	10	18	2000	100	9
25	0.5	500	10	10	160	8	18	800	50	9
10	0.5	500	4	16	160	6	18	800	50	3.6
5	0.5	500	2	18	160	6	18	400	50	1.8

Table 7.5. *Reaction conditions for MT reaction of A159VmFEN-1 with DF.*

[S] (nM)	[S] Stock (μM)	[S] Stock (nM)	[S] (ul)	1XFB (ul)	MM (ul)	10X[E] (nM)	10X[E] (ul)	Final Enzyme (pM)	EDTA (ul)	[S] in tube (nM)
5000	50	50000	5	0	40	800	4.5	80000	250	90
2500	50	50000	2.5	2.5	40	500	4.5	50000	200	56.25
1000	50	50000	1	4	40	200	4.5	20000	200	22.5
750	50	50000	0.75	4.25	40	150	4.5	15000	150	22.5
500	50	50000	0.5	4.5	40	100	4.5	10000	150	15
250	5	5000	2.5	2.5	40	50	4.5	5000	100	11.25
100	5	5000	4	16	160	40	18	4000	100	18
75	5	5000	3	17	160	30	18	3000	100	13.5
50	0.5	500	20	0	160	20	18	2000	100	9
25	0.5	500	10	10	160	16	18	1600	50	9
10	0.5	500	4	16	160	12	18	1200	50	3.6
5	0.5	500	2	18	160	12	18	1200	50	1.8

Table 7.6. *Reaction conditions for MT reaction of A159VmFEN-1 with gap.*

[S] (nM)	[S] Stock (μM)	[S] Stock (nM)	[S] (ul)	1XFB (ul)	MM (ul)	10X[E] (nM)	10X[E] (ul)	Final Enzyme (pM)	EDTA (ul)	[S] in tube (nM)
5000	50	50000	5	0	40	30	4.5	3000	250	90
2500	50	50000	2.5	2.5	40	15	4.5	1500	200	56.25
1000	50	50000	1	4	40	10	4.5	1000	200	22.5
750	50	50000	0.75	4.25	40	5	4.5	500	150	22.5
500	50	50000	0.5	4.5	40	5	4.5	500	150	15
250	5	5000	2.5	2.5	40	4	4.5	400	100	11.25
100	5	5000	4	16	160	2	18	200	100	18
75	5	5000	3	17	160	1	18	100	100	13.5
50	0.5	500	20	0	160	1	18	100	100	9
25	0.5	500	10	10	160	0.5	18	50	50	9
10	0.5	500	4	16	160	0.5	18	50	50	3.6
5	0.5	500	2	18	160	0.5	18	50	50	1.8

Table 7.7. *Reaction conditions for MT reaction of A159VmFEN-1 with DF2'OH.*

[S] (nM)	[S] Stock (μM)	[S] Stock (nM)	[S] (ul)	FB (ul)	MM (ul)	10X[E] (nM)	10X[E] (ul)	Final Enzyme (pM)	EDTA (ul)	[S] in tube (nM)
5000	50	50000	5	0	40	12	4.5	1200	250	90
2500	50	50000	2.5	2.5	40	8	4.5	800	200	56.25
1000	50	50000	1	4	40	4.8	4.5	480	200	22.5
750	50	50000	0.75	4.25	40	4.8	4.5	480	150	22.5
500	50	50000	0.5	4.5	40	3.2	4.5	320	150	15
250	5	5000	2.5	2.5	40	1.6	4.5	160	100	11.25
100	5	5000	4	16	160	0.8	18	80	100	18
75	5	5000	3	17	160	0.64	18	64	100	13.5
50	0.5	500	20	0	160	0.64	18	64	100	9
25	0.5	500	10	10	160	0.64	18	64	50	9
10	0.5	500	4	16	160	0.64	18	64	50	3.6
5	0.5	500	2	18	160	0.64	18	64	50	1.8

Table 7.8. *Reaction conditions for MT reaction of E160DmFEN-1 with DF.*

[S] (nM)	[S] Stock (μM)	[S] Stock (nM)	[S] (ul)	FB (ul)	MM (ul)	10X[E] (nM)	10X[E] (ul)	Final Enzyme (pM)	EDTA (ul)	[S] in tube (nM)
5000	50	50000	5	0	40	20	4.5	2000	250	90
2500	50	50000	2.5	2.5	40	16	4.5	1600	200	56.25
1000	50	50000	1	4	40	7	4.5	700	200	22.5
750	50	50000	0.75	4.25	40	7	4.5	700	150	22.5
500	50	50000	0.5	4.5	40	5	4.5	500	150	15
250	5	5000	2.5	2.5	40	3	4.5	300	100	11.25
100	5	5000	4	16	160	1	18	100	100	18
75	5	5000	3	17	160	1	18	100	100	13.5
50	0.5	500	20	0	160	1	18	100	100	9
25	0.5	500	10	10	160	1	18	100	50	9
10	0.5	500	4	16	160	1	18	100	50	3.6
5	0.5	500	2	18	160	1	18	100	50	1.8

Table 7.9. Reaction conditions for MT reaction of E160DmFEN-1 with gap.

[S] (nM)	[S] Stock (μM)	[S] Stock (nM)	[S] (ul)	FB (ul)	MM (ul)	10X[E] (nM)	10X[E] (ul)	Final Enzyme (pM)	EDTA (ul)	[S] in tube (nM)
5000	50	50000	5	0	40	10	4.5	1000	250	90
2500	50	50000	2.5	2.5	40	6	4.5	600	200	56.25
1000	50	50000	1	4	40	3	4.5	300	200	22.5
750	50	50000	0.75	4.25	40	2	4.5	200	150	22.5
500	50	50000	0.5	4.5	40	2	4.5	200	150	15
250	5	5000	2.5	2.5	40	1	4.5	100	100	11.25
100	5	5000	4	16	160	0.6	18	60	100	18
75	5	5000	3	17	160	0.6	18	60	100	13.5
50	0.5	500	20	0	160	0.4	18	40	100	9
25	0.5	500	10	10	160	0.4	18	40	50	9
10	0.5	500	4	16	160	0.3	18	30	50	3.6
5	0.5	500	2	18	160	0.3	18	30	50	1.8

Table 7.10. Reaction conditions for MT reaction of E160DmFEN-1 with DF2'OH.

7.3. Reaction condition for single turnover (ST) assay of WT, A159V and E160D of mFEN-1:

Loop	Mode	Switch Position	Push 1	Push 2	Delay	Time (ms)	DF	3'-SF	gap	DF2'OH
1	C	5	160	/	/	4.5	✓	✓	✗	✗
1	C	2	160	/	/	12.1	✓	✓	✗	✗
2	C	6	180	/	/	9.1	✓	✓	✓	✓
2	C	3	180	/	/	19.4	✓	✓	✓	✓
2	C	2	180	/	/	27.6	✓	✓	✓	✓
2	C	1	180	/	/	41.8	✓	✓	✓	✓
4	C	5	255	/	/	30.6	✓	✓	✓	✓
4	C	4	255	/	/	40.8	✓	✓	✓	✓
4	C	3	255	/	/	57.5	✓	✓	✓	✓
4	C	2	255	/	/	82.1	✓	✓	✓	✓
4	C	1	255	/	/	124.2	✓	✓	✓	✓
4	I	4	160	95	100	140.8	✓	✓	✓	✓
4	I	4	160	95	200	240.8	✓	✓	✓	✓
4	I	4	160	95	400	440.8	✓	✓	✓	✓
4	I	4	160	95	800	840.8	✓	✓	✓	✓
4	I	4	160	95	1600	1640.8	✓	✓	✓	✓
4	I	4	160	95	3200	3240.8	✓	✓	✓	✓
4	I	4	160	95	6400	6440.8	✓	✓	✓	✓
4	I	4	160	95	9000	9040.8	✗	✗	✗	✓
4	/	/	/	X10	1280	12841	✓	✓	✓	✓
4	/	/	/	X10	1500	15041	✗	✗	✗	✓
4	/	/	/	X10	2560	25641	✓	✓	✓	✓
4	/	/	/	X10	3000	30041	✗	✗	✗	✓
4	/	/	/	X10	5120	51241	✓	✓	✓	✓
4	/	/	/	X10	6000	60041	✗	✗	✗	✓
4	/	/	/	X10	9000	90041	✗	✗	✗	✓
4	/	/	/	X10	9999	100030	✗	✗	✗	✓

Table 7.11. *Time points and reaction conditions for ST reaction of WTmFEN-1 with DF, 3'-SF, gap and DF2'OH.*

Loop	Mode	Switch Position	Push 1	Push 2	Delay	Time (ms)	DF	gap	DF2'OH
4	C	5	255	/	/	30.6	✗	✗	✓
4	C	4	255	/	/	40.8	✗	✗	✓
4	C	3	255	/	/	57.5	✗	✗	✓
4	C	2	255	/	/	82.1	✗	✗	✓
4	C	1	255	/	/	124.2	✗	✗	✓
4	I	4	160	95	100	140.8	✓	✓	✓
4	I	4	160	95	200	240.8	✓	✓	✓
4	I	4	160	95	400	440.8	✓	✓	✓
4	I	4	160	95	800	840.8	✓	✓	✓
4	I	4	160	95	1600	1640.8	✓	✓	✓
4	I	4	160	95	3200	3240.8	✓	✓	✓
4	I	4	160	95	6400	6440.8	✓	✓	✓
4	I	4	160	95	9000	9040.8	✗	✗	✓
4	/	/	/	X10	1280	12841	✓	✓	✓
4	/	/	/	X10	1500	15041	✗	✗	✓
4	/	/	/	X10	2560	25641	✓	✓	✓
4	/	/	/	X10	3000	30041	✗	✗	✓
4	/	/	/	X10	5120	51241	✓	✓	✓
4	/	/	/	X10	6000	60041	✗	✗	✓
4	/	/	/	X10	9000	90041	✗	✗	✓
4	/	/	/	X10	9999	100030	✗	✗	✓
/	/	/	/	/	/	120000	✓	✓	✓
/	/	/	/	/	/	240000	✓	✓	✓
/	/	/	/	/	/	360000	✓	✓	✓
/	/	/	/	/	/	480000	✓	✓	✓
/	/	/	/	/	/	600000	✓	✓	✓
/	/	/	/	/	/	720000	✓	✓	✗
/	/	/	/	/	/	840000	✓	✓	✗
/	/	/	/	/	/	960000	✓	✓	✗

Table 7.12. *Time points and reaction conditions for ST reaction of A159VmFEN-1 with DF, gap and DF2'OH.*

Loop	Mode	Switch Position	Push 1	Push 2	Delay	Time (ms)	DF	gap	DF2'OH
2	C	6	180	/	/	9.1	✗	✗	✓
2	C	3	180	/	/	19.4	✗	✗	✓
2	C	2	180	/	/	27.6	✗	✗	✓
2	C	1	180	/	/	41.8	✗	✗	✓
4	C	5	255	/	/	30.6	✓	✓	✓
4	C	4	255	/	/	40.8	✓	✓	✓
4	C	3	255	/	/	57.5	✓	✓	✓
4	C	2	255	/	/	82.1	✓	✓	✓
4	C	1	255	/	/	124.2	✓	✓	✓
4	I	4	160	95	100	140.8	✓	✓	✓
4	I	4	160	95	200	240.8	✓	✓	✓
4	I	4	160	95	400	440.8	✓	✓	✓
4	I	4	160	95	800	840.8	✓	✓	✓
4	I	4	160	95	1600	1640.8	✓	✓	✓
4	I	4	160	95	3200	3240.8	✓	✓	✓
4	I	4	160	95	6400	6440.8	✓	✓	✓
4	I	4	160	95	9000	9040.8	✗	✗	✓
4	/	/	/	X10	1280	12841	✓	✓	✓
4	/	/	/	X10	1500	15041	✗	✗	✓
4	/	/	/	X10	2560	25641	✓	✓	✓
4	/	/	/	X10	3000	30041	✗	✗	✓
4	/	/	/	X10	5120	51241	✓	✓	✓

Table 7.13. *Time points and reaction conditions for ST reaction of E160DmFEN-1 with DF, gap and DF2'OH.*

7.4. Reaction condition for ECCD experiments of WT and mutated mouse and human FEN-1:

MM (uL)	FB (1X)	SB	ssDNA	DNA	100uM E	500 mM EDTA	Total	Description
620	80	100	0	0	0	0	800	Ca Blank
						40	840	EDTA Blank
620	0	100	80	0	0	0	800	10 uM ssDNA, 10mM CaCl$_2$
						40	840	9.52 uM ssDNA, 9.52 mM CaCl$_2$, 23.8 mM EDTA
620	0	100	0	80	0	0	800	10 uM DNA, 10 mM CaCl$_2$
						40	840	9.52 uM DNA, 9.52 mM CaCl$_2$, 23.8 mM EDTA
620	0	0	0	80	100	0	800	10 uM DNA, 10 mM CaCl$_2$, 12.5 uM E
						40	840	9.52 uM DNA, 9.52 mM CaCl$_2$, 11.9 uM E, 12,1 mM EDAT

Table 7.14. Reaction conditions for ECCD reaction of free ssDNA and dsDNA with and without FEN-1 enzyme using JASCO CD spectrophotometer machine.

MM (uL)	FB (1X)	SB	ssDNA	DNA	100uM E	500 mM EDTA	Total	Description
387.5	50	62.5	0	0	0	0	500	Ca^{2+} Blank
						25	525	EDTA Blank
387.5	0	62.5	50	0	0	0	500	5 uM ssDNA, 5 mM CaCl$_2$
						25	525	4.88 uM ssDNA, 4.88 mM CaCl$_2$, 12.1 mM EDTA
387.5	0	62.5	0	50	0	0	500	5 uM DNA, 5 mM CaCl$_2$
						25	525	4.88 uM DNA, 4.88 mM CaCl$_2$, 12.1 mM EDTA
387.5	0	0	0	50	62.5	0	500	5uM DNA, 5mM CaCl$_2$, 12.5 uM E
						25	525	4.88 uM DNA, 4.88 mM CaCl$_2$, 11.9 uM E, 12.1 mM EDTA

Table 7.15. Reaction conditions for ECCD reaction of free ssDNA and dsDNA with and without FEN-1 enzyme using chirascan plus CD spectrophotometer machine.

7.5. Kinetic MT and ST results of mFEN-1:
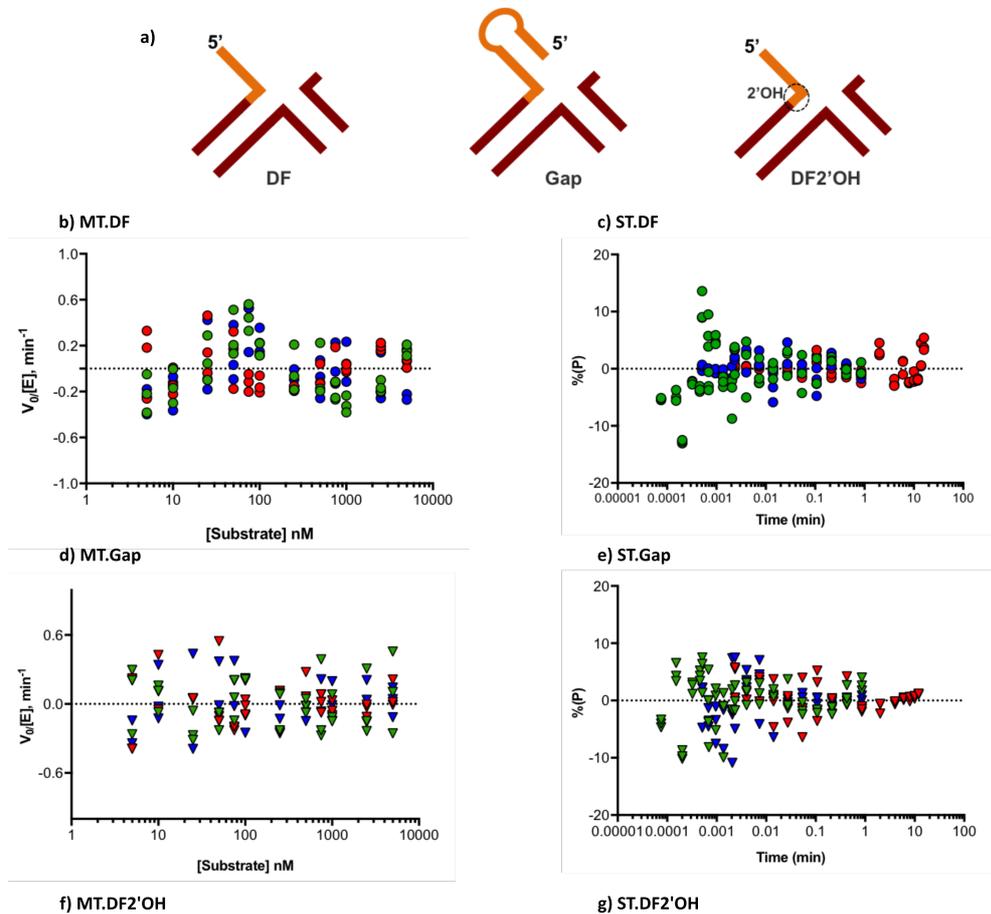
Figure 7.1. *Multiple and single turnover kinetic analysis of WT, A159V, and E160D mFEN1 upon different substrates. a) Cartoon of DF, gab and DF2'OH substrates that reacted with mFEN-1 to product minor product. b) d) and f) Comparison of weighted residual plots of Michaelis-Menten plots for cleavage of DF, gap, and DF2'OH substrates respectively by WT, A159V, and E160D mFEN1. c) e) and g) Comparison of residual plots of single turnover rate profiles for DF, gab and DF2'OH substrates respectively by WT, A159V, and E160D mFEN1.*

DNA and Protein Requirements for Substrate Conformational Changes Necessary for Human Flap Endonuclease-1-catalyzed Reaction*

Sana I. Algasaier[‡1], Jack C. Exell[‡1,2], Ian A. Bennet[‡], Mark J. Thompson[‡], Victoria J. B. Gotham[‡], Steven J. Shaw[‡],
Timothy D. Craggs[§], L. David Finger[‡], and Jane A. Grasby[‡3]

From the ‡Centre for Chemical Biology, Department of Chemistry, Krebs Institute, University of Sheffield, Sheffield S3 7HF,
United Kingdom and the §DNA:Protein Interactions Unit, School of Biochemistry, University of Bristol,
Bristol BS8 1TD, United Kingdom

Human flap endonuclease-1 (hFEN1) catalyzes the essential removal of single-stranded flaps arising at DNA junctions during replication and repair processes. hFEN1 biological function must be precisely controlled, and consequently, the protein relies on a combination of protein and substrate conformational changes as a prerequisite for reaction. These include substrate bending at the duplex-duplex junction and transfer of unpaired reacting duplex end into the active site. When present, 5′-flaps are thought to thread under the helical cap, limiting reaction to flaps with free 5′-termini *in vivo*. Here we monitored DNA bending by FRET and DNA unpairing using 2-aminopurine exciton pair CD to determine the DNA and protein requirements for these substrate conformational changes. Binding of DNA to hFEN1 in a bent conformation occurred independently of 5′-flap accommodation and did not require active site metal ions or the presence of conserved active site residues. More stringent requirements exist for transfer of the substrate to the active site. Placement of the scissile phosphate diester in the active site required the presence of divalent metal ions, a free 5′-flap (if present), a Watson-Crick base pair at the terminus of the reacting duplex, and the intact secondary structure of the enzyme helical cap. Optimal positioning of the scissile phosphate additionally required active site conserved residues Tyr[40], Asp[181], and Arg[100] and a reacting duplex 5′-phosphate. These studies suggest a FEN1 reaction mechanism where junctions are bound and 5′-flaps are threaded (when present), and finally the substrate is transferred onto active site metals initiating cleavage.

Flap endonuclease-1 (FEN1)[4] is an essential component of the DNA replicative and repair apparatus and the prototypical member of the 5′-nuclease superfamily (1–5). FEN1 removes single-stranded DNA or RNA flaps formed during DNA replication and repair as a result of strand displacement synthesis. Flapped DNAs arising in this context (*e.g.* adjacent Okazaki fragments) are equilibrating (*i.e.* migrating) structures that can have differing lengths of 5′- and 3′-single-strands, because all flaps are complementary to the continuous DNA template. However, FEN1 only processes one flapped DNA conformer, a two-way DNA junction bearing a single nucleotide (nt) 3′-flap and any length of 5′-flap (see Fig. 1, *A* and *B*) (6–8). FEN1 then catalyzes specific phosphate diester hydrolysis of the flapped DNA 1 nt into the double-strand, ensuring that the product is nicked DNA (see Fig. 1*A*). This exquisite specificity is necessary for the fidelity and efficiency of DNA replication and repair, because nicked DNA can be joined immediately by DNA ligase.

Extensive work has led to models for the origins of FEN1 reaction specificity that rely on key DNA conformational changes for substrate recognition and reaction site selection. The first selection is for two-way junction DNAs and involves the substrate bending 100° to contact two separate double-stranded DNA binding sites (see Fig. 1*B*) (7–10). One of these duplex binding sites forms a substrate-induced binding pocket that can only accommodate a 1-nt 3′-flap, which explains the preference for substrates with a single 3′-flap nucleotide.

The second requirement of hFEN1 specificity excludes the reaction of continuous single-stranded DNAs (*e.g.* template strand during replication) or flaps with bound protein. Although controversial (11), the 5′-flap is thought to pass through a hole in the protein above the active site and bordered by the helical cap (top of α4 and α5) and gateway (base of α4 and α2) (see Fig. 1, *B* and *D*) (1, 8, 12–14). The final specificity requirement is for reaction 1 nt into duplex, which is the hallmark of the 5′-nuclease superfamily that also includes the DNA repair proteins EXO1, XPG, and GEN1 (1). This selectivity is believed to involve a local DNA conformational change at the terminus of the reacting duplex (5, 8, 15–17), whereby two gating α-helices (bases of α2 and

SASBMB   VOLUME 291 • NUMBER 15 • APRIL 8, 2016

160

α4) appear to prevent access of duplex DNAs to the active site (8). It is proposed that the last two 5′ nucleotides of the reacting duplex unpair to place the scissile phosphate diester bond on the catalytic metal ions (see Fig. 1, *C* and *D*).

