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

Mechanistic studies of 5' nucleases: The FEN superfamily

Supervisor: Professor Jane Grasby

Steven J Shaw

March 2017

Declaration

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

Steven J Shaw

Acknowledgements

It comes as no surprise that a PhD is hard work, which takes plenty of time and effort to complete. However, it is an exhilarating experience to endeavour, although maybe more exhilarating after its completion, but it is an experience that cannot be done alone. Therefore, I would first like to acknowledge Professor Jane Grasby, without whose guidance and encouragement this project would not have proceeded. And to thank Dr David Finger, who has been paramount to this project since before I started, providing help where necessary and striving on into the late hours to tackle the problems of human exonuclease 1.

A large thank you to the departmental staff who ran the services necessary for this project. In particular Robb Hanson, for his help in olignonucleotide purification by HPLC, Simon Thorpe who collected any mass spectra taken and Nick Smith, who helped in various aspects from waste disposal to nitrogen canister changes. I would also like to thank Sam Fox and Karen Evans from Sciex separations for their help with and development of the capillary electrophoresis aspect of this work.

Other acknowledgements, of great importance, are to the people who kept me sane throughout this project. I am referring to the members of the Grasby group, past and present, for their entertainment and support. This includes “Dr” Ian Bennet, my Mexico companion, Dr Jack Exell, the American deserter, Rebecca Ley, the rugby enthusiast, Nur Nazihah Md Shahari, otherwise known as “Gorgeous Nazihah” to herself, and Sana Algasaier, the mothering soul who cleaned up after us. I would also like to thank Dr Mark Thompson, for his help. Finally, I would like to thank Dr Victoria Gotham, Dr Esther Allen and Alec Johns, for the time we worked together.

I would also like to thank my family and friends for all their support throughout the years, in particular my parents Barry and Carol Shaw. And last but not least, Michelle Vermeulen for her help and love in the hardest parts of my project. Thank you all.

Abstract

The flap endonuclease superfamily of 5'-nucleases have roles in DNA replication and repair. They react with DNA via a conserved nucleolytic core, co-ordinating with divalent metal ions to hydrolyse DNA in a structure-specific manner. Two examples are: flap endonuclease 1 (FEN1), which acts in Okazaki fragment maturation (OFM) during DNA replication, and exonuclease 1 (EXO1), which is important in DNA repair pathways. Although FEN1 has been extensively studied in the past, EXO1 has not been thoroughly examined. Therefore, this project focused on better understanding EXO1 by exploring commonalities and differences to FEN1.

EXO1 is primarily an exonuclease, removing nucleotides from dsDNA ends. This thesis describes biochemical analyses of the exonucleolytic capabilities of human EXO1 (hEXO1). Establishment of the enzymes substrate specificity, combined with determination of the kinetic parameters, allowed for characterisation of features of its reaction mechanism. In particular, experiments demonstrated that hEXO1 reactions were rate-limited by product release at high substrate concentrations. FEN1 processes endonucleolytic substrates by passing 5'-ssDNA flaps through a helical archway, which is conserved in EXO1. EXO1 also has suggested involvement in OFM; therefore, investigation of whether hEXO1 threads its endonucleolytic substrates was undertaken. Preventing or capturing the threaded state with a biotinylated substrate and streptavidin demonstrated that EXO1 must thread flapped substrates prior to catalysis. Further studies using multiple FEN1 mutants at residues expected to stabilise the threaded state, in combination with substrates with differing 5'-modifications, identified R104 and K132 as important residues for interaction with the +1 phosphate. Arch residue R129 was also identified as being required for efficient catalysis.

Biophysical analyses of EXO1 by CD determined that a signal change observed for hFEN1 with substrates containing tandem 2-aminopurines was not produced with hEXO1. However, the DNA base distortion hypothesised to cause the observed shift in FEN1 crystal structures is also present in EXO1 crystals. Finally, multiple N-hydroxyurea inhibitors known to inhibit hFEN1 were shown to be non-specific to FEN1 as biochemical and biophysical techniques demonstrated interaction with and inhibition of hEXO1.

Publications arising from work covered in this thesis

Algasaier SI, Exell JC, Bennet IA, Thompson MJ, Gotham VJB, **Shaw SJ**, Craggs TD, Finger LD & Grasby JA. 2016. DNA and protein requirements for substrate conformational changes necessary for human Flap Endonuclease-1-catalysed reactions. ***Journal of Biological Chemistry***, 291(15), 8258-8268.

Exell JC, Thompson MJ, Finger LD, **Shaw SJ**, Debreczeni J, Ward TA, McWhirter C, Siöberg CLB, Molina DM, Abbott WM, Jones CD, Nissink JWM, Durant ST & Grasby JA. 2016. Cellularly active *N*-hydroxyurea FEN1 inhibitors block substrate entry to the active site. ***Nature Chemical Biology***, 12, 815-821.

Tsutakawa SE, Thompson MJ, Arvai AS, Neil AJ, **Shaw SJ**, Algasaier SI, Kim JC, Finger LD, Jardine E, Gotham VJB, Sarker AS, Her MZ, Rashid F, Hamdan SM, Mirkin SM, Grasby JA & Tainer JA. 2017. Phosphate steering by Flap Endonuclease 1 promotes 5′-flap specificity and incision to prevent genome instability. ***Nature Communications***, 8, 15855.

Shaw SJ, Finger LD & Grasby JA. 2017. Human Exonuclease 1 threads 5'-flap substrates through its helical arch. ***Biochemistry,*** *56* (29), 3704–3707.

Contents

Declaration .... *i*
Acknowledgements .... *ii*
Abstract .... *iii*
Publications arising from work covered in this thesis .... *iv*
List of Abbreviations .... *viii*
  Rate constants xii
  Substrate Abbreviations xii
Chapter 1: Introduction *1*
  1.1 DNA and nucleases 2
  1.2 DNA and cell replication 3
    1.2.1 DNA replication 3
    1.2.2 Genetic recombination 4
  1.3 DNA repair mechanisms 7
    1.3.1 Mismatch repair 7
    1.3.2 Double strand break repair 10
    1.3.3 Base excision repair 12
    1.3.4 Nucleotide excision repair 18
  1.4 The flap endonuclease superfamily 20
    1.4.1 Family members 20
    1.4.2 The hydrophobic wedge 23
    1.4.3 Helical gateway and cap 24
    1.4.4 DNA sliding 25
    1.4.5 Less significant helical motifs 25
    1.4.6 Active site residues 25
    1.4.7 Enzyme specific regions 27
  1.5 Project aims 29
Chapter 2: Materials and Methods *31*
  2.1 Tables of buffers 32
    2.1.1 Media and buffers for expression and purification 32
    2.1.2 Reaction Buffers 33
  2.2 Purification of hEXO1-352 34
    2.2.1 Generation of competent cells by the Inoue method 34