Although the overall conformational changes that FEN1 substrates must undergo before reaction have been deduced, the details of these processes are still not understood and in some cases remain controversial. Here, we aim to elucidate features of the FEN1 protein and substrates required for global DNA bending and local DNA unpairing (*i.e.* transfer to the active site). We also investigate the relationship of these processes to 5′-flap accommodation and explore the orientation of the 5′-portion of substrates that is not visible in current x-ray structures. Our combined results describe substrate and protein requirements for DNA bending and unpairing, and in turn Okazaki fragment processing, providing important insights into the FEN1 catalytic cycle.

Experimental Procedures

DNA Constructs—The oligonucleotide sequences are given in Table 1. DNA oligonucleotides including those containing 5′-FAM, 5′-biotin, internal TAMRA and fluorescein, and 2-aminopurine (2AP) substitutions were purchased with HPLC purification from DNA Technology A/S. The phosphoramidite synthons used for 5′-FAM, 5′-biotin, internal TAMRA dT, and internal fluorescein dT modifications were 6-carboxyfluorescein-aminohexyl amidite, *N*-DMT-biotinyl-2-aminoethoxy-ethanol amidite, 5′-DMT-T(TEG-TAMRA), and fluorescein T amidite, respectively, and were purchased from Biosearch Technologies Inc. 2AP was incorporated using 5′-(4,4′-dimethoxytrityl)-N^2-(dimethylformamidine)-2′-deoxypurine riboside-3′-[(2-cyanoethyl)-(*N,N*-diisopropyl)]phosphoramidite obtained from Link Technologies Ltd. DNA concentrations were determined by UV absorbance at 260 nm (20 °C) using extinction coefficients generated by the Integrated DNA Technologies oligo analyzer 3.1 tool.

Substrate constructs are summarized in Table 2. FRET substrates were designed by modeling a range of different fluorophore positions using the accessible volume approach (18) on both duplex and bent hFEN1 substrate DNAs (obtained by extending the existing DNA helixes in the crystal structure of hFEN1-DNA (8)). Labeling sites were chosen to maximize the FRET change upon bending. FRET substrates (Table 2) were assembled by heating the appropriate 3′-flap, 5′-flap/exo, and template strands in 1:1.1:1 ratio in 50 mM HEPES, pH 7.5, and 100 mM KCl to 80 °C for 5 min and then cooling to room temperature. For comparison, a DNA duplex was also created as above with Tcdonor (see Table 1) and template strands in a 1:1 ratio. 2AP constructs and the kinetic substrate KDF were formed by heating the appropriate exo/5′-flap strands with the complementary template in a 1:1.1 ratio at 80 °C for 5 min in 50 mM Tris-HCl, pH 7.5, and 100 mM KCl with subsequent cooling to room temperature.

Enzymes—hFEN1 and mutants were overexpressed and purified as described (8, 13).

Florescence Resonance Energy Transfer—FRET efficiencies (*E*) were determined using the (ratio)$_A$ method (19) by measuring the enhanced acceptor fluorescence at 37 °C. The steady state fluorescent spectra of 10 nM nonlabeled (NL) trimolecular, donor-only labeled (DOL), and doubly labeled (DAL) DNA substrates (Table 2) were recorded using a Horiba Jobin Yvon FluoroMax-3® fluorometer. For direct excitation of the donor (fluorescein, DOL) or acceptor (TAMRA, AOL), the sample was excited at 490 or 560 nm (2-nm slit width), and the emission signal was collected from 515–650 or 575–650 nm (5-nm slit width). Emission spectra were corrected for buffer and enzyme background signal by subtracting the signal from the nonlabeled (NL) DNA sample. In addition to 10 nM of the appropriate DNA construct, samples contained 10 mM CaCl$_2$ or 2 mM EDTA, 110 mM KCl, 55 mM HEPES, pH 7.5, 0.1 mg/ml bovine serum albumin, and 1 mM DTT. The first measurement was taken prior to the addition of protein with subsequent readings taken on the cumulative addition of the appropriate enzyme in the same buffer, with corrections made for dilution. Transfer efficiencies (*E*) were determined according to Equations 1–3, where F_{DA} and F_D represent the fluorescent signal of the doubly labeled DNA (DAL) and donor-only labeled DNA (DOL) at the given wavelengths, respectively (*e.g.* $F_{DA}(\lambda_{EX}^D, \lambda_{EM}^A)$ denotes the measured fluorescence of acceptor emission upon excitation of the donor, for DAL DNA); ϵ^D and ϵ^A are the molar absorption coefficients of donor and acceptor at the given wavelengths; and $\epsilon^D(490)/\epsilon^A(560)$ and $\epsilon^A(490)/\epsilon^A(560)$ are determined experimentally from the absorbance spectra of doubly labeled molecules (DAL) and the excitation spectra of singly TAMRA-only labeled molecules (AOL), respectively. Energy transfer efficiency (*E*) was fitted by nonlinear regression in the Kaleidagraph program to Equation 4, where E_{max} and E_{min} are the maximum and minimum energy transfer values, [S] is the substrate concentration, [P] is the protein concentration, and K_{bend} is the bending equilibrium dissociation constant of the protein substrate [PS] complex. All experiments were repeated in triplicate.

$$E = (\text{ratio})_A \bigg/ \left(\frac{\epsilon^D(490)}{\epsilon^A(560)}\right) - \left(\frac{\epsilon^A(490)}{\epsilon^A(560)}\right) \qquad \text{(Eq. 1)}$$

where

$$(\text{ratio})_A = \frac{F_{DA}(\lambda_{EX}^D, \lambda_{EM}^A) - N \cdot F_D(\lambda_{EX}^D, \lambda_{EM}^A)}{F_{DA}(\lambda_{EX}^A, \lambda_{EM}^A)} \qquad \text{(Eq. 2)}$$

and

$$N = F_{DA}(\lambda_{EX}^D, \lambda_{EM}^D)/F_D(\lambda_{EX}^D, \lambda_{EM}^D) \qquad \text{(Eq. 3)}$$

$$E = E_{min} + \frac{(E_{max} - E_{min})}{2[S]}[([S] + [P] + K_{bend}) - \sqrt{([S] + [P] + K_{bend})^2 - 4[S][P]}] \qquad \text{(Eq. 4)}$$

Donor (fluorescein) was excited at 490 nm with emission sampled as the average value of the signal between 515 and 525 nm, and acceptor (TAMRA) was excited at 560 nm with emission averaged between 580 and 590 nm. For FRET experiments involving substrate bound to streptavidin, 5 molar equivalents of streptavidin were preincubated with the biotinylated substrate in buffer containing 10 mM CaCl$_2$, 55 mM HEPES, pH 7.5, 110 mM KCl, 1 mg/ml BSA, and 1 mM DTT for 10 min at room temperature before proceeding as above.

161
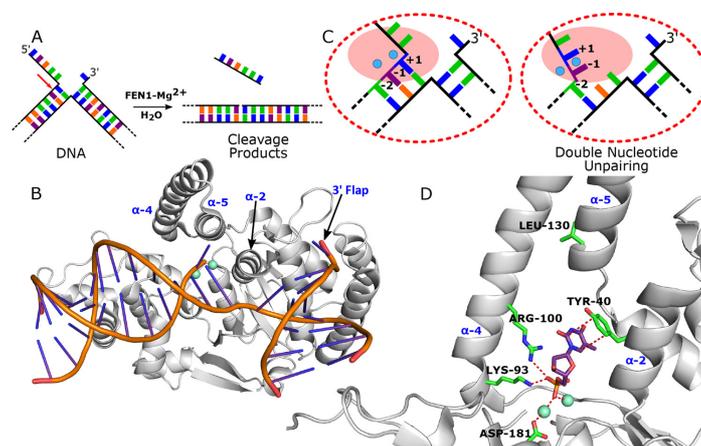
FIGURE 1. **FEN1 DNA bending and double nucleotide unpairing.** *A*, schematic of the FEN1 catalyzed hydrolysis of a double flap DNA yielding single-stranded DNA and double-stranded nicked DNA products. An *arrow* indicates the site of reaction. Each nucleobase is represented by a different color. *B*, hFEN1-product complex (Protein Data Bank code 3q8k) showing 100° bent DNA. *C*, schematic of double nucleotide unpairing proposed to position the scissile phosphodiester bond between the +1 and −1 nt on active site (*pink*) metal ions (*cyan*). *D*, cartoon representation of the active site in the FEN1-product structure (Protein Data Bank code 3q8k) showing the phosphate monoester of the unpaired −1 nt in contact with metal ions (*cyan*) and helical gateway (base $\alpha2$-$\alpha4$) and cap (top of $\alpha4$ and $\alpha5$) residues mutated in this study.

Determination of the Maximal Single Turnover Rate of Reaction (k_{STmax})—Maximal single turnover rates of reaction were determined using the KDF substrate (Table 2) and rapid quench apparatus (for WT-hFEN1 and Y40A) or manual sampling (for D181A) at 37 °C and pH 7.5, as described (20).

CD Spectroscopy—Samples containing 10 μM of the appropriate $(2AP)_2$ DNA construct (Table 2), 50 mM Tris-HCl, pH 7.5, 100 mM KCl, 1 mM DTT and, where appropriate, 12.5 μM protein and either 10 mM CaCl$_2$ or 10 mM CaCl$_2$ + 25 mM EDTA were prepared with subsequent acquisition of CD spectra (300–480 nm) at 20 °C using a JASCO J-810 CD spectrophotometer as described in detail (17). The CD spectra were plotted as $\Delta\epsilon$ per mol 2AP residue *versus* wavelength. Each measurement was independently repeated typically in triplicate.

Results

Global DNA Conformational Change

Substrate Design for DNA Bending—To study global conformational change of DNA substrates (Fig. 1*B*), we used FRET to detect duplex-duplex bending upon binding to human FEN1 (hFEN1) (Fig. 2) (7, 9, 10). Donor- and acceptor-labeled (DAL) substrates were assembled from three oligonucleotides, a TAMRA-labeled template strand, a fluorescein-labeled 3′-flap strand and an unlabeled 5′-flap/exo strand (Tables 1 and 2 and Fig. 2*A*). The positions of the fluorophores were chosen to maximize the FRET change observed upon substrate bending. In addition, donor only labeled (DOL), acceptor-only labeled (AOL), and nonlabeled (NL) versions of the substrates were also prepared (Table 2) to determine FRET efficiencies using the (ratio)$_A$ method (19). In double flap (DF) FRET substrates, the 5′-flap strand carried a terminal 5′-biotin to facilitate experiments with streptavidin; this label did not affect FRET

behavior (data not shown). To reduce any ambiguity in interpretation of our results, all substrates used in our studies were designed to be static (*i.e.* the flaps were noncomplementary to the template strand). Such static flaps permit clearer interpretation of experimental data but are known to behave identically to their equilibrating counterparts in hFEN1 reactions (6). For comparison, we also created the equivalent DAL duplex to the flapped DNAs (Table 2 and Fig. 2*A*).

To prevent cleavage of the substrate, all experiments were carried out in the absence of the viable cofactor Mg^{2+}. Because divalent metal ions are required for accommodation of the 5′-flap and for DNA conformational changes that lead to reaction (12, 17), we carried out experiments with or without catalytically nonviable Ca^{2+} ions, allowing us to investigate the relationship between DNA bending and other events of the hFEN1 catalytic cycle. Calcium ions are competitive inhibitors of Mg^{2+}-supported 5′-nuclease reactions, implying they occupy similar sites in the protein (15, 22); they have also been shown to facilitate 5′-flap threading and local DNA conformational changes (12, 17). Analysis of samples after both FRET and later CD experiments demonstrated negligible extent of reaction under all the conditions used (data not shown).

Catalytically Important Active Site Features Are Not Required for DNA Junction Bending—The FRET efficiency of DF(DAL) alone was similar ±Ca^{2+} (0.23–0.25) but was significantly greater than the corresponding duplex (0.1) (Fig. 2*D*). This indicates that the DF substrate has an overall conformation that is more bent than duplex DNA, even before the addition of protein. This is in line with single molecule observations where a double flap was seen to sample both a linear stacked and a bent conformation (9). Sequential addition of WT hFEN1

162

FIGURE 2. FRET data showing DNA bending on complexation with hFEN1 and mutants. *A*, schematic of double flap (DF, endonucleolytic) and single flap (SF, exonucleolytic) DNA constructs (Table 2) used in FRET studies, donor = fluorescein (*blue*) and acceptor = TAMRA (*red*). Nonlabeled (NL), donor only (DOL), acceptor-only (AOL), and donor and acceptor (DAL) versions of these constructs were used. *B*, variation in energy transfer efficiency of DF(DAL) upon addition of WT hFEN1 measured at pH 7.5 and 37 °C in the presence of Ca^{2+} ions (*blue*) or EDTA (*red*) fitted to Equation 4. *C*, derived (Equation 4) values of K_{bend} for the DF (double flap) and SF (single flap) substrates (Table 2) with WT and mutated hFEN1s as indicated in Ca^{2+} (*purple*) and EDTA (*pink*). MMDF contained a +1 mismatch, 5'-hydroxyl single (3') flap (*HOSF*) lacked a 5'-phosphate, and SADF had a 5'-conjugated streptavidin. Standard errors from repeat experiments are shown. *D*, derived (Equation 4) minimum (E_{min}) and maximum (E_{max}) energy transfer in Ca^{2+} (*blue*) and EDTA (*orange*) corresponding to the indicated protein with DF (double flap) or SF (single flap) substrates as in *C*. Duplex DNA was measured for comparison without protein in Ca^{2+}-containing buffer. Standard errors from repeat experiments are shown.

TABLE 1

Sequences of oligonucleotides used to construct substrates for FRET, kinetic, and ECCD experiments

A (in red), 2-aminopurine; **Bio**, biotin; **TAMRA**, tetramethylrhodamine; **Fluor**, internal fluorescein; **FAM**, fluorescein; **p**, phosphate.

Oligo	Sequence
F1	5'-FAM TTT TTA CAA GGA CTG CTC GAC AC-3'
T1	5'-GTG TCG AGC AGT CCT TGT GAC GAA GTC GTC C-3'
TEMP1	5'-CAC TCT GCC TCT TGA CAG CGA AGC TGT CC-3'
TEMP2	5'-CAC TCT GCC TTT CGA CAG CGA AGC TGT CC-3'
ssSF$_{+1-1}$	5'-**pAAG** AGG CAG AGT G-3'
ssHO-SF$_{+1-1}$	5'-**AAG** AGG CAG AGT G-3'
ssSF$_{-1-2}$	5'-**pGAA** AGG CAG AGT G-3'
ssHO-SF$_{-1-2}$	5'-**GAA** AGG CAG AGT G-3'
ssDF$_{+1-1}$	5'-TTT T**TA A**GA GGC AGA GTG-3'
ssDF$_{-1-2}$	5'-TTT TTG **AAA** GGC AGA GTC-3'
ssMM$_{+1}$DF$_{-1-2}$	5'-TTT TTC **AAA** GGC AGA GTG-3'
BssDF$_{-1-2}$	5'-**(Bio)**TTT TTT TTT TG**A A**AG GCA GAG TG-3'
Tcaccept	5'-GGT CC(**TAMRA**dT)A CTC TGC CTC AAG ACG GTC TGC TGC ACT GG-3'
Tcdonor	5-CCA G(**Fluor**dT)G CAG CAG ACC GTC C-3'
Tcflap	5'-**(5Bio)**TTT TTT TTG AGG CAG AGT AGG ACC-3'
Tcca	5'-GGT CCT ACT CTG CCT CAA GAC GGT CTG CTG CAC TGG-3'
Tccd	5'-CCA GTG CAG ACC GTC C-3'
EP-Fret	5'-**p**TTG AGG CAG AGT AGG ACC-3'
EO-Fret	5'-TTG AGG CAG AGT AGG ACC-3'
MM$_{+1}$-Fret	5'-**p**TTT TTT ATG AGG CAG AGT AGG ACC-3'
TcdonorD	5'-CCA G(**Fluor**dT)G CAG CAG ACC GTC TTG AGG CAG AGT AGG ACC-3'

TABLE 2

Oligonucleotide combinations used to make the substrate constructs for FRET, kinetic, and ECCD experiments

Construct	Composition
KDF	F1 + T1
SF$_{+1-1}$	TEMP1 + ssSF$_{+1-1}$
HO-SF$_{+1-1}$	TEMP1 + ssHO-SF$_{+1-1}$
DF$_{+1-1}$	TEMP1 + ssDF$_{+1-1}$
BDF$_{-1-2}$	TEMP2 + BssDF$_{-1-2}$
SF$_{-1-2}$	TEMP2 + ssSF$_{-1-2}$
HO-SF$_{-1-2}$	TEMP2 + ssHO-SF$_{-1-2}$
MM$_{+1}$DF$_{-1-2}$	TEMP2 + ssMM$_{+1}$DF$_{-1-2}$
DF (NL)	Tcflap + Tcca + Tccd
DF (DOL)	Tcflap + Tcca + Tcdonor
DF (AOL)	Tcflap + Tccd + Tcaccept
DF (DAL)	Tcflap + Tcdonor + Tcaccept
SF (NL)	EP-Fret + Tcca + Tccd
SF (DOL)	EP-Fret + Tcca + Tcdonor
SF (AOL)	EP-Fret + Tccd + Tcaccept
SF (DAL)	EP-Fret + Tcdonor + Tcaccept
HO-SF (NL)	EO-Fret + Tcca + Tccd
HO-SF (DOL)	EO-Fret + Tcca + Tcdonor
HO-SF (AOL)	EO-Fret + Tccd + Tcaccept
HO-SF (DAL)	EO-Fret + Tcdonor + Tcaccept
MM$_{+1}$-DF (NL)	MM$_{+1}$-Fret + Tcca + Tccd
MM$_{+1}$-DF (DOL)	MM$_{+1}$-Fret + Tcca + Tcdonor
MM$_{+1}$-DF (AOL)	MM$_{+1}$-Fret + Tccd + Tcaccept
MM-DF (DAL)	MM-Fret + Tcdonor + Tcaccept
Duplex (DOL)	TcdonorD + Tcca
Duplex (DAL)	TcdonorD + Tcaccept

to DF(DAL) produced an increase in corrected FRET signal until a plateau was reached at saturating protein, regardless of whether divalent ions were present or not (Fig. 2, *B* and *D*).

163

DNA Conformational Changes for FEN1 Catalysis

When DF (DAL) was fully bound to hFEN1 (FRET efficiency at end point), a slightly higher energy transfer value was reproducibly observed with Ca^{2+} ions present (Fig. 2, *B* and *D*). The origin of this end point difference is unknown. Nevertheless, the derived equilibrium dissociation constants $K_{bend} \pm Ca^{2+}$ only varied by a factor of two (13 ± 1.7 nM with Ca^{2+}, 21 ± 1.4 nM without), implying that the presence of divalent ions is not required for DNA to adopt a bent conformation when bound to hFEN1 (Fig. 2, *B* and *C*). Because divalent ions are required for the threading of 5′-flaps (12) and the transfer of the scissile phosphodiester to the active site (17), these results suggest that the DF substrate binds with similar affinity regardless of whether either of these conformational changes have taken place. This is consistent with the crystal structure that shows that most of the interaction surface area is with the duplex portions of the substrate (8).

To investigate the requirements for bending of DF DNA, we also tested mutated hFEN1s K93A, R100A, K93A/R100A, L130P, Y40A, and D181A (Fig. 1*D*). Superfamily conserved residues Lys[93] and Arg[100] are located at the base of α4 forming part of the hFEN1 helical gateway (8) from where they protrude into the hFEN1 active site and are not predicted to be involved in substrate interactions until the DNA is positioned to react. Leu[130] is a component of the helical cap (α5) and is removed from the active site, although the mutation L130P is presumed to interfere with formation of the secondary structure of the cap (13). Tyr[40] is an α2 gateway residue seen to interact with the +1 nucleobase (numbered relative to the scissile phosphate diester; Fig. 1*C*) of the DNA substrate when base-paired (8), whereas it stacks on the −1 nucleobase after reaction as seen in hFEN1 product structures (Fig. 1*D*). Asp[181] is an active site carboxylate in direct contact with the catalytic metal ions in hFEN1 structures (8). Mutation of Asp[181] may alter the number of metal ions bound and/or their precise positioning. Earlier studies have shown that under maximal single turnover conditions, the mutations K93A, R100A, K93A/R100A, and L130P decrease the rate of the hFEN1 reaction by factors of at least 2,000 (12, 13). To determine the effects of the Y40A and D181A mutations, we measured the maximal single turnover rate constants (k_{STmax}) using KDF substrate (Tables 1 and 2) and compared them with the WT protein (k_{STmax} = 740 min^{-1}) (data not shown). For Y40A, k_{STmax} = 7.91 ± 0.01 min^{-1}, and for D181A, k_{STmax} = 0.075 ± 0.003 min^{-1}, corresponding to rate decreases of 10^2 and 10^4, respectively. Thus, all the mutations studied have substantive and in most cases, very severe impacts on hFEN1 catalysis.

DF (DAL) adopted a bent conformation when bound to all the mutated proteins as seen by an increase in FRET signal upon addition of hFEN1. As with the wild type protein, only subtle variations in K_{bend} were observed with and without divalent metal ions (2-fold at most) (Fig. 2*C*). The exception was Y40A, where mutation stabilized the hFEN1-DNA complex in the presence of EDTA. Only small changes in K_{bend} were observed relative to the WT protein $\pm Ca^{2+}$ (less than 3-fold at most), indicating that none of the mutated residues are critical to DNA binding and bending. Like the WT protein, differences between the FRET efficiency at the end point $\pm Ca^{2+}$ were also observed with Y40A, R100A, K93A/R100A, and L130P with

titrations in Ca^{2+} buffer producing a higher value (Fig. 2*D*). In contrast, the end points with D181A and K93A remained constant $\pm Ca^{2+}$. Notably, all the altered FEN1 proteins have K_{bend} values in the low nanomolar range $\pm Ca^{2+}$, demonstrating that they will all fully bind substrate under the conditions of the local DNA unpairing (2AP)$_2$ CD experiments described later (12.5 μM protein, 10 μM DNA).

A Mismatch at the +1 Position of the Substrate Does Not Prevent Bending—Previously, we showed that double-flap substrates bearing a mismatch at the +1 position (numbering relative to scissile phosphodiester bond in the 5′-flap/exo strand; Fig. 1*C*) produced reduced reaction rates and reduced reaction-site specificity (16). This shows that the DNA base pair integrity at the +1 position is a requirement for optimal hFEN1 reaction. To determine whether a mismatch at +1 affects the ability to bind and bend substrate DNA, we prepared the appropriate construct MMDF(DAL) (Fig. 2*A*) and performed the same FRET measurements (Fig. 2*C*). Like the alteration of conserved active site residues, the presence of a mismatch at the +1 position does not prevent bending, but it does weaken substrate affinity 4–5-fold (in Ca^{2+} DF K_{bend} = 13 ± 1.7 nM, MMDF K_{bend} = 58 ± 6.8 nM).

A 5′-Flap Is Not Required for DNA Bending—An initial conundrum in the reactions of 5′-nucleases concerned their ability to carry out both endonucleolytic reactions on substrates that possessed 5′-flaps and 5′-exonucleolytic reactions on substrates that lacked such flaps. To test whether the absence of 5′-flap altered the stability of hFEN1-DNA complexes, we carried out a FRET experiment with a single flap substrate (SF(DAL)) that lacked the 5′-flap (Fig. 2*A*). Consistent with the crystal structure and the fact that hFEN1 reaction is susceptible to dsDNA (nicked) product inhibition (8, 20), the absence of a 5′-flap did not significantly alter the stability of the complex or the ability to bend (K_{bend} = 12 ± 1.1 nM with Ca^{2+}, 20 ± 2.1 nM without) (Fig. 2*C*). This is also consistent with similar K_m values observed earlier for exonucleolytic substrates bearing a 3′-flap compared with double flaps (20). However, the dissociation constant of SF substrate was sensitive to the status of the 5′-terminus. HO-SF (DAL), which lacked a 5′-phosphate monoester, was bound an order of magnitude more weakly by the protein in the presence of Ca^{2+} ions, and binding was also altered in EDTA to a lesser extent (Fig. 2*C*). This suggests that the 5′-phosphate forms an interaction with the protein facilitated by the local DNA conformational changes that occur in the presence of Ca^{2+} ions. Nevertheless, even HO-SF(DAL) would be fully bound to the protein under the conditions used to probe local DNA conformational changes by CD below. Like DF(DAL), SF(DAL) and HO-SF(DAL) also had a greater FRET value in the absence of protein (0.19–0.21) than the corresponding duplex (0.1), suggesting that the SF substrates can adopt a bent conformation in the absence of protein (Fig. 2*D*).

Accommodation of the 5′-Flap Is Not Required for DNA Bending—Although FEN1 substrates correctly positioned to react have yet to be observed crystallographically, it is suggested that the 5′-flap departs from the active site passing underneath the helical cap through the hole created by the cap (top of α4 and α5) and gateway (base of α4 and α2) (Fig. 1, *B* and *D*) (1, 8, 12–14). Evidence for this so-called threading hypothesis came

164

from experiments where streptavidin is added to 5′-biotin-labeled substrates before or after binding to the protein (12, 14). Prior conjugation—assumed to "block" substrate threading—severely retards FEN1 action, but conjugation to preformed DNA-protein complex does not affect the reaction rate. Furthermore, only this latter "trapped" substrate cannot exchange with competitor DNA.

We wished to ascertain whether, when present, accommodation of the 5′-flap is necessary for global substrate bending. A 5′-strepavidin complex with DF(DAL) (12) was used (blocked SADF) and showed a higher FRET efficiency in the absence of protein (Fig. 2*D*). This suggests a more bent overall conformation than uncomplexed DNA, likely because of the presence of a bulky streptavidin homotetramer conjugated to the 5′-terminus. Nevertheless, the blocked SADF with hFEN1-Ca^{2+} had a similar FRET efficiency at end point as the unmodified substrate, albeit with a 5-fold increase in K_{bend} (Fig. 2*C*). This result demonstrates that accommodation of the 5′-flap underneath the helical cap is not required for global substrate bending.

Local DNA Conformational Change of the Reacting Duplex

A Substrate 5′-Flap Is Not Required for Local DNA Conformational Change—In hFEN1-product structures, the −1 nt is unpaired and extrahelical (Fig. 1, *B* and *D*) such that its 5′-phosphate monoester contacts active site metal ions, whereas the adjacent −2 nt remains base-paired (8) (numbering of 5′-flap/exo strand, (Fig. 1*C*)). In contrast, structures of hFEN1-substrate DNA, where the substrate has no 5′-phosphate monoester, showed a base-paired substrate close to but not in the active site. Thus, it was deduced that 2 nt of the substrate unpair to allow the scissile phosphate to contact active site ions. We previously studied this local DNA conformational change using substrate or product constructs labeled with tandem 2APs at the −1 and −2 positions (DF$_{-1-2}$ and P$_{-1-2}$, respectively) (17). An exciton coupling between the adjacent 2APs produces a signal in the low energy region of the CD spectrum, the magnitude of which varies depending upon the relative orientation of the electronic transition dipole moments of the nucleobases. This exciton-coupled CD (ECCD) signal is readily followed because it is partially visible in a region of the spectrum where unmodified DNA bases and protein are transparent (24, 25). When either DF$_{-1-2}$ or P$_{-1-2}$ was bound to hFEN1 in EDTA buffer, a strong ECCD signal was observed (λ_{max} = 326 nm), consistent with the 2APs remaining stacked in the duplex. In the presence of hFEN1-Ca^{2+}, the signal was dramatically reduced to nearly zero. This was deduced to reflect the DNAs adopting a conformation of the kind seen in the product crystal structure, with transfer of the 5′-nucleotide of product, or the +1 and −1 nt of substrate, to the active site (Fig. 1*D*).

By analogy to these earlier experiments, 2APs were located at the −1 and −2 positions of a SF substrate (SF$_{-1-2}$) to test whether exonucleolytic substrates lacking the 5′-flap were also unpaired by hFEN1-Ca^{2+} (Tables 1 and 2 and Fig. 3). As seen previously, the ECCD signal of the isolated (2AP)$_2$ single-strand (ssSF$_{-1-2}$) was increased in magnitude, and the maximum was red-shifted to 326 nm upon forming the duplex SF$_{-1-2}$ (Fig. 3*A*) (17). On addition of hFEN1-Ca^{2+} to this substrate, the sig-

nal was dramatically reduced to near zero (Fig. 3, *A* and *B*). This behavior is similar to that observed earlier with DF$_{-1-2}$ or P$_{-1-2}$ (17). When EDTA was added to the hFEN1-Ca^{2+}·SF$_{-1-2}$ sample, a strong ECCD signal at 326 nm was restored. This demonstrates that the 5′-flap is not required for a change in respective orientation of the −1 and −2 nt, while confirming the presence of active site divalent metal ion(s) is essential. Moreover, both exonucleolytic (SF) and endonucleolytic (DF) substrates undergo analogous local DNA conformational changes.

FEN1 Conserved Residues Are Not Required for −1−2 Local DNA Conformational Change—Similar experiments were conducted with SF$_{-1-2}$ and mutant hFEN1 proteins. Fig. 3*B* shows the magnitude of the ECCD signal at 326 nm for each mutated protein ±Ca^{2+}. K93A, R100A, Y40A, and K93A/R100A were all capable of effecting local conformational change of SF$_{-1-2}$ in the presence of Ca^{2+}, with K93A most closely matching the spectra obtained with WT protein in Ca^{2+}. As seen previously with DF$_{-1-2}$ (13, 17), spectra of SF$_{-1-2}$ produced by R100A, Y40A, and K93A/R100A with Ca^{2+} contained an additional minimum at 310 nm (data not shown). This suggests an altered orientation of the −1 and −2 nt to that produced by WT and K93A hFEN1s. We found that D181A-Ca^{2+} was able to bring about an analogous conformational change to WT protein (Fig. 3*B*), which was surprising given that no active site metal ions were visible in an x-ray structure of D181A bound to SF DNA substrate in the presence of Ca^{2+} and the DNA remained base-paired (8). In contrast, the ECCD signal at 326 nm with L130P was similar ±Ca^{2+}, indicating that this protein does not facilitate the local DNA conformational change. Together, these results demonstrate that conserved residues are not required to bring about a change in the orientation of the −1 and −2 nt in exonucleolytic DNA substrates, although the intact secondary structure of the helical cap is. The results obtained with mutated hFEN1s strongly resemble those previously obtained with DFs (13, 17), underscoring that there are no overall differences between the behaviors of exonucleolytic (without 5′-flap) and endonucleolytic (with 5′-flap) hFEN1 substrates.

A 5′-Phosphate Is Not Required for Local DNA Conformational Change Monitored at the −1 and −2 nt—In the exonucleolytic FEN1 substrate SF$_{-1-2}$, the +1 nt has a terminal 5′-phosphate, whereas the double flap substrate DF$_{-1-2}$ has a 5′-phosphate diester (followed by the flap) in the corresponding position. Both substrates underwent a similar local DNA conformational change when bound by hFEN1-Ca^{2+}. A SF substrate lacking a 5′-phosphate (*i.e.* 5′-OH) crystallized with hFEN1 in base-paired form, despite the presence of active site metal ions (8). Furthermore, we previously reported that SF substrates lacking the 5′-phosphate monoester showed a 20-fold decrease in reaction efficiency, and we hypothesized that this was due to the inability to affect the local conformational change. To test whether the 5′-phosphate monoester is required for reorientation of the −1 and −2 nt, we created a substrate lacking the 5′-phosphate, HO-SF$_{-1-2}$. Surprisingly, we observed that this substrate underwent a change in orientation of the 2APs upon addition of hFEN1-Ca^{2+} with the signal

165
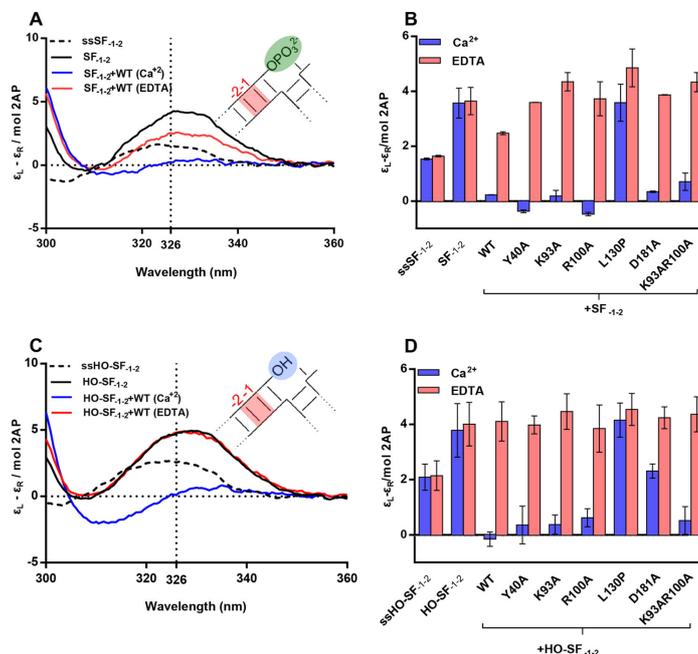
FIGURE 3. **hFEN1 and mutant mediated conformational change of 2AP-containing single flap SF$_{-1-2}$ monitored by ECCD.** All measurements were carried out at 20 °C and pH 7.5. *A*, divalent metal ion-dependent reduction in 2AP exciton coupling signal occurred when substrate SF$_{-1-2}$ was bound to hFEN1, indicative of local substrate conformational change. Unbound SF$_{-1-2}$ (*black*), the corresponding single strand (ssSF$_{-1-2}$, *dashed line*) and SF$_{-1-2}$ bound to hFEN1 (*blue*) all in Ca^{2+}-containing buffer. SF$_{-1-2}$ bound to hFEN1 in buffer containing 25 mM EDTA (*red*). *B*, comparison of molar ellipticity per 2AP residue at 326 nm of SF$_{-1-2}$ bound to WT- and mutant hFEN1s in Ca^{2+} (*purple*) and EDTA (*pink*) buffers. Standard errors from repeat experiments are shown. *C*, divalent metal ion-dependent reduction in 2AP exciton coupling signal occurred when substrate HO-SF$_{-1-2}$, which lacks a 5'-phosphate, was bound to hFEN1, indicative of local substrate conformational change. Unbound HO-SF$_{-1-2}$ (*black*), the corresponding single strand (ssHO-SF$_{-1-2}$, *dashed line*) and HO-SF$_{-1-2}$ bound to hFEN1 (*blue*) all in Ca^{2+}-containing buffer. HO-SF$_{-1-2}$ bound to hFEN1 in buffer containing 25 mM EDTA (*red*). *D*, comparison of molar ellipticity per 2AP residue at 326 nm of single flap HO-SF$_{-1-2}$ free or bound to WT and mutant hFEN1s in Ca^{2+} (*purple*) and EDTA (*pink*) buffers. The unbound corresponding single strand is also shown.

reducing close to zero at 326 nm (Fig. 3*C*). However, unlike the 5'-phosphorylated SF$_{-1-2}$, the spectra also contained a minimum at 315 nm. Thus, the presence of a 5'-phosphate is not required for the WT protein to bring about local DNA conformational change involving the -1 and -2 nt in substrate DNAs, but the orientation of the 2APs may differ from that adopted by the 5'-phosphorylated form (Fig. 3, *A* and *C*). Additionally, all the mutated proteins had the same response to HO-SF$_{-1-2}$ asSF$_{-1-2}$ with the exception of D181A, where hFEN1-Ca^{2+} reduced the ECCD signal to a lesser extent (Fig. 3*D*).

Streptavidin Blocking of 5'-Flaps Prevents Local DNA Conformational Change—To test whether the severely reduced reaction rates observed with 5'-streptavidin blocked substrates resulted from inability to transfer substrate to the active site, we created a 5'-biotinylated double flap with 2AP at -1 and -2, BDF$_{-1-2}$. The addition of biotin did not alter behavior of the substrate in ECCD experiments (Fig. 4*A*), but its behavior when the 5'-flap was blocked with streptavidin was markedly different. In this case, addition of hFEN1-Ca^{2+} did not alter the ECCD signal, indicating that local substrate conformational

change is prevented by the addition of the streptavidin block. In contrast, when streptavidin was added to trap a preformed complex of hFEN1-Ca^{2+}·BDF$_{-1-2}$, the ability to change the conformation of the substrate was retained. These results demonstrate that proper accommodation of the 5'-flap of the DNA substrate is required for the local conformational change necessary for reaction.

A Watson-Crick Base Pair Is Required at the Terminus of the Hydrolyzed Duplex—To test whether the decreased rate and specificity with mismatched substrates could be attributed to inhibition of the local DNA conformational change, we created a double flap substrate with a $+1$ C-C mismatch retaining 2APs at positions -1 and -2, denoted MM$_{+1}$DF$_{-1-2}$ (Fig. 4*B*). The ECCD signal produced by WT hFEN1-Ca^{2+} and MM$_{+1}$DF$_{-1-2}$ was decreased slightly compared with that for the mismatch substrate alone or the same sample in EDTA but did not approach the nearly zero signal produced with fully base-paired substrate under these conditions. This implies that although the local DNA structure of the mismatched substrate may be subtly altered by
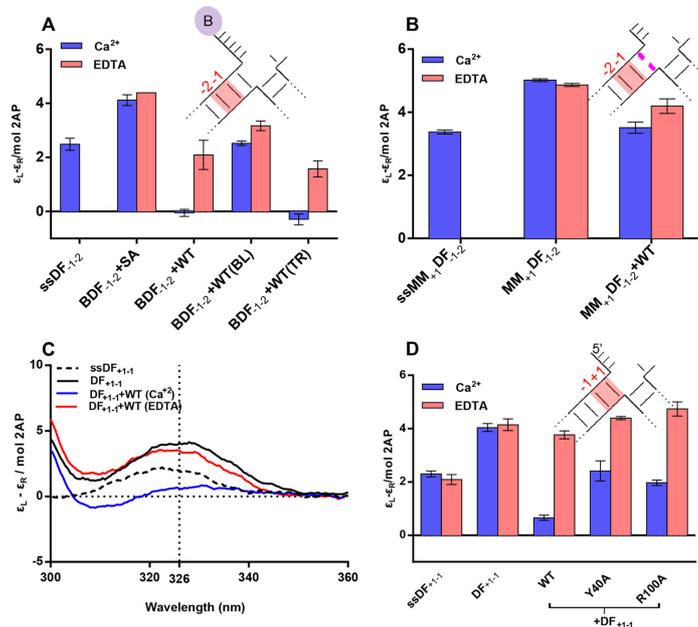
166

FIGURE 4. **ECCD monitored conformational change of +1−1 2AP and 5′-modified −1−2 double flap substrates.** All measurements were carried out at 20 °C and pH 7.5. *ss*, single strand. Standard errors from repeat experiments are shown. *A*, comparison of molar ellipticity per 2AP residue at 326 nm of 5′-streptavidin blocked (*BL*) and free and bound to hFEN1 and streptavidin-trapped (*TR*) complexes in Ca^{2+} (*purple*) and EDTA (*pink*) buffers. Blocked complex was formed by adding streptavidin to the substrate before the addition of hFEN1, whereas trapped was formed by adding streptavidin to the preformed hFEN1-Ca^{2+}-BDF complex. *B*, comparison of molar ellipticity per 2AP residue at 326 nm of a doubled flap substrate with a +1 mismatch ($MMDF_{-1-2}$) when free and bound to WT-hFEN1 in Ca^{2+} (*purple*) and EDTA (*pink*) buffers. The corresponding single strand is also shown. *C*, divalent metal ion-dependent reduction in 2AP exciton coupling signal occurred when substrate DF_{+1-1} was bound to hFEN1, indicative of local substrate conformational change. Unbound DF_{+1-1} (*black*), the corresponding single strand ($ssDF_{+1-1}$, *dashed line*) and DF_{+1-1} bound to hFEN1 (*blue*) all in Ca^{2+}-containing buffer. DF_{+1-1} bound to hFEN1 in buffer containing 25 mM EDTA (*red*). *D*, comparison of molar ellipticity per 2AP residue of double flap DF_{+1-1} at 326 nm when free and bound to WT and R100A hFEN1s in Ca^{2+} (*purple*) and EDTA (*pink*) buffers. The corresponding single strand is also shown. Standard errors from repeat experiments are shown.

hFEN1-Ca^{2+}, it does not adopt the same conformation as the Watson-Crick base-paired substrate, or there is a significant change in the partition between the base-paired and active site positioned forms.

Local DNA Conformational Change at the +1−1 Position Requires Conserved Residues and a +1 Phosphate—Placing the scissile phosphodiester bond on hFEN1 active site metal ions is presumed to require that both the +1 and −1 nt of the substrate unpair from duplex (Fig. 1C). Because there are currently no x-ray structures of hFEN1 in complex with substrate positioned to react, the relative juxtaposition of the −1 and +1 nucleobases in this catalytically competent state are unknown. To use ECCD to inform on this state, we created single flap $SF_{+1−1}$ and double flap $DF_{+1−1}$ substrates containing tandem 2APs at the −1 and +1 positions (Fig. 4, *C* and *D*, and 5, *A* and *B*). In both cases, addition of Ca^{2+} to hFEN1 complexes with the respective substrates substantially decreased the ECCD signal at 326 nm. This implies that in the presence of hFEN1-Ca^{2+} stacking interactions between the +1 and −1 nt are significantly altered.

When the same mutated FEN1s detailed above were employed, K93A most closely resembled the behavior of the WT protein with $SF_{+1−1}$ (Fig. 5B). Both R100A-Ca^{2+} and K93A/R100A-Ca^{2+} also reduced the ECCD signal of $SF_{+1−1}$, although not to the same extent (Fig. 5B). However, Y40A, L130P, and D181A did not significantly alter the signal with $SF_{+1−1}$ at 326 nm ±Ca^{2+}. With double flap substrates and identically positioned 2APs ($DF_{+1−1}$), R100A-Ca^{2+} and Y40A-Ca^{2+} both reduced the ECCD signal but not to the same extent as the WT protein in Ca^{2+} buffer (Fig. 4D).

When the 5′-phosphate was removed from the SF substrate (HO-$SF_{+1−1}$) ECCD signals were significantly altered. A smaller decrease in ECCD signal at 326 nm was observed in the presence of divalent ions and WT protein relative to the same sample in EDTA (Fig. 5C). Moreover, the maximum of the signal with hFEN1-Ca^{2+} was blue-shifted relative to free HO-$SF_{+1−1}$. When mutated hFEN1s interacted with HO-$SF_{+1−1}$, only K93A was able to mimic the small change of WT hFEN1-Ca^{2+} with other proteins producing negligible effects within error.