    2.2.2 Cloning and expression of hEXO1-352 34
    2.2.3 Purification of hEXO1-352 35
    2.2.4 Optimised purification of hEXO1-352 36
  2.3 Oligonucleotide sequences and constructs used herein 38
    2.3.1 Oligonucleotide strand sequences 38
    2.3.2 DNA constructs 39
    2.3.3 Purification of oligonucleotides 43
  2.4 Multiple turnover kinetics 44
    2.4.1 Steady-state kinetics 44
    2.4.2 Steady state kinetics by capillary electrophoresis 45
    2.4.3 Determination of Michaelis-Menten Parameters 46
    2.4.4 Observation of reactions by denaturing polyacrylamide gel electrophoresis (PAGE) 47
  2.5 Single turnover kinetics 47
    2.5.1 Determination of k_{ST} for hEXO1-352 47
    2.5.2 Determination of threading by trapping and blocking with streptavidin 48
  2.6 Circular dichroism spectroscopy 49
  2.7 Förster resonance energy transfer and fluorescence anistropy 50
    2.7.1 Förster resonance energy transfer (FRET) 50
    2.7.2 Fluorescence anisotropy (FA) 51
  2.8 Inhibitor studies 52
    2.8.1 Multiple turnover studies 52
    2.8.2 Differential scanning fluorimetry (DSF) 53
Chapter 3: Studies into the Specificity and Kinetic Parameters of Exonuclease 1 as an Exonuclease . 54
  3.1 Evaluation of the substrate specificity of hEXO1 55
  3.2 Determination of the steady state parameters 58
  3.3 The mechanism of reaction and determination of the single turnover parameters 60
  3.4 The implications of processivity in the mechanism of hEXO1 63
  3.5 Summary of exonuclease studies 65
Chapter 4: Mechanistic Studies into how Flapped Substrates are Processed by EXO1 and FEN1 66
  4.1 Studies of the endonuclease capability of hEXO1 67
  4.2 Determination of hEXO1's requirement to thread flapped substrates 70
  4.3 The importance of conserved basic residues in +1 phosphate interaction in hFEN1 80
  4.4 Summary of endonuclease studies 91

Chapter 5: The Role of Substrate Dynamics in Catalysis 92
5.1 Enzyme-induced conformational change 93
5.2 The limitations of substrate binding techniques with hEXO1 103
5.3 Summary of substrate dynamics 108
Chapter 6: Determination of the specificity of a range of N-hydroxyurea inhibitors for hFEN1 110
6.1 Identification of a group of N-hydroxyurea inhibitors for hFEN1 111
6.2 Determination of the compounds specificity for hFEN1 112
6.3 The effects of PCNA on FEN1 inhibition 115
6.4 Summary of the inhibition experiments 118
Chapter 7: Conclusions 120
7.1 The conclusions of this research 121
7.2 Future work 125
7.3 Implications of this research on the flap endonuclease superfamily 125
References 127
Appendices 135
Publications 139

List of Abbreviations

Abbreviation	**Term**
2-AP	2-Aminopurine
OVH	3' overhang
8-oxoG	8-oxoguanine
AP	Abasic
APE1	Abasic endonuclease 1
AOL	Acceptor only labelled
AID	Activation-induced cytidine deaminase
APS	Ammonium persulphate
$(NH_4)_2SO_4$	Ammonium sulphate
Amp	Ampicillin
AIM	Autoinduction media
BER	Base excision repair
BLM	Bloom syndrome RecQ helicase
BD	Blunt duplex
BSA	Bovine serum albumin
$CaCl_2$	Calcium chloride
CE	Capillary electrophoresis
CETSA	Cellular thermal shift assay
Cm	Chloramphenicol
CD	Circular dichroism
CV's	Column volumes
C	Competitor
DNA	Deoxyribonucleic acid
DSF	Differential scanning fluorimetry
DMSO	Dimethyl sulphoxide
Na_2HPO_4	Disodium hydrogen phosphate
DTT	Dithiothreitol
DOL	Donor only labelled
DAL	Donor-acceptor labelled

Abbreviation	**Term**
DF	Double flapped duplex
DNU	Double nucleotide unpairing
DSB	Double strand break
DSBR	Double strand break repair
Endo	Endonucleolytic
E	Enzyme, macromolecule
EQ	Enzyme-duplex product
ES	Enzyme-substrate complex
ES'	Enzyme-substrate intermediate complex
ESAS	Enzyme-substrate-streptavidin blocked complex
ESSA	Enzyme-substrate-streptavidin trapped complex
EDTA	Ethylenediaminetetraacetic acid
ECCD	Exciton-coupled circular dichroism
EXO1	Exonuclease 1
Exo	Exonucleolytic
XRN	Exoribonuclease
FF	Fast flow
FEN1	Flap endonuclease 1
FEN supefamily	Flap endonuclease superfamily
FAM	Fluorescein
FA	Fluorescence anisotropy
FB	Folding buffer
FRET	Förster Resonance Energy Transfer
GEN1	Gap endonuclease 1
H2TH	Helix two turn helix
HPLC	High pressure liquid chromatography
His	Histidine
HJ	Holliday junction