Combined results imply that changes in the relative orientation of the +1 and −1 nt occur consistent with reduced stacking of these nucleobases once unpaired and extrahelical. These changes evidently require the presence of the +1 5′-phosphate,

167
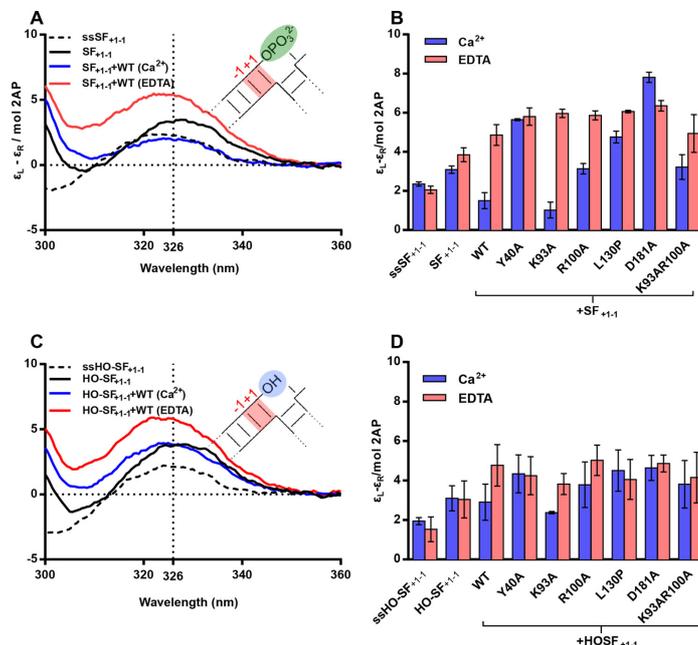
FIGURE 5. **ECCD monitored conformational change of single flap +1−1 2AP containing substrates upon binding hFEN1 and mutants.** All measurements were carried out at 20 °C and pH 7.5. *ss*, single strand. *A*, divalent metal ion-dependent reduction in 2AP exciton coupling signal occurred when single flap SF_{+1-1} was bound to hFEN1, indicative of local substrate conformational change. Unbound SF_{+1-1} (*black*), the corresponding single strand (*$ssSF_{+1-1}$*, *dashed line*) and SF_{+1-1} bound to hFEN1 (*blue*) all in Ca^{2+}-containing buffer. SF_{+1-1} bound to hFEN1 in buffer containing 25 mM EDTA (*red*). *B*, comparison of molar ellipticity per 2AP residue of SF_{+1-1} at 326 nm when free and bound to WT and mutant hFEN1s in Ca^{2+} (*purple*) and EDTA (*pink*) buffers. The corresponding single strand is also shown. Standard errors from repeat experiments are shown. *C*, a small divalent metal ion-dependent reduction in 2AP exciton coupling signal occurred when single flap $HO-SF_{+1-1}$ that lacks a 5′-phosphate was bound to hFEN1, indicative of deficiency in bringing about local substrate conformational change. Unbound $HO-SF_{+1-1}$ (*black*), the corresponding single strand ($ssHO-SF_{+1-1}$, *dashed line*) and $HO-SF_{+1-1}$ bound to hFEN1 (*blue*) all in Ca^{2+}-containing buffer. $HO-SF_{+1-1}$ bound to hFEN1 in buffer containing 25 mM EDTA (*red*). *D*, comparison of molar ellipticity per 2AP residue of single flap $HO-SF_{+1-1}$ at 326 nm when free and bound to WT and mutant hFEN1s in Ca^{2+} (*purple*) and EDTA (*pink*) buffers. The corresponding single strand is also shown. Standard errors from repeat experiments are shown.
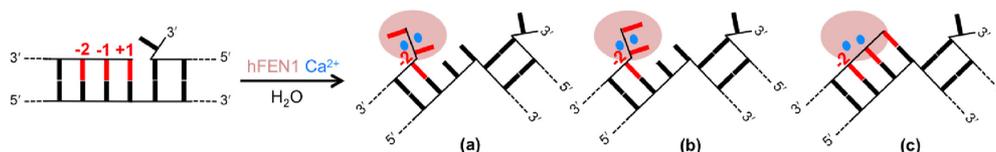
FIGURE 6. **Schematic model summarizing the responses of hFEN1-substrate complexes to addition of divalent metal ions based on ECCD results.** *Part (a)*, in the presence of divalent ions, unmodified substrates interacting with WT and K93A hFEN1s adopt an orientation of the −1 and −2 nt that is unstacked consistent with unpaired DNA. Also, stacking between the −1 and +1 nt is substantially reduced, suggesting control of their relative positions after unpairing. This observed conformational ordering of nucleobases is presumed to effect optimal contact between the scissile bond and active site metal ions and catalytic residues. *Part (b)*, a divalent metal ion-induced substrate state where there is a gross change in the orientation of the −1 and −2 nt suggestive of local DNA unpairing is adopted by R100A, D181A, and Y40A with unmodified substrates and by all proteins (except L130P) with substrates lacking a 5′-phosphate. In these cases, however, there is evidence that stacking reminiscent of ssDNA remains between the −1 and +1 nt, suggesting an unpaired DNA state that is not optimally positioned for reaction. *Part (c)*, the L130P mutation, modifications of the substrate that prevent accommodation of the 5′-flap under the helical cap (*i.e.* streptavidin conjugation to terminus of 5′-flap), or a mismatch at the +1 position prevent a DNA conformational change on addition of divalent ions. In these cases, the substrate is assumed to remain base-paired.

Tyr^{40}, Arg^{100}, and Asp^{181} (Fig. 6*A*). We presume this reflects a conformation of the unpaired substrate that allows optimal orientation of the scissile phosphate relative to active site metal ions, basic residues, and attacking hydroxide. However, changes involving the −1 and −2 nt do not require these substrate and protein features, suggesting that in addition to requirements to effect unpairing of the substrate, additional residues are important to optimally position the unpaired DNA for reaction. (Fig. 6*B*). In contrast, perturbation of the secondary structure of the helical cap (L130P), prevention of substrate

168

threading with a 5′-streptavidin block, or the inclusion of a +1 mismatch abolishes the ability of the protein-substrate complex to undergo the usual local DNA conformational changes when divalent metal ions are added (Fig. 6C).

Discussion

Selection of both the correct DNA substrate and the correct phosphate diester bond for hydrolysis are key to hFEN1 biological function during replication and repair. Incorrect hydrolysis by hFEN1 would endanger genome integrity and necessitate the action of DNA repair mechanisms. The data presented here begin to reveal the details, interrelationships, and complexity of this process. The DNA junction itself is first recognized by its ability to bend 100°. This bent substrate conformation allows recognition of a single-nt 3′-flap and places the 5′-end of the reacting duplex close to the hFEN1 active site. However, the FRET results presented here demonstrate that junction bending does not require the 5′-portion of substrates to be accommodated by the protein either by threading 5′-flaps under the helical cap or by transfer to the active site metal ions (Fig. 2). Substrates that cannot transfer to the active site because metal ions are not present are still bent when bound to hFEN1 protein. Similarly, substrates that lack a 5′-flap or where the 5′-flap is prevented from threading underneath the helical cap are also bent, albeit with modestly reduced stability in the case of the streptavidin-blocked substrate. Thus, although global DNA bending must precede the local DNA conformational change necessary for reaction, it is not required to occur concomitantly with this process.

The key process in enforcing hFEN1 reaction site specificity is the transfer of the scissile phosphate diester located one nt into the reacting duplex onto active site metal ions. ECCD of $(2AP)_2$ containing DNAs demonstrates that FEN1 substrates do not require a 5′-flap to enable this change (Fig. 3), underscoring the fact that exonucleolytic and endonucleolytic reactions of FEN1 substrates proceed by a common mechanism. However, DNAs with a mismatch at the end of the reacting duplex of the substrate are deficient in local DNA conformational changes (Figs. 4B and 6C). Similarly, the status of the 5′-termini of 5′-flaps is a determinant of the ability to bring about local DNA conformational change. Notably, the DNA substrate cannot position for reaction when the protein cannot properly accommodate 5′-flaps, as demonstrated by 5′-streptavidin blocking (Figs. 4A and 6C). Thus, when 5′-flaps with bound protein (e.g. RPA) or lacking free 5′-termini (continuous DNA of template strand) are encountered, reaction is prevented because the scissile phosphodiester bond cannot access the active site.

Alongside a requirement for threading of 5′-flaps demonstrated here, earlier work examining changes in orientation of the −1 and −2 nt in a $(2AP)_2$ DF substrate concluded that individual conserved residues of the hFEN1 protein played little part in this DNA conformational change. However, the presence of active site divalent metal ions and the intact structure of the helical cap were essential for this reorientation (12, 17). Here, we show that this is also the case with exonucleolytic substrates lacking a 5′-flap and that a 5′-phosphate is not required for this −1 and −2 substrate distortion in these SF substrates (Fig. 3). The orientation of the +1 and −1 nt is also dependent on an intact helical cap and the presence of active site divalent metal ions (Figs. 4, C and D, and 5). However, the local conformational changes that occur with +1 and −1 nt are markedly altered by changes in both the substrate and protein.

Despite clear evidence of hFEN1-Ca^{2+} reorientation of the −1 and −2 nt when SF substrates lack a 5′-phosphate (Fig. 3C), only a small change is observed in the +1 and −1 ECCD signal (Fig. 5C). Assuming that the position adopted by the substrate in the presence of hFEN1-Ca^{2+} reflects the catalytically viable conformation, the 5′-phosphate monoester of SF substrates must form a key interaction required to assemble this state. Contacts to the 5′-phosphate monoester are also implied from the FRET studies (Fig. 2C), and although the substrate could still adopt the bent state, removal of the 5′-phosphate monoester of the SF substrate (HO-SF) increased the magnitude of K_{bend} substantially in Ca^{2+} buffer. With DF substrates, interactions with the equivalent 5′-phosphate diester (+1 position, i.e. the next phosphate 5′ in the chain to the scissile phosphate) presumably also play a key role in productive substrate positioning. This would explain earlier work demonstrating that neutralization of the charge of this +1 5′-phosphate by conversion to methyl phosphonate is detrimental to reaction (21). Thus, both ECCD and FRET behaviors reported here are consistent with earlier work in suggesting key interactions involving the substrate 5′-phosphate monoester/diester when DNA is positioned to react within the active site.

The mutation of conserved residues did not produce any substantive variation in the value of K_{bend} in the presence of Ca^{2+} (Fig. 2C). However, several of these residues were implicated in active site substrate positioning by studies of the +1 and −1 ECCD signal (Figs. 5B and 6). When the hFEN1 protein was altered to Y40A, there was no change in +1 and −1 ECCD signal in the presence of divalent metal ions compared with their absence with SF substrate and a substantially reduced effect with DF substrate compared with that seen with WT protein (Fig. 4D). Because Tyr^{40} forms stacking interactions with either the +1 or −1 nucleobases in substrate and product structures, respectively, these interactions are likely in the catalytically competent state. Previous fluorescence studies have revealed evidence for unusually fast quenching of substrate 2APs at both the +1 or −1 positions when bound to hFEN1-Ca^{2+}, consistent with an interaction with Tyr^{40} (17). This was interpreted as an equilibrium between paired and unpaired forms of the substrate with Tyr^{40} interacting with the 2AP at +1 in paired and −1 in unpaired conformations. The data presented here support the idea that the Tyr^{40} residue plays an important role in optimal substrate positioning, and its mutation to alanine was found to reduce the rate of cleavage of DF substrate by a factor of 100.

There was also no change in +1 and −1 ECCD signal with D181A-Ca^{2+}, and the lack of reorientation of the nucleobases in this instance may be related to metal ion positioning in the mutated protein (because Asp^{181} is directly coordinated to one of the active site M^{2+} ions). In addition, Arg^{100} appears to play a role in reorientation of the +1 and −1 nt because with this mutant the ECCD signal was reduced in the presence of Ca^{2+}, but to a lesser extent than with WT hFEN1. Because the Arg^{100}

169

residue contacts the cleaved phosphate monoester in product structures, it may well position the scissile phosphate diester in active site positioned substrate complexes. In contrast, Lys[93] does not play a role in substrate positioning, and the impact of its mutation to alanine seems to be entirely related to catalysis (23).

Overall, these studies unravel the interrelationships between events in the hFEN1 catalytic cycle. Global DNA bending involving interactions with the duplex regions of substrates is essential to position the reacting duplex close to the active site. This facilitates accommodation of the 5′-flap (when present) and the local DNA conformational change required for reaction, but neither of these events is a prerequisite for the initial DNA interaction, suggesting that they occur after binding the substrate duplex regions. Once substrate is bound in a bent conformation, 5′-flaps, if present, are threaded underneath the cap. Threading is a prerequisite for transfer of the scissile phosphodiester to the active site in double flap substrates. Finally, the substrate adopts a single-stranded catalytically competent conformation traveling through the helical gateway (base of $\alpha 4$ and $\alpha 2$) contacting active site metal ions. ECCD results with -1 and -2 $(2AP)_2$ substrates show that metal ions are sufficient to draw the substrate toward the active site providing the cap can adopt a helical state and that 5′-flaps can be threaded (Fig. 6*B*). However, ECCD data with $+1$ and -1 $(2AP)_2$ DNAs demonstrate that the precise positioning of substrate is dependent on interaction with Tyr[40] and Arg[100] residues of the helical gateway and requires the presence of active site Asp[181] and contacts to $+1$ phosphate of substrate (Fig. 6*A*).

Author Contributions—T. D. C. and L. D. F. designed the FRET experiments, which were performed by S. I. A. and I. A. B. J. A. G., L. D. F., S. I. A., and J. C. E. designed the ECCD experiments, which were performed by S. I. A., J. C. E., M. J. T., V. J. B. G., and S. J. S. The proteins were purified by M. J. T., L. D. F., and J. C. E. All authors analyzed the data and contributed to preparation of the manuscript.

References

1. 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

2. Balakrishnan, L., and Bambara, R. A. (2013) Flap endonuclease 1. *Annu. Rev. Biochem.* **82,** 119–138

3. 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

4. Finger, L. D., Atack, J. M., Tsutakawa, S., Classen, S., Tainer, J., Grasby, J., and Shen, B. (2012) The wonders of flap endonucleases: structure, function, mechanism, and regulation. In *The Eukaryotic Replisome: A Guide to Protein Structure and Function* (MacNeill, S., ed) pp. 301–326, Springer-Verlag, New York Inc., New York

5. Tomlinson, C. G., Atack, J. M., Chapados, B., Tainer, J. A., and Grasby, J. A. (2010) Substrate recognition and catalysis by flap endonucleases and related enzymes. *Biochem. Soc. Trans.* **38,** 433–437

6. Kaiser, M. W., Lyamicheva, N., Ma, W., Miller, C., Neri, B., Fors, L., and Lyamichev, V. I. (1999) A comparison of eubacterial and archaeal structure-specific 5′-exonucleases. *J. Biol. Chem.* **274,** 21387–21394

7. 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

8. 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., Cooper, P. K., Grasby, J. A., and Tainer, J. A. (2011) Human flap endonuclease structures, DNA double-base flipping, and a unified understanding of the FEN1 superfamily. *Cell* **145,** 198–211

9. Craggs, T. D., Hutton, R. D., Brenlla, A., White, M. F., and Penedo, J. C. (2014) Single-molecule characterization of Fen1 and Fen1/PCNA complexes acting on flap substrates. *Nucleic Acids Res.* **42,** 1857–1872

10. Sobhy, M. A., Joudeh, L. I., Huang, X., Takahashi, M., and Hamdan, S. M. (2013) Sequential and multistep substrate interrogation provides the scaffold for specificity in human flap endonuclease 1. *Cell Reports* **3,** 1785–1794

11. Orans, J., McSweeney, E. A., Iyer, R. R., Hast, M. A., Hellinga, H. W., Modrich, P., and Beese, L. S. (2011) Structures of human exonuclease 1 DNA complexes suggest a unified mechanism for nuclease family. *Cell* **145,** 212–223

12. 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

13. 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

14. Gloor, J. W., Balakrishnan, L., and Bambara, R. A. (2010) Flap endonuclease 1 mechanism analysis indicates flap base binding prior to threading. *J. Biol. Chem.* **285,** 34922–34931

15. 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 T5 flap endonuclease catalysed DNA hydrolysis. *J. Biol. Chem.* **283,** 28741–28746

16. Beddows, A., Patel, N., Finger, L. D., Atack, J. M., Williams, D. M., and Grasby, J. A. (2012) Interstrand disulfide crosslinking of DNA bases supports a double nucleotide unpairing mechanism for flap endonucleases. *Chem. Comm.* **48,** 8895–8897

17. 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

18. Kalinin, S., Peulen, T., Sindbert, S., Rothwell, P. J., Berger, S., Restle, T., Goody, R. S., Gohlke, H., and Seidel, C. A. (2012) A toolkit and benchmark study for FRET-restrained high-precision structural modeling. *Nat. Methods* **9,** 1218–1225

19. Clegg, R. M., Murchie, A. I., Zechel, A., Carlberg, C., Diekmann, S., and Lilley, D. M. (1992) Fluorescence resonance energy transfer analysis of the structure of the four-way DNA junction. *Biochemistry* **31,** 4846–4856

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. Allawi, H. T., Kaiser, M. W., Onufriev, A. V., Ma, W. P., Brogaard, A. E., Case, D. A., Neri, B. P., and Lyamichev, V. I. (2003) Modelling of flap endonuclease interactions with DNA substrate. *J. Mol. Biol.* **328,** 537–554

22. Tomlinson, C. G., Syson, K., Sengerová, B., Atack, J. M., Sayers, J. R., Swanson, L., Tainer, J. A., Williams, N. H., and Grasby, J. A. (2011) Neutralizing mutations of carboxylates that bind metal 2 in T5 flap endonuclease result in an enzyme that still requires two metal ions. *J. Biol. Chem.* **286,** 30878–30887

23. Sengerová, B., Tomlinson, C., Atack, J. M., Williams, R., Sayers, J. R., Williams, N. H., and Grasby, J. A. (2010) Brønsted analysis and rate-limiting steps for the T5 flap endonuclease catalyzed hydrolysis of exonucleolytic substrates. *Biochemistry* **49,** 8085–8093

24. Jose, D., Datta, K., Johnson, N. P., and von Hippel, P. H. (2009) Spectroscopic studies of position-specific DNA "breathing" fluctuations at replication forks and primer-template junctions. *Proc. Natl. Acad. Sci. U.S.A.* **106,** 4231–4236

25. Johnson, N. P., Baase, W. A., and Von Hippel, P. H. (2004) Low-energy circular dichroism of 2-aminopurine dinucleotide as a probe of local conformation of DNA and RNA. *Proc. Natl. Acad. Sci. U.S.A.* **101,** 3426–3431

170

Phosphate steering by Flap Endonuclease 1 promotes 5′-flap specificity and incision to prevent genome instability

Susan E. Tsutakawa[1,*], Mark J. Thompson[2,*], Andrew S. Arvai[3,*], Alexander J. Neil[4,*], Steven J. Shaw[2], Sana I. Algasaier[2], Jane C. Kim[4], L. David Finger[2], Emma Jardine[2], Victoria J.B. Gotham[2], Altaf H. Sarker[5], Mai Z. Her[1], Fahad Rashid[6], Samir M. Hamdan[6], Sergei M. Mirkin[4], Jane A. Grasby[2] & John A. Tainer[1,7]

DNA replication and repair enzyme Flap Endonuclease 1 (FEN1) is vital for genome integrity, and FEN1 mutations arise in multiple cancers. FEN1 precisely cleaves single-stranded (ss) 5′-flaps one nucleotide into duplex (ds) DNA. Yet, how FEN1 selects for but does not incise the ss 5′-flap was enigmatic. Here we combine crystallographic, biochemical and genetic analyses to show that two dsDNA binding sites set the 5′polarity and to reveal unexpected control of the DNA phosphodiester backbone by electrostatic interactions. Via 'phosphate steering', basic residues energetically steer an inverted ss 5′-flap through a gateway over FEN1's active site and shift dsDNA for catalysis. Mutations of these residues cause an 18,000-fold reduction in catalytic rate *in vitro* and large-scale trinucleotide (GAA)$_n$ repeat expansions *in vivo*, implying failed phosphate-steering promotes an unanticipated lagging-strand template-switch mechanism during replication. Thus, phosphate steering is an unappreciated FEN1 function that enforces 5′-flap specificity and catalysis, preventing genomic instability.

[1] Molecular Biophysics and Integrated Bioimaging, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. [2] Centre for Chemical Biology, Sheffield Institute for Nucleic Acids (SInFoNiA), Department of Chemistry, University of Sheffield, Sheffield S3 7HF, UK. [3] Department of Molecular Biology, The Scripps Research Institute, La Jolla, California 92037, USA. [4] Department of Biology, Tufts University, Medford, Massachusetts 02155, USA. [5] Biological Systems and Engineering, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. [6] Division of Biological and Environmental Sciences and Engineering, King Abdullah University of Science and Technology, Thuwal 23955, Saudi Arabia. [7] Department of Molecular and Cellular Oncology, The University of Texas M.D. Anderson Cancer Center, Houston, Texas 77030, USA. * These authors contributed equally to this work. Correspondence and requests for materials should be addressed to S.M.M. (email: sergei.mirkin@tufts.edu) or to J.A.G. (email: j.a.grasby@sheffield.ac.uk) or to J.A.T. (email: jtainer@mdanderson.org).

171

T he structure-specific nuclease, flap endonuclease-1 (FEN1) plays a vital role in maintaining genome integrity by precisely processing intermediates of Okazaki fragment maturation, long-patch excision repair, telomere maintenance, and stalled replication forks. During DNA replication and repair, strand displacement synthesis produces single-stranded (ss) 5′-flaps, at junctions in double-stranded (ds) DNA. During replication in humans, FEN1 removes ∼50 million Okazaki fragment 5′-flaps with remarkable efficiency and selectivity to maintain genome integrity[1–3]. Consequently, FEN1 deletion is embryonically lethal in mammals[4], and functional mutations can lead to cancer[5]. FEN1 also safeguards against DNA instability responsible for trinucleotide repeat expansion diseases[6]. As FEN1 is overexpressed in many cancer types[7,8], it is an oncological therapy target[9,10].

Precise FEN1 incision site selection is central to DNA replication fidelity and repair. FEN1 preferentially binds to double flap substrates with a one nt 3′-flap and any length of 5′-flap, including zero. It catalyses a single hydrolytic incision one nucleotide (nt) into dsDNA (Fig. 1a) to yield nicked DNA ready for direct ligation[11,12]. Thus, FEN1 acts on dsDNA as both an endonuclease (with 5′-flap) and an exonuclease (without 5′-flap). Recent single molecule experiments show that FEN1 binds both ideal and non-ideal substrates but decisively incises only its true substrate[13]. In contrast to homologs in bacteriophage[14–16] and some eubacteria[17], eukaryotic FEN1s do not hydrolyse within 5′-flap ssDNA.

However, key features of FEN1 substrate selection remain unclear. FEN1 must efficiently remove 5′-flaps at discontinuous ss-dsDNA junctions yet avoid genome-threatening action on continuous ss–ds junctions, such as ss gaps or Holliday junctions. Paradoxically, other FEN1 5′-nuclease superfamily members[3] are specific for continuous DNA junctions: namely, ERCC5/XPG (nucleotide excision repair), which acts on continuous ss-ds bubble-like structures; and GEN1 (Holliday junction resolution), which processes four-way junctions. Structures determined with DNA of eukaryotic superfamily members lack ss-ds junction substrate with 5′-ssDNA or the attacking water molecule leaving cardinal questions unanswered[18–22]. For example, structures of FEN1 and Exo1 go from substrate duplex DNA with the scissile phosphodiester far from the catalytic metals to an unpaired terminal nt in the product; is the unpairing occurring before or after incision?