Abbreviation	**Term**
HR	Homologous recombination
HEPES	Hydroxyethylpiperazineethane sulphonic acid
IMAC	Immobilised metal ion affinity chromatography
IDL's	Insertion-deletion loops
ID	Internal diameter
LIF	Laser induced fluorescence
L_D	Length to detection window
LB	Luria bertani
$MgCl_2$	Magnesium chloride
Mg^{2+}	Magnesium ions
$MgSO_4$	Magnesium sulphate
mRNA	Messenger Ribonucleic acid
MMR	Mismatch repair
MRN	MRE11-Rad50-NBS1
TEMED	N-, N-, N’-, N’- tetramethylethylenediamine
ND	Nicked duplex
NL	Non labelled
NHEJ	Non-homologous end joining
NER	Nucleotide excision repair
PBS	Phosphate buffer saline
Pol's	Polymerases
KCl	Potassium chloride
KH_2PO_4	Potassium dihydrogen phosphate
P, Q	Product
PCNA	Proliferating cell nuclear antigen
PY	Pseudo-Y
QF	Quenchflow
RB	Reaction buffer

Abbreviation	**Term**
ROS	Reactive oxygen species
RRB	Reduced reaction buffer
RPC	Replication factor C
RPA	Replication protein A
SF	Single 3' flapped duplex
ST	Single turnover
sn-BER	Single-nucleotide base excision repair
NaN_3	Sodium azide
NaCl	Sodium chloride
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
NaOH	Sodium hydroxide
SA	Streptavidin
S	Substrate
SSA	Substrate-streptavidin complex
TB	Terrific broth
TAMRA	Tetramethylrhodamine
tRNA	Tranfer Ribonucleic acid
TNR's	Trinucleotide repeats
Tris	Trisaminomethane
TBE	Tris-Borate-EDTA
UV	Ultraviolet radiation
XP *"N"*	Xeroderma pigmentosum type "N" (N = A, B, C, D, E, F or G)
βME	β-mercaptoethanol

Rate constants

Constant	**Term**
K_{bend}	Substrate bending constant
k_{cat}	Maximal rate constant
k_{CC}	Conformational change rate
k_{chem}	Chemistry rate
K_D	Dissociation constant
K_M	Michaelis-Menten constant
k_{off}	Off rate
k_{on}	On rate
k_{PA}	Product association rate
$k_{release}$	Product release rate
k_{ST}	Single turnover rate constant
k_{UCC}	Reverse conformational change rate

Substrate Abbreviations

Full substrate details available in *section 2.3*.

Substrate	**Meaning**
OVH1	3' Overhang: 5' FAM labelled
OVH2	3' Overhang: 3' FAM labelled, 5' phosphate
OVH3	3' Overhang: 5' phosphate, competitor
OVH_{+1-1}	3' Overhang: +1, -1, 2-AP
OVH_{-1-2}	3' Overhang: -1, -2, 2-AP
UOVH_{+1-1}	Unimolecular 3' overhang: +1, -1, 2-AP
UOVH_{-1-2}	Unimolecular 3' overhang: -1, -2, 2-AP
BD1	Blunt duplex: 5' FAM labelled
BD2	Blunt duplex: 3' FAM labelled, 5' phosphate
DF1	Double flapped duplex: 5' FAM labelled
DF2	Double flapped extended duplex: 5' FAM labelled
DF_{+1-1}	Double flapped duplex: +1, -1, 2-AP
DF_{-1-2}	Double flapped duplex: -1, -2, 2-AP