Models of FEN1 specificity must address how ss–ds junctions are recognized and how 5′-flaps, as opposed to continuous ssDNA are recognized. There are threading and kinking models. To exclude continuous DNAs, 5′-flaps may thread through a 'tunnel'[21,23–25] formed by two superfamily-conserved helices flanking the active site, known as the 'helical gateway,' topped by a 'helical cap' (Fig. 1b). Due to cap and gateway disorder in DNA-free FEN1, they are thought disordered during threading and to undergo a disorder-to-order transition on 3′-flap binding[21,24,26]. In this threading model, however, ssDNA passes through a tunnel without an energy source and directly over the active site, risking non-specific incision. These issues prompted an alternative clamping model where the ss flap kinks away from the active site[11,20] (Fig. 1b). Whereas these models explain selection against continuous DNA junctions, FEN1 exonuclease activity does not require a 5′-flap. Furthermore, how FEN1 prevents off target incisions and moves the dsDNA junction onto the metal ions are not explained by these models.

Here crystallographic analyses uncover an unprecedented electrostatic steering of an inverted 5′-flap through the human FEN1 (hFEN1) helical gateway. Gateway and cap positively-charged side chains are positioned to 'steer' the phosphodiester backbone across the active site, energetically promoting threading

and preventing nonspecific hydrolysis within the 5′-flap. Mutational analysis of these positively charged 'steering' residues revealed an added role of phosphate steering in moving dsDNA towards the catalytic metal ions for reaction. Moreover, phosphate steering mutations efficiently blocked Rad27 (S. cerevisae homolog of hFEN1) function, causing a compromised response to DNA damaging agents and dramatically increased expandable repeat instability.

Results

FEN1 selects for 5′-flaps by steering flap through a gateway. To obtain structures of hFEN1 with a ss 5′-flap substrate for insight into ss 5′-flap selection, we crystallized three hFEN1 active site mutants D86N, R100A and D233N with a double-flap (DF) substrate and with Sm^{3+} (Fig. 1 and Supplementary Figs 1 and 2A)[27]. Mg^{2+} is the physiological cofactor. D86N, R100A and D233N mutations slow the hFEN1 catalysed reaction rate by factors of 530, 7,900 and 16 respectively (Supplementary Fig. 2B). The DF substrates in the crystal structures had a ss 5′-flap (4–5 nt) and a 1 nt 3′-flap, termed S4,1 or S5,1 (Supplementary Fig. 1). The DNA-enzyme complex structures for hFEN1-D86N, hFEN1-R100A, and hFEN1-D233N were determined to 2.8, 2.65, and 2.1 Å resolution, respectively (Figs 1c,d and 2, Supplementary Fig. 2 and Table 1). In all cases, the overall protein resembled wild-type (wt) hFEN1 (with product DNA, PDB code 3Q8K)[21], with root mean square deviation (RMSD) values of 0.26 for hFEN1-R100A, 0.22 for hFEN1-D233N, and 0.42 for hFEN1-D86N.

These structures show that FEN1 interacts primarily (88% by PISA interface analysis[28]) with two regions of ∼100° bent dsDNA supporting prior observations[21], rather than to the ss 5′-flap in these structures (Figs 1c,d and 2, Supplementary Movie 2). FEN1 binding to dsDNA is mediated by four regions: (1) a hydrophobic wedge (composed of helix 2 and helix 2–3 loop) and β pin (formed between β strands 8 and 9) which sandwich upstream and downstream dsDNA portions at the bending point of the two-way junction with Tyr40 packing at the ss/dsDNA junction; (2) a C-terminal helix-hairpin-helix motif binds upstream dsDNA and the one nt 3′-flap and is absent from superfamily-related members hEXO1 (ref. 20) and bacteriophage 5′-nuclease structures[29]; (3) the helix-2turn-helix (H2TH) motif with bound K^+ ion and positive side chains bind downstream dsDNA; and (4) the two-metal ion active site near the 5′-flap strand. Much of the interaction (43% by PISA analysis) is to the strand complementary to the flap strands, reinforcing dsDNA specificity.

The dsDNA binding sites on either side of the active site, the $K+$ and the hydrophobic wedge, are spaced one helical turn apart (Supplementary Movie 2). Their spacing enforces the specificity for helical dsDNA and places the 5′-flap in the active site, selecting against unstructured ssDNA or 3′-flaps that would require a narrower spacing. Additionally, the minor-groove phosphate backbone is recognized by superfamily-conserved Arg70 and Arg192 pair spaced 14 Å apart (Fig. 2, Supplementary Movie 1). Unique to FEN1, cap positive side chains (Lys125, Lys128, Arg129) interact with the template strand at the ss/dsDNA junction (Supplementary Fig. 3, Supplementary Movie 1). Lys128 and Arg47 pack against each other, linking the 3′-flap pocket to the gateway helices. The active site consists of seven superfamily-conserved metal-coordinating carboxylate residues plus invariant Lys93 and Arg100 from gateway helix 4 and Gly2 at the processed N-terminus (Figs 2 and 3c, Supplementary Movie 4). An ordered gateway and cap formed by helices 2, 4 and 5 are observed above the active site in these three structures. Helix 2 Tyr40 forms part of the hydrophobic wedge and packs the duplex DNA at the bend.

172

In the hFEN1-D86N and hFEN1-R100A structures, the ssDNA (5′-flap) region of the substrate threaded through the tunnel formed by the gateway/cap (Figs 1d and 2a, and Supplementary Fig. 2A and Supplementary Movie 1). This observation explains how FEN1 excludes continuous DNA like Holliday junctions and DNA bubbles. The third independent hFEN1-D233N crystal structure captures two cleaved nts from the 5′-flap bound on the other side of the tunnel from the dsDNA, consistent with threading (Fig. 3a and Supplementary Fig. 2A). Together, these three distinct structures support the threading model to validate substrates have a ss 5′-flap.

Phosphate steering inverts the ss phosphodiester backbone. In both threaded substrate structures, the ss 5′-flap phosphodiester backbone is 'inverted' between the $+1$ and $+2$ positions, with the $+2$ and $+3$ phosphates facing away from the active site metals and the DNA bases facing the metals (Fig. 1b,d and Supplementary Movie 1). (We denote the plus and minus positions relative to the scissile phosphate). This inversion would place the flap phosphodiesters away from the catalytic metals and thereby logically reduce inadvertent incision within the ssDNA. In both structures, the inverted $+1$ phosphodiester is directly between the gateway helices with the bases on either side of the
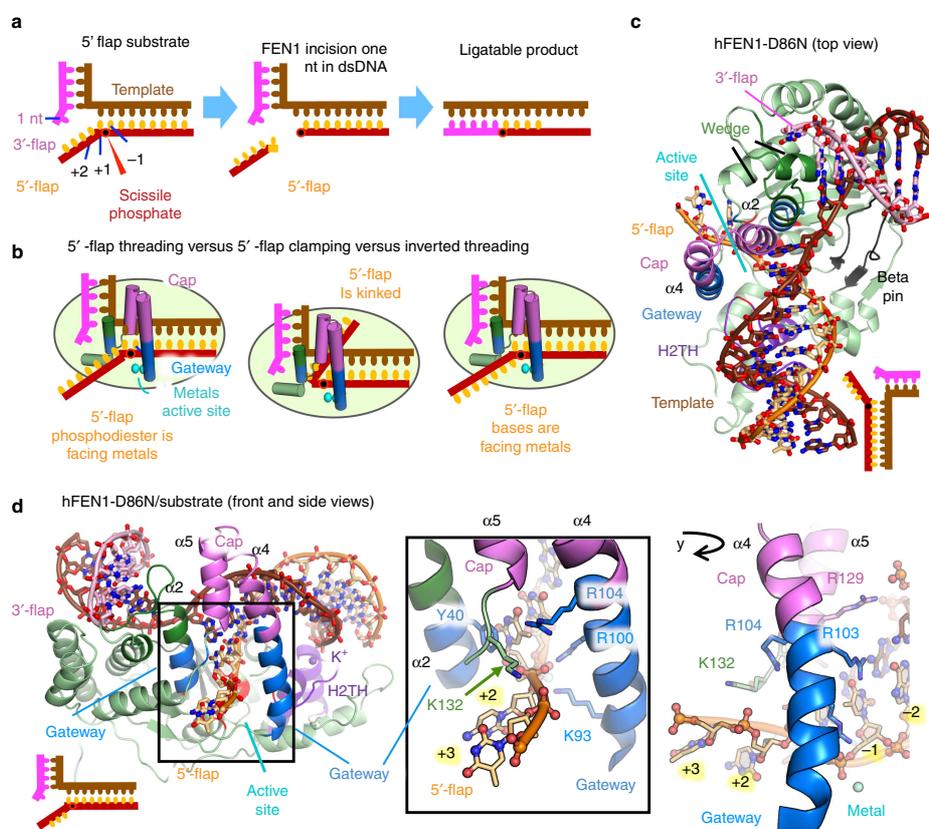
Figure 1 | Specificity and inverted threading of ss5′-flap in hFEN1 D86N substrate structure. (**a**) Schematic FEN1 incision on an optimal double-flap substrate, incising 1 nt into dsDNA to ensure a ligatable product. (**b**) Proposed models for ssDNA selection. (**c**) Top view of hFEN1-D86N crystal structure showing extensive interaction to dsDNA arms of 5′ flap substrate. The 5′-flap substrate is composed of three strands; the 5′-flap strand (orange), the template strand (brown), and the 3′-flap strand (pink). Functionally critical regions in FEN1 include the gateway (blue) and the cap (violet) for selecting substrates with ss-5′-flaps, the hydrophobic wedge between the 3′-flap binding site and the gateway/cap (dark green), the K$^+$ ion and H2TH (purple) that interacts with the downstream DNA, the beta pin (grey) that locks in the DNA at the bend. Relative DNA orientation shown in schematic on lower right. (**d**) Front and side views of hFEN1-D86N crystal structure showing helical gateway and cap architecture position positively-charged residues to steer ss 5′-flaps through a protecting gateway in an inverted orientation across the active site. Relative DNA orientation is shown in schematic. The inverted 5′-flap ssDNA is threaded between gateway helices (blue) and under the helical cap (violet). The inverted threading reveals charged interactions to basic sidechains in the cap and van der Waals interactions to ssDNA. See also Supplementary Figs 1–3; Table 1, and Supplementary Movies 1 and 2.

gateway. In the hFEN1-D86N structure, two basic residues of the gateway/cap, Arg104 and Lys132, were within 4–7 Å of the +1, +2 and +3 phosphodiester. These residues are positioned to

energetically promote threading and an inverted orientation. They are conserved in FEN1 and semi-conserved across the 5′-nuclease superfamily and shown important for incision activity
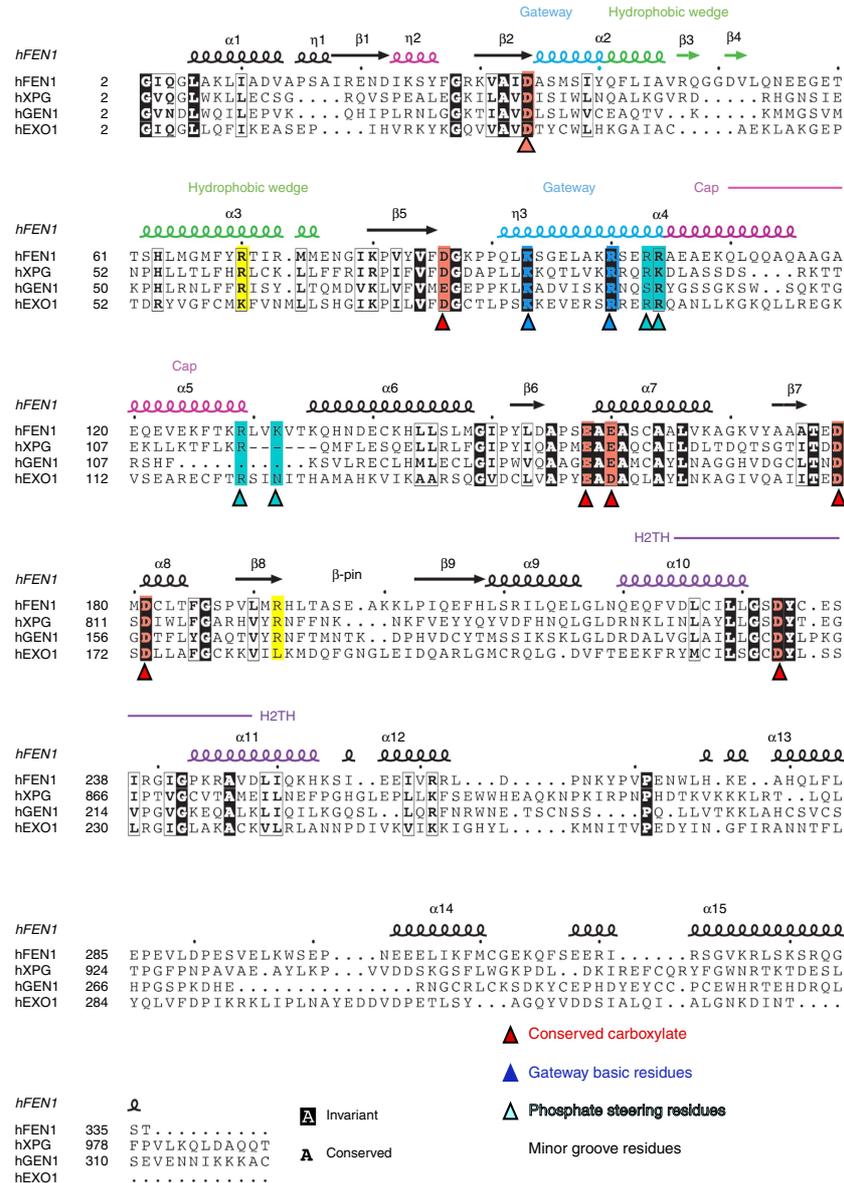
Figure 2 | FEN1 superfamily sequence and secondary structure alignment. Map of FEN1 secondary structure (PDB code 3q8k), structural elements, and mutants to a sequence alignment of FEN superfamily human members. XPG residues 117–766 were removed (dash) to facilitate alignment.

174

Table 1 | X-ray data collection and refinement statistics (molecular replacement).

	hFEN1-D86N	hFEN1-R100A	hFEN1-D233N
Data collection			
Space group	P 31 2 1	P 31 2 1	P 31 2 1
Cell dimensions			
$\quad a, b, c$ (Å)	105.8 105.8 100.7	105.2 105.25 104.1	105.2 105.2 104.5
$\quad \alpha, \beta, \gamma$ (°)	90 90 120	90 90 120	90 90 120
Resolution (Å)	18.2–2.8 (2.9–2.8)	37.0–2.6 (2.7–2.6)	34.5–2.1 (2.2–2.1)
R_{meas}	0.08 (1.0)	0.07 (0.70)	0.09 (0.65)
$I/\sigma I$	22.7 (1.95)	12.7 (2.9)	73.9 (5.00)
Completeness (%)	1.00	0.98	0.99
Redundancy	8.6 (8.5)	5.2 (5.4)	13.8 (10.4)
Refinement			
Resolution (Å)	18.2–2.8	37.0–2.6	34.5–2.1
No. reflections	30,758 (3,031)	34,943 (4,282)	39,092 (3,705)
R_{work}/R_{free}	0.21/0.26	0.22/0.25	0.18/0.22
No. heavy atoms	3,534	3,703	4,183
$\quad$ Protein/DNA	3,492	3,519	3,545
$\quad Sm^{3+}$, K^+	7	6	9
$\quad$ Water	36	173	625
B-factors		91	64
$\quad$ Protein/DNA	113.97	92	63
$\quad Sm^{3+}$, K^+	144	91.5	93.3
$\quad$ Water	110	77.3	69
R.m.s. deviations			
$\quad$ Bond lengths (Å)	0.004	0.005	0.003
$\quad$ Bond angles (°)	0.56	0.60	0.54

Related to Figs 1 and 3.
One crystal was used for each mutant. Values in parentheses are for highest-resolution shell.

(Fig. 2 and Supplementary Fig. 3)[3,21,30]. The +2 and +3 nt of the 5′-flap were sandwiched between the main chain (residues 86–89 and residues 132–135) on one side of the channel and Leu97 on the other (Supplementary Fig. 2C) by non-sequence specific van der Waals contacts. The overall inverted flap orientation resembles the hFEN1-R100A structure with the +1 phosphate remaining within 7 Å of Arg104, but shifted towards Arg103, presumably due to Arg100 removal. Together, these substrate structures suggest that basic residues enable a phosphate steering mechanism, which we here define as electrostatic interactions that can dynamically position the phosphodiester backbone.

Shifting of the scissile phosphate and the catalytic mechanism. In the hFEN1-D86N structure, we were surprised that the scissile phosphate was within catalytic distance of the active site while surrounding bases remained basepaired to the template strand (Fig. 3). This contradicts the prevailing hypothesis that surrounding bases must unpair for the scissile phosphate to move into the active site for incision[3]. Similarities and functionally-significant differences appeared on closer examination of hFEN1-D86N, hFEN1-R100A, and an earlier structure of FEN1-substrate with no 5′-flap or +1 phosphate (PDB code 3Q8L). In all three substrate structures, the dsDNA major groove is widened as it approaches the active site, and DNA bases flanking the scissile phosphate are stacked with one another, with the +1 base packed against Tyr40. However, the basepairing, the scissile phosphodiester bond location and the Tyr40 rotamer are distinctly different in the respective complexes, despite containing the same dsDNA sequence. In the 3Q8L structure, the DNA remained fully basepaired, and the scissile phosphodiester was positioned ∼6 Å away from catalytic metals. In the hFEN1-R100A structure, the scissile phosphodiester bond was ∼4–5 Å away from the metal ions, although −1 and +1 nts

have moved towards the active site and away from the template strand. The −1 and −2 nts display less base overlap (stacking), and the +1 and −1 nts are no longer hydrogen bonded to the template strand (4–6 Å apart).

In striking contrast, the scissile phosphodiester bond was directly coordinated to the one active site metal ion in the hFEN1-D86N structure (Fig. 3c). Furthermore, the +1 and −1 nts remained unexpectedly basepaired to the template strand, which is shifted relative to the other substrate structures via a dsDNA distortion surrounding the scissile phosphodiester (Fig. 3b and Supplementary Fig. 2D,E). There is no base stacking between −1 and −2 nts in the 5′-flap strand; instead, an unusual interstrand base stacking interaction occurs between the −2 nt of the 5′-flap strand and the template strand opposite of the −1 nt. We had hypothesized that unpairing of the +1 and −1 was required to move the scissile phosphate to within catalytic distance of the active site metals[3,21,31,32]. This new hFEN1-D86N substrate structure shows instead that basic residues can rotate dsDNA into the active site with basepairing intact (Fig. 3, Supplementary Movie 3). Moreover, since the DNA in 3Q8L, which was the furthest from the active site, lacked a 5′-flap or +1 phosphate, the DNA movements observed in the hFEN1-R100A and hFEN1-D86N structures are likely partly a consequence of either the 5′-flap and/or the +1 phosphate.

In concert with the DNA rotation in hFEN1-D86N, Tyr40 is in a different rotamer conformation from all other substrate or product bound or DNA-free hFEN1 structures (Fig. 3a,b and Supplementary Movie 3)[21,26]. This Tyr40 rotamer shift tracks duplex DNA rotation into the active site. The Tyr ring is fully stacked on the +1 base, and its side chain hydroxyl forms a hydrogen bond to the +1 phosphate. Notably, as duplex DNA is not shifted close to the catalytic metal in the R100A structure, this structure may represent a pre-reactive substrate form. Its Tyr40 stacks at a 50° angle with the +1 nt and resembles the other hFEN1 structures, suggesting that the Tyr40 rotamer is linked
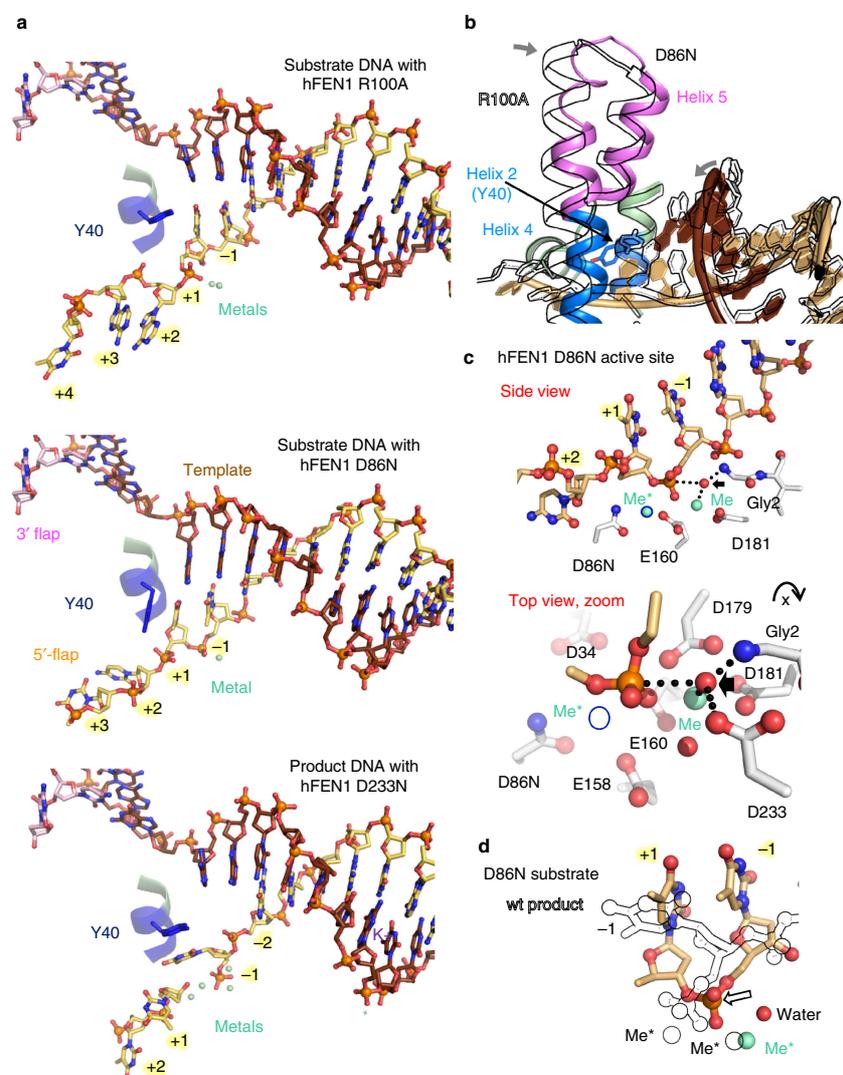
Figure 3 | Three FEN1 crystal structures show threading through the capped gateway. (**a**) DNA from hFEN1-D86N, hFEN1-R100A, and hFEN1-D233N structures showed threading not clamping. Tyr40 (stick model) changes its rotamer to track DNA movement through the gateway in the hFEN1-D86N structure. (**b**) A protein chain overlay between hFEN1-R100A (outline) and hFEN1-D86N (coloured) highlights how the helical gateway and cap and the DNA rotate closer together in the hFEN1-D86N complex. See Supplementary Movie 3. (**c**) The hFEN1-D86N active site revealed a water molecule positioned for linear attack on the scissile phosphate. Orthogonal views are shown. The 2nd metal position (outlined in black and denoted by Me*) is not observed in the hFEN1-D86N and is shown by overlaying the protein from the wt-product structure (PDB code 3Q8K). See also Supplementary Figs 1–4, and Supplementary Movie 4. (**d**) Protein chain overlay of hFEN1-D86N-substrate (coloured) and wt-product (outline, PDB code 3Q8K) structures shows that the scissile phosphate is shifted in the active site ~2 Å (demarked by arrow).

to shifting duplex DNA into a catalytic position. In the hFEN1-D86N-substrate structure, cap and gateway helices 4 and 5 are shifted 1–3 Å towards the dsDNA relative to all other hFEN1 crystal structures with DNA (Fig. 3b and Supplementary Movie 3). The backbone of helix 2 (which contains Tyr40) does not change position.

Close examination of the hFEN1-D86N structure revealed a water molecule 3.3 Å from the scissile phosphate (Fig. 3c, Supplementary Fig. 4, and Supplementary Movie 4). This water is positioned for a linear attack on the scissile phosphate and for its evident activation by the catalytic metal and the Gly2 at the FEN1 N terminus, which was proposed to replace the 'third' metal in bacteriophage FEN[33]. Asp233 is 3 Å from the attacking water and contributes modestly to catalysis; the D233N mutant has 16-fold reduced but still substantial catalytic activity compared to mutants of other invariant carboxylates, such as D181A[21] and D86N (Supplementary Fig. 2B). When a second metal ion is modeled by overlay with the hFEN1-product structure (PDB code 3Q8K), the structure is reminiscent of the classical two-metal-ion catalysis[34]. Moreover, superfamily conserved and catalytically required[21] Lys93 and Arg100 sidechains point towards the scissile phosphodiester bond, poised to assist metal ion mediated hydrolysis. On the basis of the hFEN1-R100A structure, Arg100 is also likely essential for shifting of the scissile phosphate into direct contact with the catalytic metals. Notably, the scissile phosphate has moved ~1–2 Å between hFEN1-D86N-substrate and wt hFEN1-product (Fig. 3d). An analogous metal movement into more optimal coordination geometry in an RNaseH-product structure was proposed to favour product formation[35]. We cannot exclude a possible third metal ion as time-resolved experiments on other enzymes show metals ions can appear and disappear during reaction[36–40].

Together these structures reinforce and extend biochemical data that suggest that FEN1 checks for the ss 5′-flap by threading it through a tunnel formed between the active site and capped gateway helices (Fig. 1d)[24,41]. The substrate structures imply the 5′-flap is (1) electrostatically steered through the capped gateway by conserved basic residues in the gateway and cap and (2) positioned in an inverted orientation.

Biochemically testing phosphate steering. If the gateway/cap region basic residues steer the phosphodiester backbone as implied by the structures, then their mutation should affect 5′-flap substrate incision rates. On the basis of the hFEN1-R100A structure, we mutated three basic residues (Arg103, Arg104 and Lys132) positioned to guide the phosphodiester backbone and stabilize the inverted ssDNA orientation (Fig. 1d). We also mutated Arg129, which is adjacent the other residues and could act in steering. When the helical cap is structured, Arg129 makes a long-range electrostatic interaction with a phosphate of the template strand[21], a distance shortened in hFEN1-D86N by template strand relocation. Strikingly, these four basic residues are conserved across all FEN1s including yeast and archaeal, except for the less-specific phage 5′ nucleases (Supplementary Fig. 3). As the helical gateway and cap regions are flexible before productive DNA binding[22,24,26], specific interactions would seem unlikely during the flap threading process but electrostatic guidance is possible. Notably, as these side chains range from 10 to 19 Å from the target phosphate, they are unlikely to impact FEN1 activity by aspects other than electrostatic guidance and substrate-positioning. To test this idea, we mutated them to alanine or glutamate to either remove the attractive positive charge or provide a repulsive charge, respectively.

These charge mutations all reduced specific incision activity on a 5′-flap substrate, S5,1 (Fig. 4a), indicating an important functional role. Under multiple turnover conditions, single mutations R103A and K132A moderately decreased the reaction rate relative to wt hFEN1 by 3- and 5-fold, respectively, whereas a 20-fold decrease was observed with either R104A or R129A (Fig. 4b and Supplementary Fig. 5A,B). These rate decreases are consistent with a single residue electrostatic guidance interaction[42]. Double mutant R104AK132A showed an additive effect with 200-fold reduced activity and, significantly, the corresponding repulsive mutant R104EK132E was far more severely compromised with a rate reduction of 11,000 compared to the wt enzyme. Importantly, the substrate dissociation constant (K_d) for each of these double mutants was only modestly raised (Supplementary Fig. 6). This suggests deficient substrate positioning, not poor binding, as the major contributing factor to diminished activity. Similarly, double mutants R103AR129A or R103ER129E showed reductions in reactivity of 70- or 5,000-fold, respectively,
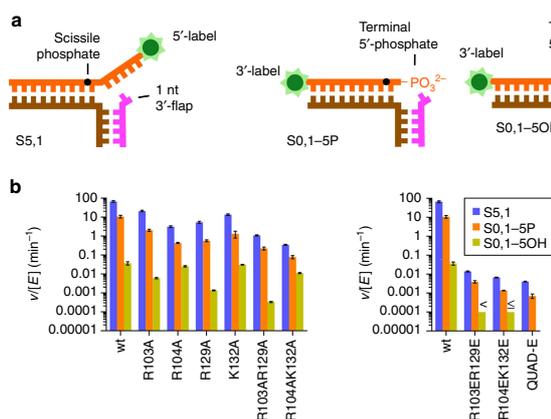
Figure 4 | Phosphate steering residue mutants show reduced activity. (**a**) Schematic of substrates used with fluorescent positions. (**b**) Comparison of multiple turnover rates for cleavage of each substrate at 50 nM. Reaction rates for glutamate mutants with S0,1-5OH were too slow to measure accurately, so threshold values are indicated. See also Supplementary Figs 1, 5 and 6. Error bars are shown as a function of s.e.m., with replicate number given in Supplementary Fig. 5E.

without any substantial effect on K_d. Analogous trends in rate effects were observed under single turnover kinetic conditions (Supplementary Fig. 5C,D).

Mutating all four gateway/cap residues to glutamate ('QUAD-E' mutant) severely impaired activity (18,000-fold slower than wt FEN1). Strikingly, the K_d increased only 17-fold showing the enzyme was folded and capable of substrate binding. This large rate decrease is remarkable for mutation of residues not acting in catalysis and distant from the active site: it resembles the penalty for streptavidin added to 5′-biotinylated substrates, which would prevent 5′-flap gateway/cap threading[24].

If the FEN1 basic cap residues are primarily required for ss 5′-flap steering, then their mutation should not be deleterious to incision activity on an exonucleolytic substrate lacking a 5′-flap but with a 5′-phosphate (S0,1-5P; Fig. 4a). This substrate was hydrolysed sevenfold more slowly than the DF S5,1 by wt hFEN1 (Fig. 4b and Supplementary Fig. 5A,B) showing that threading the 5′-flap facilitates access to the catalytically competent conformation, as well as being a key mechanism in substrate selection. For reaction rates expressed relative to wt hFEN1 to normalize for this sevenfold difference, the gateway mutants all proved similarly defective on both the exonucleolytic (S0,1-5P) and endonucleolytic (S5,1) substrates (relative rates given in Supplementary Fig. 4B). These results unmask a key universal role for $+1$ phosphate steering in the FEN1 incisions of both exonucleolytic and endonucleolytic substrates (since this phosphate is present in both substrates).

Given the results with the exonucleolytic substrate and the observation that DNA movement towards the active site required a $+1$ phosphate[43], we reasoned that some basic residues were electrostatically interacting with the $+1$ phosphodiester of dsDNA to facilitate this movement[44]. To test this hypothesis, we measured reaction rates with an analogous exonucleolytic substrate lacking the 5′-phosphate at the $+1$ position (S0,1-5OH; Fig. 4a). This substrate was bound 20-fold more weakly and incised 300-fold more slowly than S0,1-5P by wt hFEN1 (Fig. 4b, Supplementary Figs 4 and 5). These data indicate that 5′-phosphate ($+1$ phosphate) interactions stabilize the enzyme-substrate complex and contribute to catalysis. Combined and individual mutations of R103A and R129A all decreased incision rates of S0,1-5OH analogously to the other substrates. However, R104A, K132A and R104AK132A all processed S0,1-5OH at a similar rate to wt hFEN1. These results imply that the $+1$ phosphate group functionally interacts with Arg104 and Lys132, consistent with the phosphate steering hypothesis, but that Arg103 and Arg129 (along with Arg100 and Lys93) have long-range interactions to other parts of the DNA substrate, including the scissile phosphate itself.

A role for phosphate steering in genome stability. To test the biological importance of the basic cap and gateway residues, we made equivalent mutations in the helical gateway/cap region of Rad27, the *S. cerevisiae* homolog of hFEN1, to analyse their role in genome integrity *in vivo*. Alanine or glutamate mutations were introduced at Rad27 Arg104, Arg105, Arg127 and Lys130; equivalent to hFEN1 Arg103, Arg104, Arg129 and Lys132, respectively (Fig. 5a). Growth characteristics of double- or quadruple-mutant yeast strains were compared to wild type *RAD27* (wt) strain and *rad27-D179A* (corresponding to human D181A whose incision rate is given in Supplementary Fig. 2B), a severely catalytically impaired mutant, which displays an equivalent phenotype to the *rad27* null strain (that is, sensitivity to hydroxyurea, a replication inhibitor and DNA-damaging UV light[1,45]).
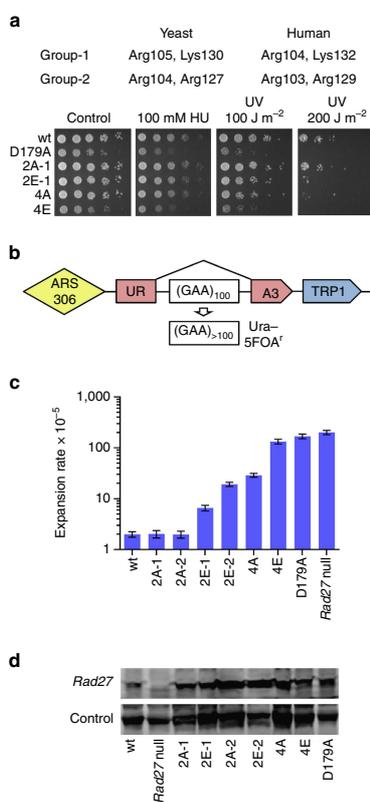
Figure 5 | Phenotypic and DNA repeat expansion defects of *Rad27* basic cap residue mutations in yeast *S. cerevisiae*. (a) Table of the tested basic residues in yeast (and their human counterpart) and spot-test (serial fivefold dilutions) for yeast growth with and without exposure to hydroxyurea (replication inhibitor) or UV light (DNA-damaging). **(b)** Experimental system to measure the rates of large-scale repeat expansions in yeast. The $(GAA)_{100}/(TTC)_{100}$ repeat is incorporated into the intron of an artificially split *URA3* gene. Addition of ≥ 10 extra repeats inhibits reporter's splicing, which allows cells with repeat expansions to grow on 5-FOA-containing media. **(c)** Effect of active site control and phosphate steering mutations in the *RAD27* gene on repeat expansion rates (error bars represent 95% confidence intervals of calculated expansion rates). **(d)** Rad27 protein expression was not substantially altered in the mutated strains. See also Supplementary Fig. 7 and Supplementary Table 1.

Even without exogenous treatment, quadruple glutamate mutant (Fig. 5a) showed growth inhibition resembling that for the active site mutant *rad27-D179A*. Replication stress induced by hydroxyurea greatly accentuated this effect. Moderate UV irradiation ($100 \, J \, m^{-2}$) was strongly deleterious to both strains, and higher-dose irradiation ($200 \, J \, m^{-2}$) further revealed UV sensitivity for the double glutamate (2E-1; R105EK130E) and quadruple alanine (4A) mutants (Fig. 5a). Thus, electrostatic interactions of gateway/cap basic residues with DNA are critical for flap endonuclease biological function, with particular
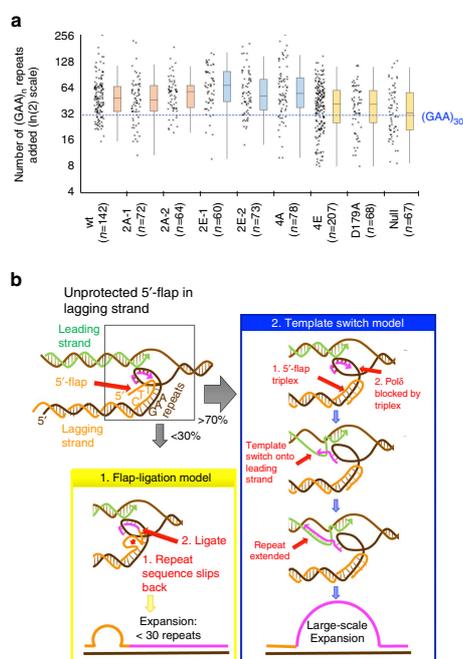
Figure 6 | FEN1 phosphate steering is essential for lagging strand precision at DNA repeats. (**a**) Graphed distributions of repeat expansion lengths shows that the majority of expansions in the wt, phosphate steering and D179A Rad27 mutants are >30 repeats. The numbers of added repeats in each strain are shown as scatter plots alongside box-and-whisker plots with 5 and 95% whiskers. The number of colonies tested are given in the parentheses. (**b**) Two models for repeat expansions driven by the presence of an unprocessed 5′-flap. In model 1 (left panel) the repeat on the 5′-flap ligates to the 3′-end of the oncoming Okazaki fragment followed by its equilibration into a loop. After the next round of replication, up to ∼30 repeats can be added (see text for details). In model 2 (right panel), the 5′-flap folds back forming a triplex, which blocks Pol(δ) DNA synthesis along the lagging strand template and promotes its switch to the nascent leading strand. This template switch mechanism explains the accumulation of large-scale repeat expansions >30 repeats.

deleterious effects on cells under replication stress and/or with damaged DNA.

Rad27 inactivation in yeast stimulates expansion of trinucleotide repeats relevant to human disease[46–48]. We therefore tested the effect of phosphate steering mutations on expansion rates of $(GAA)_n$ repeats using our system (Fig. 5b), which contains a $(GAA)_{100}$ tract situated in the intron of a *Ura3* reporter gene[49,50]. Addition of 10 or more repeats to the $(GAA)_{100}$ tract effectively blocks splicing, resulting in gene inactivation and rendering the yeast resistant to 5-fluoroorotic acid (5-FOA). The repeat expansion rates in the *rad27* knockout and in the severely catalytically impaired D179A active site metal ligand mutant was increased by ∼100-fold compared to wt (Fig. 5c,d). Strikingly for a non-active site mutant, phosphate steering 4E mutant exhibited a quantitatively similar phenotype. The double glutamate (2E-1, 2E-2) and 4A mutants showed intermediate (∼10-fold) increases

in repeat expansion rates. These results match growth characteristics of these mutants and emphasize the role of electrostatic interactions of the gateway basic residues with DNA in repeat-mediated genome instability.

Ligation of unprocessed 5′-flaps to the 3′-end of the approaching Okazaki fragment is proposed to cause the elevated repeat expansions in Rad27 mutants[48,51,52]. In this scenario, one expects added repeat lengths to be relatively short: less than the size of an Okazaki fragment. In fact, the major mutations caused by disruption of the *RAD27* gene in yeast were repeat-related expansions of 5–108 bases[53]. Recently, the median size of the unprocessed 5′-flap in *S. pombe FEN1* knockout was measured as 89 nts[54]. Given these numbers, the median expansion size of GAA repeats in our experimental system should be ∼30 repeats in Rad27 mutants.

To define the size distribution of expansion products, we measured the scale of repeat expansions in wt and Rad27 mutants described above via PCR (Fig. 6a). In the wt strain, median expansion size corresponded to 47 triplets[49]. The *rad27* knockout was different: median expansion size was 32 repeats, and Kolmogorov-Smirnov (KS) comparison confirms a significant difference from the wt strain ($P < 0.001$), which agrees with known flap size in *FEN1* knockouts[54]. The expansion scale in near-catalytic-dead (D179A) and 4E Rad27 mutants lies between the wt and knockout mutant: the median is 40 repeats and KS shows significant difference from wt ($P < 0.05$). Finally, the scale of expansions in 2E and 4A mutants is greater than wt with medians from 50 to 66 added repeats. Thus, the 100-fold increase in expansions (Fig. 5c) in phosphate steering mutants cannot be explained by an increase in small-scale expansions alone (caused by simple 5′-flap ligation), but is a consequence of larger expansions. Thus, most expansions in the Rad27 phosphate steering mutants originate via mechanisms distinct from simple 5′-flap ligation (see Discussion). Overall, these Rad27 results suggest that functional phosphate steering of 5′-flaps and dsDNA is vital for genome integrity: in promoting normal growth, in response to DNA damage, and in preventing trinucleotide repeat expansions.

Discussion

We sought to understand the mechanism whereby FEN1s binds and precisely incises ss-dsDNA junctions yet excludes hydrolysis of continuous DNA substrates, reasoning that this specificity was key to FEN1 functions during replication and repair. These investigations resolve controversies and improve our understanding of how FEN1-DNA interactions provide specificity and genome stability.

First, elucidation of a 5′-flap DNA threaded through the helical gateway/cap answers a longstanding question in eukaryotic FEN1 function and explains the selection of 5′-flap substrates with free 5′-termini. Although threading occurs in other enzymes, phosphate steering and inverted threading are extraordinary. For example, bacteriophage T5 5′-nuclease threads substrates[29], but positions the 5′ flap primarily through hydrophobic interactions to the 5′ flap nucleobases. The phosphodiester is closer to the metals than the nucleobases, consistent with its lower incision site specificity and tendency to cleave within the ssDNA 5′-flap. In other enzymes, threading selects for free ss 5′-termini that will undergo incision and there is no inversion. However, FEN1 preserves, rather than degrades, the threaded nucleic acid.

Second, our results uncover an essential function in FEN1 specificity and catalysis for phosphate steering, which we define as electrostatic interactions that dynamically control the phosphodiester backbone. The parallel effects of steering mutations on either endonucleolytic or exonucleolytic reactions
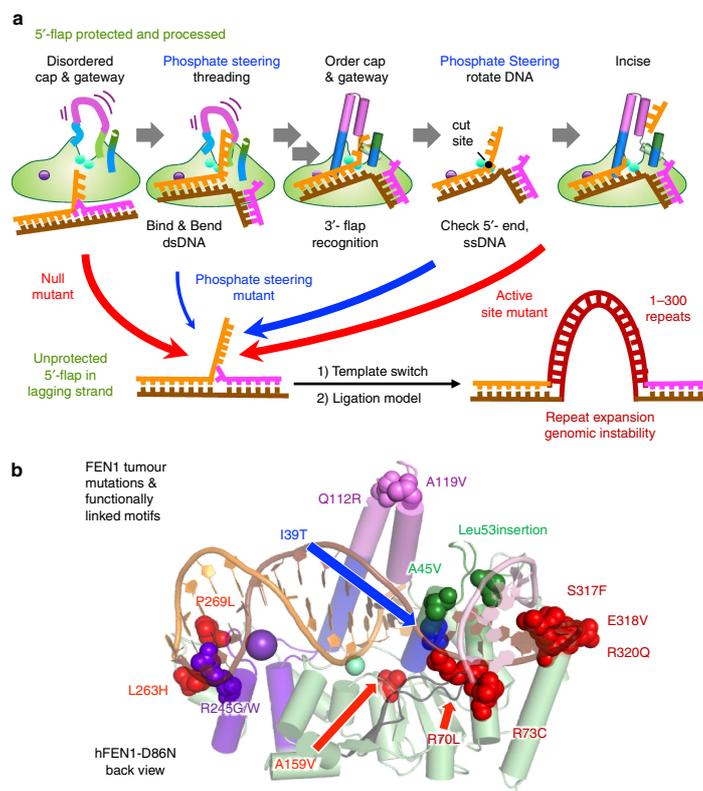
Figure 7 | Multiple motifs for FEN1 substrate recognition and hydrolysis ensure accurate incision activity and prevent genomic instability.
(**a**) Schematic model of the FEN1 mechanism emphasizing the functional role of phosphate steering in the dynamic processes of 5'-flap inverted threading and shifting of the duplex DNA towards the catalytic metals. (**b**) Tumour-associated mutations from breast, lung, skin, kidney, colorectal, ovarian and testicular cancers map to functionally-important structural motifs: dsDNA binding (P269L, L263H, R245G/W, R70L and R73G), 3'-flap binding (Leu53ins, A45V, S317F, E318V and R320Q), helical gateway/cap (I39T, Q112R and A119V) or active site (A159V).

(that is, on substrates with or without a 5'-flap) indicated involvement of basic gateway/cap residues in a rate-limiting step in the FEN1 catalytic pathway, that is, in moving the target phosphodiester bond from the ss-ds junction onto catalytic metal ions. Thus, phosphate steering may act in orienting the ss 5'-flap during threading (negative design to avoid off-target reactions) and moving the scissile phosphate into catalytic distance of the metals (positive design to enhance target reactions) (Fig. 7a). Notably, steering residue Arg104, is semiconserved throughout the superfamily suggesting that phosphate positioning occurs in other members.