Substrate	**Meaning**
DF_{AOL}	Double flapped duplex: Acceptor-only labelled
DF_{DAL}	Double flapped duplex: Donor-acceptor labelled
DF_{DOL}	Double flapped duplex: Donor-only labelled
DF_{NL}	Double flapped duplex: Non-labelled
ND1	Nicked duplex: 5' FAM labelled,
ND2	Nicked duplex: 3' FAM labelled, 5' phosphate
ND3	Nicked duplex: 3' FAM labelled, 5' hydroxyl
ND_{+1-1}	Nicked duplex: +1, -1, 2-AP
ND_{-1-2}	Nicked duplex: -1, -2, 2-AP
ND_{AOL}	Nicked duplex: Acceptor-only labelled
ND_{DAL}	Nicked duplex: Donor-acceptor labelled
ND_{DOL}	Nicked duplex: Donor-only labelled
ND_{NL}	Nicked duplex: Non-labelled
PY1	Pseudo-Y: 5' FAM labelled
PY2	Pseudo-Y: 3' FAM labelled
PY3	Pseudo-Y: 5' FAM labelled, 5' biotinylated
PY_{+1-1}	Pseudo-Y: +1, -1, 2-AP
PY_{-1-2}	Pseudo-Y: -1, -2, 2-AP
SF1	Single 3' flap duplex: 5' FAM labelled
SF2	Single 3' flap duplex: 3' FAM labelled, 5' phosphate
SF3	Single 3' flap duplex: 3' FAM labelled, 5' hydroxyl
SF_{+1-1}	Single 3' flap duplex: +1, -1, 2-AP
SF_{-1-2}	Single 3' flap duplex: -1, -2, 2-AP

Chapter 1: Introduction

1.1 DNA and nucleases

1.2 DNA and cell replication

1.3 DNA repair mechanisms

1.4 The flap endonuclease superfamily

Chapter 1: Introduction

1.1 DNA and nucleases

Deoxyribonucleic acid (DNA) has been well established as the genetic material for decades [1-2]. Important information is stored as DNA which is transcribed into messenger ribonucleic acids (mRNA's) before being translated into proteins via codon matching with transfer RNA (tRNA) at the ribosome [3]. The polymeric structure of DNA is composed of monomeric nucleotides; nucleotides consist of a negatively charged monophosphate connected to a deoxyribose sugar [4]. Each nucleotide contains one of four nucleobases; adenine, guanine, cytosine and thymine, connected to the deoxyribose at carbon-1; three nucleotides code for an amino acid. Nucleotides form a polymeric structure which is the basis of DNA; each phosphate forms a phosphodiester bond with the hydroxyl group at carbon-3 of the next nucleotides deoxyribose. The phosphodiester bond is very stable, but it can be broken by DNA hydrolysis (*Figure 1.1*). Natural decay of this bond is very slow under biological conditions, with a half-life of approximately 30 million years. However, enzymes known as nucleases catalyse DNA phosphodiester hydrolysis on the biologically relevant time scale of milliseconds [5].

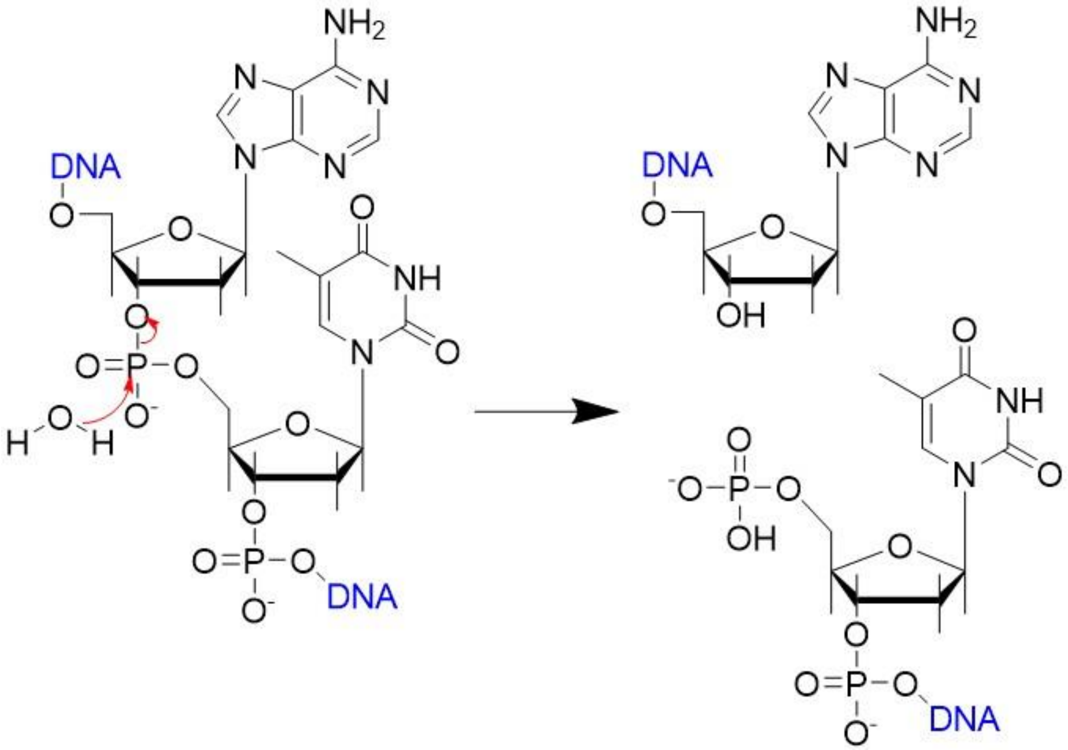


Figure 1.1: DNA hydrolysis. *The phosphodiester bond is broken via a pentavalent intermediate or transition state with a water molecule acting as the nucelophile.*

The capability of nucleases to hydrolyse DNA is utilised for various mechanisms in the cell. They have roles in DNA replication, repair mechanisms and DNA metabolism. Nucleases have two modes of specificity. Sequence-specific nucleases [6], such as restriction endonucleases, target a certain sequence within the DNA, are thought to have evolved as a defence to bacteriophages and spread by horizontal gene transfer [7]. Structure-specific nucleases, as the name would suggest, target a specific structure in the DNA, such as a double flap junction [8]. This thesis will focus on structure-specific nucleases.

1.2 DNA and cell replication

1.2.1 DNA replication

DNA replication is a highly regulated process that generates DNA copies precisely and efficiently. For example, the human genome consists of roughly 3 billion pairs of either GC and AT nucleotide pairs which encode genes, promoters, and various other components. The replication of the human genome is achieved by DNA polymerases (Pols), which chemically link nucleotides together. DNA elongation is a bidirectional process which is efficient at synthesising DNA in a 5' to 3' polarity, producing a long continuous strand known in the case of the leading strand. On the other hand, DNA synthesis of the opposing strand is more difficult as the strand only becomes accessible as the replication fork is unwound by a DNA helicase. Replication is still a 5' to 3' process, but it must be constructed in fragments as the DNA becomes accessible. In eukaryotes, primase and Pol α introduce primers which are extended by Pol δ and proliferating cell nuclear antigen (PCNA), generating ~50 million non-continuous strands otherwise known as Okazaki fragments [9-11]. This process is also known as lagging strand synthesis.

The lagging strand constructed from these Okazaki fragments in the process of Okazaki fragment maturation [12-13]. Elongation of the RNA primers generates overlapping sequences between the fragments which form “flapped” structures that must be removed before ligation by DNA ligase. Early work with cell extract isolated maturation factor 1 which was identified as being necessary for this process [14]. This factor was later aptly renamed flap endonuclease 1 (FEN1), a structure-specific nuclease responsible for the removal of these flaps, with a substrate requirement for a single nucleotide 3'-flap [12]. It has been suggested that RNase H cleaves the primer to a single ribonucleotide, and then FEN1 displaces the remainder of the primer into a 5'-flap and cuts one nucleotide into the duplex producing a nicked duplex which can be ligated forming a continuous sequence [13, 15-16].