Third, the proposed requirement for double base unpairing for the dsDNA to reach the active site metal ions[3] needs re-evaluation. Our observation of basepaired DNA contacting an active site metal ion with a water molecule positioned for in-line attack, would generate the arrangement for 'two-metal-ion' catalysis. This basepaired catalytically competent conformation appears at odds with spectroscopic characterization of FEN1 and GEN1 substrate complexes[19,32,44], and the inability of FEN1 to

process duplexes cross-linked at the terminal basepair[31,55], consistent with an unpairing mechanism. Yet, the DNA distortion seen in structures here (Supplementary Fig. 2E) provides an alternative explanation implying dsDNA can remain basepaired and roll onto the active site metal ions aided by Tyr40 rotation and by positive side chains on the helical gateway and cap.

Whereas replication fidelity is canonically based on sequence, it furthermore depends on sequence-independent specificity in FEN1. Importantly, structural elements critically involved in FEN1 function, including phosphate steering and inverted threading, require key residues distant from the active site metal ions. Indeed, clinically relevant FEN1 mutations compiled by The Cancer Genome Atlas (TCGA) and others[5,56,57] map to these structural elements (Fig. 7b). So, although tumour mutation data has been called 'a bewildering hodgepodge of genetic oddities'[58], for FEN1, there is a clear link of structurally-mapped mutations to compromised function, genomic instability and cancer. Although these mutations may retain nuclease activity, even

tiny off target activity risks toxicity and genomic instability, and replication mutations account for two-thirds of the mutations in human cancers[59].

We uncovered a role for phosphate steering in triplet $(GAA)_n$ repeat expansions, that also implicates template switching from the lagging strand due to FEN1 defects. Most expansions in Rad27 phosphate steering mutants were large-scale (> 30 repeats; Fig. 6a) which is difficult to explain by the canonical flap-ligation model for repeat instability[46,48,51,60]. In this model, an unprocessed 5′-flap is ligated to the 3′-end of the approaching Okazaki fragment (Fig. 6b, left), limiting the length of expansions to the size of those flaps. Recently, the median size of the 5′-flap in a FEN1 knockout was found to be 89 nts[54], that is, ∼30 triplet repeats. Since median expansion size in phosphate steering mutants is >30 repeats, we propose that besides a flap-ligation model, a template switch between nascent repetitive strands occurs as a replication fork stumbles through the repeat sequence[50] (Fig. 6b, right). Unprocessed $(TTC)_n$ 5′-flaps of the Okazaki fragments may form a stable triplex[61] with the downstream repetitive run. This could block displacement synthesis by the lagging strand polymerase[62] and prompt it to search for a new template. Large-scale repeat expansions would then occur when the polymerase switches template—continuing DNA synthesis along the nascent leading strand. As a starting repeat gets longer, larger expansions become feasible, consistent with the progressive increase in expansion amplitudes with the length of original repeat tract, as observed in human pedigrees[63].

The profound stimulation of large-scale expansions in the phosphate steering mutants unexpectedly sheds light on the molecular mechanism of template switching. A priori, either a nascent leading strand can switch onto the nascent lagging strand to use it as a template[49,50], or the nascent lagging strand can switch onto the nascent leading strand serving as a template[47]. Since it is the lagging strand synthesis and specifically Okazaki fragment maturation that are unraveled in Rad27 mutants, the sheer magnitude of their effects on large-scale repeat expansions implies that the lagging strand likely switches onto the nascent leading strand accounting for the repeat instability, and this merits further biochemical investigation. Another question emerges from biochemical studies of FEN1 functions during long-patch base excision repair where expansions occur on dysregulation of DNA handoffs from polymerase β to FEN1 (ref. 64), which suggests studies to investigate whether phosphate steering may prevent expansions during long-patch repair.

In summary, we find FEN1 phosphate steering energetically promotes dsDNA rotation into the active site and inverted threading of the 5′-flap to enforce efficiency and fidelity in replication and repair. Interestingly, elevated FEN1 expression safeguards against repeat instability in somatic tissues[6]. Phosphate steering mutations could thus be the trans-modifiers of repeat expansions during either somatic, or intergenerational transmissions in human disease[65]. Moreover, as the basic residues implicated in phosphate steering are largely conserved in the 5′-nuclease superfamily, control over the $+1$ and -1 phosphates may be a superfamily-conserved mechanism.

Methods

Site-directed mutagenesis. Plasmids for expression of mutant proteins were prepared from either the pET29b-hFEN1Δ336(wt) or pET28b-hFEN1-(His)$_6$ constructs, as indicated above, following the protocol outlined in the QuikChange site-directed mutagenesis kit (Agilent Technologies, Inc.). Mutagenic primers were purchased from Fisher Scientific, with desalting, then reconstituted in ultrapure water and used as supplied. Mutagenic primer sequences were as follows: D86N, 5′-ggcggcttgccattaaagacatacacggggctt-3′ and 5′-aagcccgtgtatgtctttaatggcaagccgcc-3′; R103A, 5′-caaacgcagtgaggcgcgggctgaggca-3′ and 5′-tgcctcagcccgcgcctcactgcgtttg-3′; R103E, 5′-ccaaacgcagtgaggagcgggctgaggcag-3′ and 5′-ctgcctcagcccgctcctcactgcgtt tgg-3′; R104A, 5′-ctctgcctcagcccgcccgctcactgcgt-3′ and 5′-acgcagtgagcgggcggctgaggc

agag-3′; R104E, 5′-acgcagtgagcgggaggctgaggcagag-3′ and 5′-ctctgcctcagcctcccgctcac tgcgt-3′; R129A, 5′-ttagtgaccttcaccagcgccttagtgaatttttccacctc-3′ and 5′-gaggtggaaaaa ttcactaaggcgctggtgaaggtcactaa-3′; R129E, 5′-ttagtgaccttcaccagctccttagtgaatttttccac ctc-3′ and 5′-gaggtggaaaaattcactaaggagctggtgaaggtcactaa-3′; K132A, 5′-cactaagcgg ctggtggcggtcactaagcagcac-3′ and 5′-gtgctgcttagtgaccgccaccagccgccttagtg-3′; K132E, 5′-gctgcttagtgacctccaccagccgcttagt-3′ and 5′-actaagcggctggtggaggtcactaagcagc-3′; R103ER104E, 5′-gcttctctgcctcagcctcctcctcactgcgtttggcca-3′ and 5′-tggccaaacgcagtga ggaggaggctgaggcagagaagc-3′; R129EK132E, 5′-gtgctgcttagtgacctccaccagctccttagtgaa tttttccacc-3′ and 5′-ggtggaaaaattcactaaggagctggtggaggtcactaagcagcac-3′.

Protein expression. Plasmids encoding R100AΔ336 and D233NΔ336 human FEN1 for crystallography were generated by site-directed mutagenesis from the pET29b-hFEN1Δ336(wt) construct bearing a PreScission protease site and (His)$_6$-tag after residue 336 of the wt sequence[21]. Full length wt hFEN1 was encoded using the pET28b-hFEN1-(His)$_6$ vector reported previously[12], and all reported mutants were generated from this by site-directed mutagenesis. Proteins were expressed in Rosetta (DE3)pLysS competent cells grown in $2 \times$ YT media or Terrific Broth to an OD$_{600}$ of 0.6–0.8 at 37 °C then induced by addition of 1 mM IPTG, followed by incubation at 18 °C for 18–24 h. Cells were collected by centrifugation at 6,000 g/4 °C, washed with PBS, then resuspended in buffer IMAC-A1 (20 mM Tris pH 7.0, 1.0 M NaCl, 5 mM imidazole, 0.02% NaN$_3$, 5 mM β-mercaptoethanol supplemented with SIGMAFAST protease inhibitor tablets and 1 mg ml^{-1} chicken egg white lysozyme). Each suspension was kept on ice for 2 h then stored frozen at -20 °C until further processing, as detailed below.

Purification of hFEN1 D86NΔ336 and R100AΔ336 and D233NΔ336. All steps were carried out at 4 °C. Chromatography was on an ÄKTA system with flow rate of 5.0 ml min^{-1} unless stated otherwise. Columns were from GE Healthcare, unless stated otherwise. Frozen lysates were thawed on ice and homogenized by sonication. Next, 0.1 volume of a 10% v/v TWEEN 20 solution was added. The mixture was clarified by centrifugation at 30,000 g for 30 min. Supernatant was loaded onto a Ni-IDA affinity column, which was then washed with 5 column volumes (CV) of buffer IMAC-A1, 5 CV of buffer IMAC-A2 (20 mM Tris pH 7.0, 0.5 M NaCl, 40 mM imidazole, 0.02% NaN$_3$, 0.1% v/v TWEEN 20, 5 mM β-mercaptoethanol). FEN1 was eluted with 5 CV of buffer IMAC-B1 (250 mM imidazole pH 7.2, 0.5 M NaCl, 0.02% NaN$_3$, 5 mM β-mercaptoethanol). Pooled fractions were diluted 1:5 with water and then loaded onto a HiPrep Heparin FF 16/10 column. The column was washed with 5 CV buffer HEP-A1 (25 mM Tris pH 7.5, 1 mM CaCl$_2$, 0.02% NaN$_3$, 20 mM β-mercaptoethanol). FEN1 was eluted with a linear gradient of 100% HEP-A1 to 100% HEP-A2 (25 mM Tris pH 7.5, 1 mM CaCl$_2$, 1.0 M NaCl, 0.02% NaN$_3$, 20 mM β-mercaptoethanol) in 20 CV. Pooled FEN1 fractions were diluted by slow addition of two volumes of 3.0 M (NH$_4$)$_2$SO$_4$ at 4 °C. The solution was loaded onto a HiPrep Phenyl FF (high sub) 16/10 phenylsepharose column. The column was washed with 7 CV buffer P/S-B1 (25 mM Tris pH 7.5, 2.0 M (NH$_4$)$_2$SO$_4$, 2 mM CaCl$_2$, 0.02% NaN$_3$, 20 mM β-mercaptoethanol). FEN1 was eluted with a gradient of 100% P/S-B1 to 100% P/S-A1 (25 mM Tris pH 7.5, 10% v/v glycerol, 1 mM CaCl$_2$, 0.02% NaN$_3$, 20 mM β-mercaptoethanol) in 20 CV. Pooled fractions were concentrated to ∼7 ml using an Amicon stirred cell (Merck Millipore), then passed through 5×5 ml HiTrap Desalting columns arranged in tandem, injected in 1.5 ml portions. The desalting columns were equilibrated in $1 \times$ TBS supplemented with 1 mM EDTA and 1 mM DTT, and eluted with the same buffer. Combined protein-containing eluent (35–40 ml) was treated with PreScission protease (20 μl of activity 10 U μl^{-1}) and incubated at 4 °C overnight. Complete cleavage of the (His)$_6$ tag was verified by SDS-PAGE, then the protein solution concentrated to 5 ml using a Vivaspin 20 Centrifugal Concentrator (10,000 MWCO). A final purification step at a 0.5 ml min^{-1} flow rate with a Sephacryl S-100 HR column, equilibrated with 2 CV of $2 \times$ SB (100 mM HEPES pH 7.5, 200 mM KCl, 2 mM CaCl$_2$, 10 mM DTT, 0.04% NaN$_3$). FEN1 fractions were pooled and the protein concentration determined by A$_{280}$, using the calculated OD$_{280}$. The solution was concentrated to >200 μM using a Vivaspin 20 Centrifugal Concentrator (10,000 MWCO). Finally, the solution was mixed 1:1 v/v with cold glycerol, placed on a roller mixer until homogenous, then divided into 1 ml aliquots and stored as a 100 μM stock solution at -20 °C.

Crystallography of mutant FEN1-DNA complexes. hFEN1 mutants were crystallized with DF substrates (S5,1) or (S4,1) of slightly different sequence (desalted purity from IDT, Supplementary Fig. 1). hFEN1-D86NΔ336 (19 mg ml^{-1}) was mixed in volumetric ratio 1:2:1 with 4.25 mM SmSO$_4$, and 1.3 mM substrate S5,1-D86N. This mixture was in turn combined 1:1 with 12% mPEG 2,000, 20% saturated KCl, 5% ethylene glycol, 100 mM HEPES pH 7.5. Crystals were collected after 5 days at 15 °C. hFEN1-R100AΔ336 (19 mg ml^{-1}) was mixed in volumetric ratio 1:2:1 with 3.75 mM SmSO$_4$, and 1.3 mM substrate S4,1-R100A. This mixture was in turn combined 1:1 with 22% mPEG 2000, 20% saturated KCl, 5% ethylene glycol, 100 mM HEPES pH 7.5. Crystals were collected after ∼3 weeks at 15 °C. hFEN1-D233NΔ336 (8.2 mg ml^{-1}) with 1.6 mM SmSO$_4$, 0.25 mM substrate S4,1-D233N was mixed 1:1 with 24% mPEG 2000, 20% satu-rated KCl, 5% ethylene glycol, 100 mM HEPES pH 7.5. hFEN1-D86N data was collected at 0.98 Å (SSRL beamline 12-2) and processed with HKL2000. hFEN1-R100A data was collected at 0.98 Å (SSRL beamline 9-2) and processed with XDS.

hFEN1-D233N data was collected at 1.12 Å (ALS beamline 12.3.1) and processed with HKL2000. hFEN1-D86N, hFEN1-R100 and hFEN1-D233N crystals diffracted to 2.8, 2.65 and 2.1 Å, respectively. Structures were solved by molecular replacement using PHASER[66] with human FEN1 protein as the search model and refined in PHENIX[67] with rounds of manual rebuilding in COOT[68]. For hFEN1-R100A, we refined the model using higher diffraction data to 2.1 Å, based on the theory that cutting off resolution at an arbitrary point leads to series termination errors. Flexible regions became more visible and we could follow the path of the 5′-flap more easily. The R and R$_{free}$ measures dropped substantially. We used a higher resolution structure (PDB code: 3Q8K) for reference in refinement. For the three structures, anomalous differences from the Sm^{3+} atoms were used in refinement and modelling. In the active sites of the hFEN1-D86N, hFEN1-R100A and hFEN1-D233N structures there were, respectively, one, three and four Sm^{3+} atoms, with partial occupancy. For all structures, there were no Ramachandran outliers. For hFEN1-D86n, 95% were favoured and 5% were allowed. For hFEN1-R100A, 96% were favoured and 4% were allowed. For hFEN1-D233N, 98% were favoured and 2% were allowed. Structure figures were created in PyMol (Schrödinger, LLC). Movies were created in Chimera[69].

Protein purification of full-length FEN1 proteins. All steps were carried out using an ÄKTA FPLC system at 4 °C, at a flow rate of 5.0 ml min^{-1} unless stated otherwise. Frozen/thawed lysates were loaded onto a Ni-IDA column, followed by washing with 4 CV buffer IMAC-A1, 4 CV buffer IMAC-A2, a gradient of 100% IMAC-A2 to 100% IMAC-B1 in 2 CV, then 4 CV IMAC-B1. Pooled fractions were diluted 1:1 with 20 mM β-mercaptoethanol and loaded onto a 5 ml HiTrap Q FF column to remove nucleic acid contamination, with a 20 CV elution gradient from 0 to 1.0 M NaCl in 20 mM Tris pH 8.0, 1 mM EDTA, 0.02% NaN$_3$, 20 mM β-mercaptoethanol. The flow-through containing FEN1 was diluted 1:4 with 20 mM β-mercaptoethanol and passed through the HiPrep Heparin FF 16/10 column as above. The purified FEN1 was exchanged into 2 × SB using a HiPrep 26/10 Desalting column, concentrated and prepared for storage as detailed above. Proteins requiring further purification (wt hFEN1 and D233N) were passed through the HiPrep Phenyl FF (high sub) 16/10 phenylsepharose column, as above. Protein-containing fractions were pooled and concentrated to 5 ml using an Amicon stirred cell, subjected to gel filtration and prepared for storage as outlined above.

Oligonucleotide synthesis. The DNA oligonucleotides used for crystallization (Supplementary Fig. 1) were purchased from IDT as desalted oligonucleotides. They were resuspended in 10 mM HEPES 7.5, 50 mM KCl, 0.5 mM EDTA and annealed at ~1–2 mM. The DNA oligonucleotides used to construct the kinetic substrates (Supplementary Fig. 1) were purchased from DNA Technology A/S (Denmark) with HPLC purification. Except for E1 and E2 (Supplementary Fig. 1A), the oligonucleotides as supplied were reconstituted in ultrapure water and concentrations of stock solutions determined using calculated extinction coefficients (OD$_{260}$). Oligonucleotides E1 and E2 required additional HPLC purification, which was carried out using an OligoSep GC cartridge (Transgenomic; #NUC-99-3860) using buffers A (100 mM triethylammonium acetate pH 7.0, 0.025% v/v acetonitrile) and B (100 mM triethylammonium acetate pH 7.0, 25% acetonitrile) and a gradient of 5–50% B over 18 min, at 50 °C and a flow rate of 1.5 ml min^{-1}. Purified oligonucleotide in solution was loaded onto a 5 ml HiTrap DEAE FF column equilibrated with 3 CV of buffer C (10 mM Tris pH 7.5, 100 mM NaCl, 1 mM EDTA, 0.02% NaN$_3$). The column was washed with a further 3 CV of buffer C, then eluted using a step gradient of 100% buffer C-100% buffer D (10 mM Tris pH 7.5, 1.0 M NaCl, 1 mM EDTA, 0.02% NaN$_3$) in 3 CV. Fractions containing DNA were desalted into ultrapure water using NAP-25 columns. Desalted samples were dried then reconstituted as above. DNA constructs were annealed in 1 × FB (50 mM HEPES pH 7.5, 100 mM KCl) for at least 5 min at 95 °C, then left at ambient temperature for 30 min.

FRET binding assay. Values for K_d were obtained using sequential titration of the appropriate enzyme into a 10 nM solution of the appropriate DNA construct, according to the reported protocol[44]. FRET efficiencies (E) were determined using the (ratio)$_A$ method by measuring the enhanced acceptor fluorescence at 37 °C. The steady state fluorescent spectra of 10 nM non-labelled (NL) trimolecular, donor-only labelled (DOL) and doubly labelled (DAL) DNA substrates (Supplementary Fig. 1A,B) were recorded using a Horiba Jobin Yvon FluoroMax-3 fluorometer. For direct excitation of the donor (fluorescein, DOL) or acceptor (TAMRA, AOL), the sample was excited at 490 nm or 560 nm (2 nm slit width) and the emission signal collected from 515–650 nm or 575–650 nm (5 nm slit width). Emission spectra were corrected for buffer and enzyme background signal by subtracting the signal from the NL DNA sample. In addition to 10 nM of the appropriate DNA construct, samples contained 10 mM CaCl$_2$ or 2 mM EDTA, 110 mM KCl, 55 mM HEPES pH 7.5, 0.1 mg ml^{-1} bovine serum albumin and 1 mM DTT. The first measurement was taken before the addition of protein with subsequent readings taken on the cumulative addition of the appropriate enzyme in the same buffer, with corrections made for dilution. Transfer efficiencies (E) were determined according to equation (1), where F_{DA} and F_D represent the fluorescent signal of the DAL and DOL DNA at the given wavelengths, respectively (for

example, $F_{DA}(\lambda^D_{EX}, \lambda^A_{EM})$, denotes the measured fluorescence of acceptor emission on excitation of the donor, for DAL DNA); ε^D and ε^A are the molar absorption coefficients of donor and acceptor at the given wavelengths; and $\varepsilon^D(490)/\varepsilon^A(560)$ and $\varepsilon^A(490)/\varepsilon^A(560)$ are determined experimentally from the absorbance spectra of DAL and the excitation spectra of singly TAMRA-AOL, respectively. Energy transfer efficiency (E) was fitted by non-linear regression in the Kaleidagraph program to equation (2), where E_{max} and E_{min} are the maximum and minimum energy transfer values, [S] is the substrate concentration, [P] is the protein concentration and K_{bend} is the bending equilibrium dissociation constant of the protein substrate [PS] complex.

$$E = (ratio)_A / \left(\frac{\varepsilon^D(490)}{\varepsilon^A(560)} \right) - \left(\frac{\varepsilon^A(490)}{\varepsilon^A(560)} \right), \qquad (1)$$

Where $(ratio)_A = \frac{F_{DA}(\lambda^A_{EX}, \lambda^A_{EM}) - N \cdot F_D(\lambda^D_{EX}, \lambda^A_{EM})}{F_{DA}(\lambda^A_{EX}, \lambda^A_{EM})}$,

And $N = F_{DA}(\lambda^D_{EX}, \lambda^D_{EM}) / F_D(\lambda^D_{EX}, \lambda^D_{EM})$,

$$E = E_{min} + \frac{(E_{max} - E_{min})}{2[S]} \left[([S] + [P] + K_{bend}) - \sqrt{([S] + [P] + K_{bend})^2 - 4[S][P]} \right]$$

$$(2)$$

Donor (fluorescein) was excited at 490 nm with emission sampled as the average value of the signal between 515 and 525 nm, and acceptor (TAMRA) was excited at 560 nm with emission averaged between 580 and 590 nm.

Multiple turnover rates. Reaction mixtures (final volume 180 μl) were prepared in 1.5 ml microcentrifuge tubes with 50 nM final substrate concentration (S5,1; S0,1-5P; S0,1-5OH; or S0,1-5FAM) and incubated at 37 °C before addition of enzyme to initiate the reaction. The final composition of each reaction mixture was 1 × RB (55 mM HEPES pH 7.5, 110 mM KCl, 8 mM MgCl$_2$, 0.1 mg ml^{-1} BSA) supplemented with 1 mM DTT. Enzyme concentrations were chosen to give ~15% cleavage after 20 min, and any data points showing greater cleavage were discarded due to effects of substrate depletion. For substrates S5,1 and S0,1-5FAM, aliquots (20 μl) of each reaction mixture were quenched into 250 mM EDTA (50 μl) at seven different time points—typically 2, 4, 6, 8, 10, 12 and 20 min—and reaction progress monitored by dHPLC analysis using a WAVE system equipped with an OligoSep cartridge (4.6 × 50 mm; ADS Biotec). The 6-FAM label was detected by fluorescence (excitation 494 nm, emission 525 nm) and product(s) separated from unreacted substrate using the following gradient: 5–30% B over 1 min; 30–55% B over 4.5 min; 55–100% B over 1.5 min; 100% B for 1.4 min; ramp back to 5% B over 0.1 min; hold at 5% B for 2.4 min, where A is 0.1% v/v MeCN, 1 mM EDTA, 2.5 mM tetrabutylammonium bromide and B is 70% v/v MeCN, 1 mM EDTA, 2.5 mM tetrabutylammonium bromide[12]. Initial rates (v, nM min^{-1}) were determined by linear regression of plots of product concentration versus time and adjusted for enzyme concentration to give normalized rates (v/[E], min^{-1}). For analysis of exonucleolytic activity, reactions with substrates S0,1-5P and S0,1-5OH were run as above but quenched in 98% deionised formamide containing 10 mM EDTA. Time points and enzyme concentrations were selected to give 10–15% cleavage at the reaction end point (≥20 min). The quenched samples were analysed by capillary electrophoresis as detailed below, then rates determined and normalized as above.

Analysis of reaction aliquots by capillary electrophoresis. Capillary electrophoresis was performed on the P/ACE MDQ Plus system (Beckman Coultier) using the ssDNA 100-R Kit (AB SciEx UK Limited; #477480) according to the manufacturer's instructions. Briefly, the supplied capillary (ID 100 μm, 30 cm long; 20 cm to detection window) was loaded with the commercially supplied gel using 70 psi of pressure for 5 min. The capillary was then equilibrated between two buffer vials containing Tris-Borate-Urea buffer (AB SciEx UK Limited; #338481) at 3, 5 and 9.3 kV for 2, 2 and 10 min, respectively, with a ramp time of 0.17 min. Samples were then run using a 5 s electrokinetic injection preceded by a 1 s plug injection of deionised water, before separation over 20 min with a voltage of 9.3 kV applied between two buffer vials; runs were carried out at 50 °C with constant pressure of 40 psi maintained on both sides of the capillary. The gel was replaced every five sample runs and running buffer was replaced every 20 sample runs. Peak detection was by laser induced fluorescence (LIF) using an excitation wavelength of 488 nm and a 520 nm filter to measure the emission. The electrophoretograms were integrated to determine the concentration of product formed at each time point. Initial rates of reaction (v, nM min^{-1}) were then obtained using linear regression, and converted to the reported normalized rates (v/[E], min^{-1}) as above.

Single turnover rapid quench experiments. Rapid quench experiments for determination of single turnover rate were carried out for wt hFEN1 and the mutants R104A, K132A, R103AR129A and D233N. Reactions were carried out at 37 °C using an RQF-63 device (HiTech Limited, Salisbury, UK)[12,70]. Premix stock solutions of enzyme and substrate were prepared at 2 × final concentration in reaction buffer (55 mM HEPES pH 7.5, 110 mM KCl, 8 mM MgCl$_2$, 2.5 mM DTT and 0.1 mg ml^{-1} BSA) and kept on ice until use. For individual reactions, the two 80 μl sample injection loops of the instrument (lines A and B) were filled with aliquots of enzyme and substrate stock, respectively. The syringe feeding the

quench line contained 1.5 M NaOH, 20 mM EDTA. Individual reactions were carried out using a controlled time delay of between 0.0091 and 51.241 s before quenching, with final concentrations of 5 nM substrate S5,1 and either 400 nM or 1,000 nM enzyme, as indicated (Supplementary Fig. 4C,D). Quenched reaction mixtures were analysed by dHPLC as described above for multiple turnover reactions, and rates were derived from curves consisting of at least 14 individual time points. The single turnover rate of the reaction was obtained as the first-order rate constant (k_{ST}) derived using nonlinear least squares regression for a one- or two-phase exponential in GraphPad Prism 6.05 (GraphSoftware, Inc.). Model selection was by statistical analysis using Aikake's Information Criteria (AIC).

Benchtop single turnover experiments. For the remaining proteins—hFEN1 mutants R104AK132A, R103ER129E, R104EK132E, QUAD-E (R103E/R104E/R129E/K132E) and D181A—reactions to determine single turnover rates were carried out using manual sampling, as described for the multiple turnover reactions above, except using 5 nM substrate S5,1 and an enzyme concentration of either 400 nM or 1,000 nM as indicated in each case (Supplementary Fig. 4C,D). A final reaction volume of 360 µl was used, permitting sampling of 14 time points per tube, which were typically chosen to span a reaction duration of at least 20 half-times. Quenched samples were analysed by dHPLC as detailed above, then single turnover rates were derived as described for the rapid quench experiments.

Yeast strain construction. To construct the individual yeast mutants, the *hphMX4* hygromycin resistance marker was first integrated downstream of *Rad27*, replacing genomic region ChrXI:224,681–224,712, in a strain containing the *Ura3-(GAA)₁₀₀* cassette[49] derived from parent strain CH1585 (MATa *leu2-Δ1, trp1-Δ63, ura3–52*, and *his3–200*). The rate of (GAA)₁₀₀ expansion in this strain (designated *Rad27-Hyg*) was indistinguishable from the wild type strain not carrying the downstream *hphMX4* cassette. Genomic DNA from *Rad27-Hyg* was used as a template for PCR with a ~100 bp forward primer containing the specific mutations and a reverse primer downstream of the *hphMX4* cassette. These PCR products were used to transform the wt (GAA)₁₀₀ strain with selection on 200 µg ml⁻¹ hygromycin. Transformants were screened by PCR and/or restriction digest, and the full-length sequences of the mutated *Rad27* alleles were verified by Sanger sequencing. The length of the starting (GAA)₁₀₀ tract in the mutant strains was confirmed by PCR using primers A2 (5′-CTCGATGTGCAGAACCTGAAGC TTGATCT-3′) and B2 (5′-GCTCGAGTGCAGACCTCAAATTCGATGA-3′).

Yeast spot assay. Fivefold serial dilutions were made on an equivalent starting number of cells for each strain. A 2.5 µl aliquot of each dilution was spotted onto YPD, YPD with 10 µg ml⁻¹ camptothecin, or YPD with 100 mM hydroxyurea. For UV treatment, cells spotted onto YPD were immediately irradiated using a UV Stratalinker 1,800 (Stratagene).

Fluctuation assay and expansion rates. At least two independent isolates of each yeast mutant were diluted from frozen stocks and grown for 40 h on solid rich growth media (YPD) supplemented with uracil. 16 individual colonies (8 per isolate) were dissolved in 200 µl of water and serially diluted. Appropriate dilutions were plated on synthetic complete media containing 0.09% 5-fluoro-orotic acid (5FOA) to select for large-scale expansion events or YPD to assess total cell number. Colonies on each plate were counted after three days of growth. For each mutant, at least 96 representative 5FOA colonies (8–12 per plate) were analysed for large-scale GAA expansion via PCR using primers A2 and B2 followed by agarose gel electrophoresis (1.5% agarose in 0.5X TBE). To determine a true expansion rate (as opposed to a gene inactivation rate), the number of 5FOA-r colonies counted per plate was adjusted by the overall percentage of GAA expansion events observed for that mutant. Expansion rates were calculated using the Ma-Sandri-Sarkar maximum likelihood estimator method with a correction for plating efficiency determined as $z-1/z\ln(z)$, where z is the fraction of the culture analysed (Rosche and Foster, 2000). PCR product lengths for the calculation of GAA expansion size were determined using cubic spline interpolation on Total Lab Quant software. Kolmogorov-Smirnov comparison of expansion lengths between strains was conducted using SPSS software—non-parametric testing of independent samples. Genotype information for each strain used is shown in Supplementary Table 1.

Extraction of Rad27 proteins and western blotting. Wt and mutant strains in mid-log phase (OD₆₀₀ 0.6–0.8, 10 mls) were pelleted, washed with water and frozen. Pellets were resuspended in 150 µl of distilled water, mixed with an equal volume of 0.6 M NaOH with a 10 min incubation at room temperature. After low speed centrifugation (153 g) for 5 min, the supernatant was removed and each pellet resuspended in SDS sample loading buffer (60 mM Tris-HCl pH 6.8, 4% β-mercaptoethanol, 4% SDS, 0.01% bromophenol blue, 5% glycerol). The samples were boiled for 3 min, then 10 µl of each separated by 4–12% SDS–PAGE gel (Invitrogen) followed by Western blotting using anti-RAD27 goat polyclonal antibody (1:125 dilution; Santa Cruz Biotechnology, #sc-26719), donkey anti-goat IgG-HRP secondary antibody (1:2,500; Santa Cruz Biotechnology; #sc-2020) and visualized using an ECL detection kit (GE Healthcare). A nonspecific band present in all lanes was used as a loading control (Fig. 5d).

Data availability. Coordinates and structure factors are deposited with the Protein Data Bank (PDB) under the accession codes: 5UM9 (D86N), 5KSE (R100A), and 5K97 (D233N). The data that support the findings of this study are available from the corresponding authors on request.

References

1. Balakrishnan, L. & Bambara, R. A. Flap endonuclease 1. *Annu. Rev. Biochem.* **82**, 119–138 (2013).
2. Finger, L. D. *et al.* in *The Eukaryotic Replisome: A Guide to Protein Structure and Function* (ed. MacNeill, S. Ch. 16 (Springer, 2012).
3. 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.* **37**, 74–84 (2012).
4. Larsen, E., Gran, C., Sæther, B. E., Seeberg, E. & Klungland, A. Proliferation failure and gamma radiation sensitivity of FEN1 null mutant mice at the blastocyst stage. *Mol. Cell. Biol.* **23**, 5346–5353 (2003).
5. Zheng, L. *et al.* FEN1 mutations result in autoimmunity, chronic inflammation and cancers. *Nat. Med.* **13**, 812–819 (2007).
6. Mason, A. G. *et al.* Expression levels of DNA replication and repair genes predict regional somatic repeat instability in the brain but are not altered by polyglutamine disease protein expression or age. *Hum. Mol. Genet.* **23**, 1606–1618 (2014).
7. Lam, J. S. *et al.* Flap endonuclease 1 is overexpressed in prostate cancer and is associated with a high Gleason score. *BJU Int.* **98**, 445–451 (2006).
8. Singh, P. *et al.* Overexpression and hypomethylation of flap endonuclease 1 gene in breast and other cancers. *Mol. Cancer Res.* **6**, 1710–1717 (2008).
9. van Pel, D. M. *et al.* An Evolutionarily conserved synthetic lethal interaction network identifies FEN1 as a broad-spectrum target for anticancer therapeutic development. *PLoS Genet.* **9**, e1003254 (2013).
10. Exell, J. C. *et al.* Cellularly active N-hydroxyurea FEN1 inhibitors block substrate entry to the active site. *Nat. Chem. Biol.* **12**, 815–821 (2016).
11. Chapados, B. R. *et al.* Structural basis for FEN-1 substrate specificity and PCNA-mediated activation in DNA replication and repair. *Cell* **116**, 39–50 (2004).
12. Finger, L. D. *et al.* The 3′-flap pocket of human flap endonuclease 1 is critical for substrate binding and catalysis. *J. Biol. Chem.* **284**, 22184–22194 (2009).
13. Rashid, F. *et al.* Single-molecule FRET unveils induced-fit mechanism for substrate selectivity in flap endonuclease 1. *Elife* **6**, e21884 (2017).
14. Bhagwat, M. & Nossal, N. G. Bacteriophage T4 RNase H removes both RNA primers and adjacent DNA from the 5′ end of lagging strand fragments. *J. Biol. Chem.* **276**, 28516–28524 (2001).
15. Garforth, S. J., Ceska, T. A., Suck, D. & Sayers, J. R. 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 (1999).
16. Mitsunobu, H., Zhu, B., Lee, S.-J., Tabor, S. & Richardson, C. C. Flap endonuclease activity of gene 6 exonuclease of bacteriophage T7. *J. Biol. Chem.* **289**, 5860–5875 (2014).
17. Allen Lee, M., Hodskinson Michael, R. G. & Sayers Jon, R. Active site substitutions delineate distinct classes of eubacterial flap endonuclease. *Biochem. J.* **418**, 285–292 (2009).
18. Lee, S. H. *et al.* Human Holliday junction resolvase GEN1 uses a chromodomain for efficient DNA recognition and cleavage. *Elife* **4**, e12256 (2015).
19. Liu, Y. *et al.* Crystal structure of a eukaryotic GEN1 resolving enzyme bound to DNA. *Cell Rep.* **13**, 2565–2575 (2015).
20. Orans, J. *et al.* Structures of human exonuclease 1 DNA complexes suggest a unified mechanism for nuclease family. *Cell* **145**, 212–223 (2011).
21. Tsutakawa, S. E. *et al.* Human flap endonuclease structures, DNA double-base flipping, and a unified understanding of the FEN1 superfamily. *Cell* **145**, 198–211 (2011).
22. Mietus, M. *et al.* Crystal structure of the catalytic core of Rad2: insights into the mechanism of substrate binding. *Nucleic Acids Res.* **42**, 10762–10775 (2014).
23. Gloor, J. W., Balakrishnan, L. & Bambara, R. A. Flap endonuclease 1 mechanism analysis indicates flap base binding prior to threading. *J. Biol. Chem.* **285**, 34922–34931 (2010).
24. Patel, N. *et al.* 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 (2012).
25. Hwang, K. Y., Baek, K., Kim, H.-Y. & Cho, Y. The crystal structure of flap endonuclease-1 from Methanococcus jannaschii. *Nat. Struct. Mol. Biol.* **5**, 707–713 (1998).
26. Sakurai, S. *et al.* Structural basis for recruitment of human flap endonuclease 1 to PCNA. *EMBO J.* **24**, 683–693 (2005).
27. Shen, B., Nolan, J. P., Sklar, L. A. & Park, M. S. Functional analysis of point mutations in human flap endonuclease-1 active site. *Nucleic Acids Res.* **25**, 3332–3338 (1997).
28. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).

29. AlMalki, F. A. *et al.* Direct observation of DNA threading in flap endonuclease complexes. *Nat. Struct. Mol. Biol.* **23,** 640–646 (2016).

30. Qiu, J. *et al.* Interaction interface of human flap endonuclease-1 with its DNA substrates. *J. Biol. Chem.* **279,** 24394–24402 (2004).

31. Beddows, A. *et al.* Interstrand disulfide crosslinking of DNA bases supports a double nucleotide unpairing mechanism for flap endonucleases. *Chem. Commun.* **48,** 8895–8897 (2012).

32. Finger, L. D. *et al.* Observation of unpaired substrate DNA in the flap endonuclease-1 active site. *Nucleic Acids Res.* **41,** 9839–9847 (2013).

33. Syson, K. *et al.* Three metal ions participate in the reaction catalyzed by T5 flap endonuclease. *J. Biol. Chem.* **283,** 28741–28746 (2008).

34. 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.* **10,** 25–33 (1991).

35. Nowotny, M. & Yang, W. Stepwise analyses of metal ions in RNase H catalysis from substrate destabilization to product release. *EMBO J.* **25,** 1924–1933 (2006).

36. Freudenthal, B. D., Beard, W. A., Shock, D. D. & Wilson, S. H. Observing a DNA polymerase choose right from wrong. *Cell* **154,** 157–168 (2013).

37. Molina, R. *et al.* Visualizing phosphodiester-bond hydrolysis by an endonuclease. *Nat. Struct. Mol. Biol.* **22,** 65–72 (2015).

38. Nakamura, T., Zhao, Y., Yamagata, Y., Hua, Y. J. & Yang, W. Watching DNA polymerase eta make a phosphodiester bond. *Nature* **487,** 196–201 (2012).

39. Gao, Y. & Yang, W. Capture of a third Mg^{2+} is essential for catalyzing DNA synthesis. *Science* **352,** 1334–1337 (2016).

40. Yang, W., Weng, P. J. & Gao, Y. A new paradigm of DNA synthesis: three-metal-ion catalysis. *Cell Biosci.* **6,** 51 (2016).

41. Sobhy, M. A., Joudeh, L. I., Huang, X., Takahashi, M. & Hamdan, S. M. Sequential and multistep substrate interrogation provides the scaffold for specificity in human flap endonuclease 1. *Cell Rep.* **3,** 1785–1794 (2013).

42. Getzoff, E. D. *et al.* Electrostatic recognition between superoxide and copper, zinc superoxide dismutase. *Nature* **306,** 287–290 (1983).

43. Tsutakawa, S. E. & Tainer, J. A. Double strand binding-single strand incision mechanism for human flap endonuclease: implications for the superfamily. *Mech. Ageing Dev.* **133,** 195–202 (2012).

44. Algasaier, S. I. *et al.* DNA and protein requirements for substrate conformational changes necessary for human flap endonuclease-1-catalyzed reaction. *J. Biol. Chem.* **291,** 8258–8268 (2016).

45. Reagan, M. S., Pittenger, C., Siede, W. & Friedberg, E. C. Characterization of a mutant strain of Saccharomyces cerevisiae with a deletion of the RAD27 gene, a structural homolog of the RAD2 nucleotide excision repair gene. *J. Bacteriol.* **177,** 364–371 (1995).

46. Callahan, J. L., Andrews, K. J., Zakian, V. A. & Freudenreich, C. H. Mutations in yeast replication proteins that increase CAG/CTG expansions also increase repeat fragility. *Mol. Cell. Biol.* **23,** 7849–7860 (2003).

47. Zhang, Y. *et al.* Genome-wide screen identifies pathways that govern GAA/TTC repeat fragility and expansions in dividing and nondividing yeast cells. *Mol. Cell* **48,** 254–265 (2012).

48. Singh, P., Zheng, L., Chavez, V., Qiu, J. & Shen, B. Concerted action of exonuclease and Gap-dependent endonuclease activities of FEN-1 contributes to the resolution of triplet repeat sequences (CTG)n- and (GAA)n-derived secondary structures formed during maturation of Okazaki fragments. *J. Biol. Chem.* **282,** 3465–3477 (2007).

49. Shah Kartik, A. *et al.* Role of DNA polymerases in repeat-mediated genome instability. *Cell Rep* **2,** 1088–1095 (2012).

50. Shishkin, A. A. *et al.* Large-scale expansions of Friedreich's ataxia GAA repeats in yeast. *Mol. Cell* **35,** 82–92 (2009).

51. Gordenin, D. A., Kunkel, T. A. & Resnick, M. A. Repeat expansion—all in a flap? *Nat. Genet.* **16,** 116–118 (1997).

52. Kim, J. C. & Mirkin, S. M. The balancing act of DNA repeat expansions. *Curr. Opin. Genet. Dev.* **23,** 280–288 (2013).

53. Tishkoff, D. X., Filosi, N., Gaida, G. M. & Kolodner, R. D. A novel mutation avoidance mechanism dependent on S. cerevisiae RAD27 is distinct from DNA mismatch repair. *Cell* **88,** 253–263 (1997).

54. Liu, B., Hu, J., Wang, J. & Kong, D. Direct Visualization of RNA-DNA Primer Removal from Okazaki Fragments Provides Support for Flap Cleavage and Exonucleolytic Pathways in Eukaryotic Cells. *J. Biol. Chem.* **292,** 4777–4788 (2017).

55. Pizzolato, J., Mukherjee, S., Schärer, O. D. & Jiricny, J. FANCD2-associated nuclease 1, but not exonuclease 1 or flap endonuclease 1, is able to unhook DNA interstrand cross-links *in vitro. J. Biol. Chem.* **290,** 22602–22611 (2015).

56. Sun, H. *et al.* The FEN1 L209P mutation interferes with long-patch base excision repair and induces cellular transformation. *Oncogene* **36,** 194–207 (2017).

57. Liu, S. *et al.* Okazaki fragment maturation involves alpha-segment error editing by the mammalian FEN1/MutSalpha functional complex. *EMBO J.* **34,** 1829–1843 (2015).

58. Ledford, H. End of cancer-genome project prompts rethink. *Nature* **517,** 128–129 (2015).

59. Tomasetti, C., Li, L. & Vogelstein, B. Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention. *Science* **355,** 1330–1334 (2017).

60. Kokoska, R. J. *et al.* Destabilization of yeast micro- and minisatellite DNA sequences by mutations affecting a nuclease involved in Okazaki fragment processing (rad27) and DNA polymerase delta (pol3-t). *Mol. Cell. Biol.* **18,** 2779–2788 (1998).

61. Potaman, V. N. *et al.* Length-dependent structure formation in Friedreich ataxia (GAA)n*(TTC)n repeats at neutral pH. *Nucleic Acids Res.* **32,** 1224–1231 (2004).

62. Krasilnikov, A. S. *et al.* Mechanisms of triplex-caused polymerization arrest. *Nucleic Acids Res.* **25,** 1339–1346 (1997).

63. Mirkin, S. M. Expandable DNA repeats and human disease. *Nature* **447,** 932–940 (2007).

64. Liu, Y. *et al.* Coordination between polymerase beta and FEN1 can modulate CAG repeat expansion. *J. Biol. Chem.* **284,** 28352–28366 (2009).

65. Morales, F. *et al.* Somatic instability of the expanded CTG triplet repeat in myotonic dystrophy type 1 is a heritable quantitative trait and modifier of disease severity. *Hum. Mol. Genet.* **21,** 3558–3567 (2012).

66. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).

67. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66,** 213–221 (2010).

68. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66,** 486–501 (2010).

69. Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).

70. Patel, N. *et al.* Proline scanning mutagenesis reveals a role for the flap endonuclease-1 helical cap in substrate unpairing. *J. Biol. Chem.* **288,** 34239–34248 (2013).

Author contributions

J.A.G., L.D.F., M.J.T., S.E.T., F.R., S.M.H., S.M.M. and J.A.T. designed the experiments. M.J.T., L.D.F., S.I.A., S.J.S., V.J.B.G., M.Z.H. and E.J. made mutant proteins and performed biochemical analyses. A.S.A., S.E.T. and J.A.T. did crystallographic analyses. J.C.K., A.J.N. and S.M.M. carried out the *in vivo* yeast studies. A.S.S. did western blot analysis.

Additional information

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Tsutakawa, S. E. *et al.* Phosphate steering by Flap Endonuclease 1 promotes 5′-flap specificity and incision to prevent genome instability. *Nat. Commun.* **8,** 15855 doi: 10.1038/ncomms15855 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.