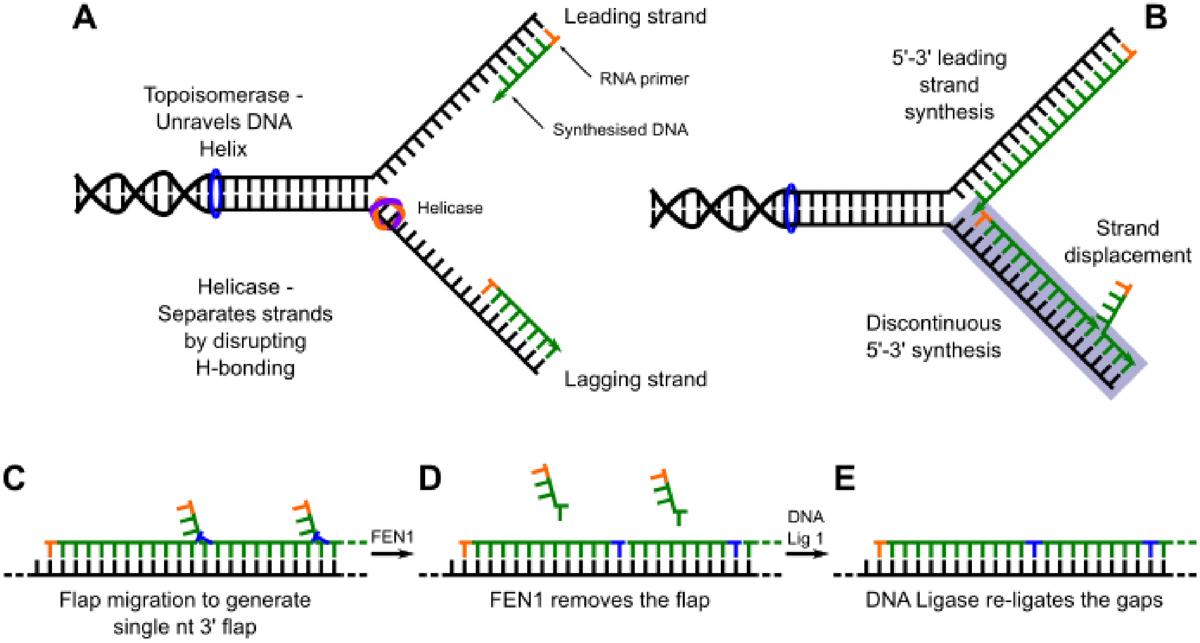


Figure 1.2: DNA synthesis and Okazaki Fragment Maturation. *DNA synthesis has to occur in two directions; continuous 5' to 3' leading strand synthesis and discontinuous 5' to 3' lagging strand synthesis.* ***A.*** *A topoisomerase unravels the DNA helix and a helicase melts the DNA strands to allow for synthesis of both parent strands.* ***B***. *The strands are elongated 5' to 3' by DNA polymerase δ from an RNA primer, however lagging strand synthesis can only occur in short Okazaki fragments that require combination.* ***C.*** *A flap is displaced by strand displacement synthesis which requires flap migration to form the optimal substrate of FEN1.* ***D & E***. *After the removal of the flap by FEN1 one nucleotide into the duplex the major product only requires ligation by DNA ligase.*

1.2.2 Genetic recombination

Cell replication requires equal division of DNA that is replicated during S-phase (as above); mitosis allows for the formation of two daughter cells from an initial mother cell, which are clones of the original [17]. Meiosis is a process similar to mitosis; the genetic material is also split evenly between cells but eventually, after two rounds, provides sex cells with one full set of chromosomes (half that of a somatic cell) [18]. Meiosis I separates chromosome pairs that are aligned along the spindle fibres before cell division, this produces cells with one set of chromosomes but double the DNA content (2n). Meiosis II is a second division event which provides cells with a single set of chromosomes, which can develop into sperm or eggs.

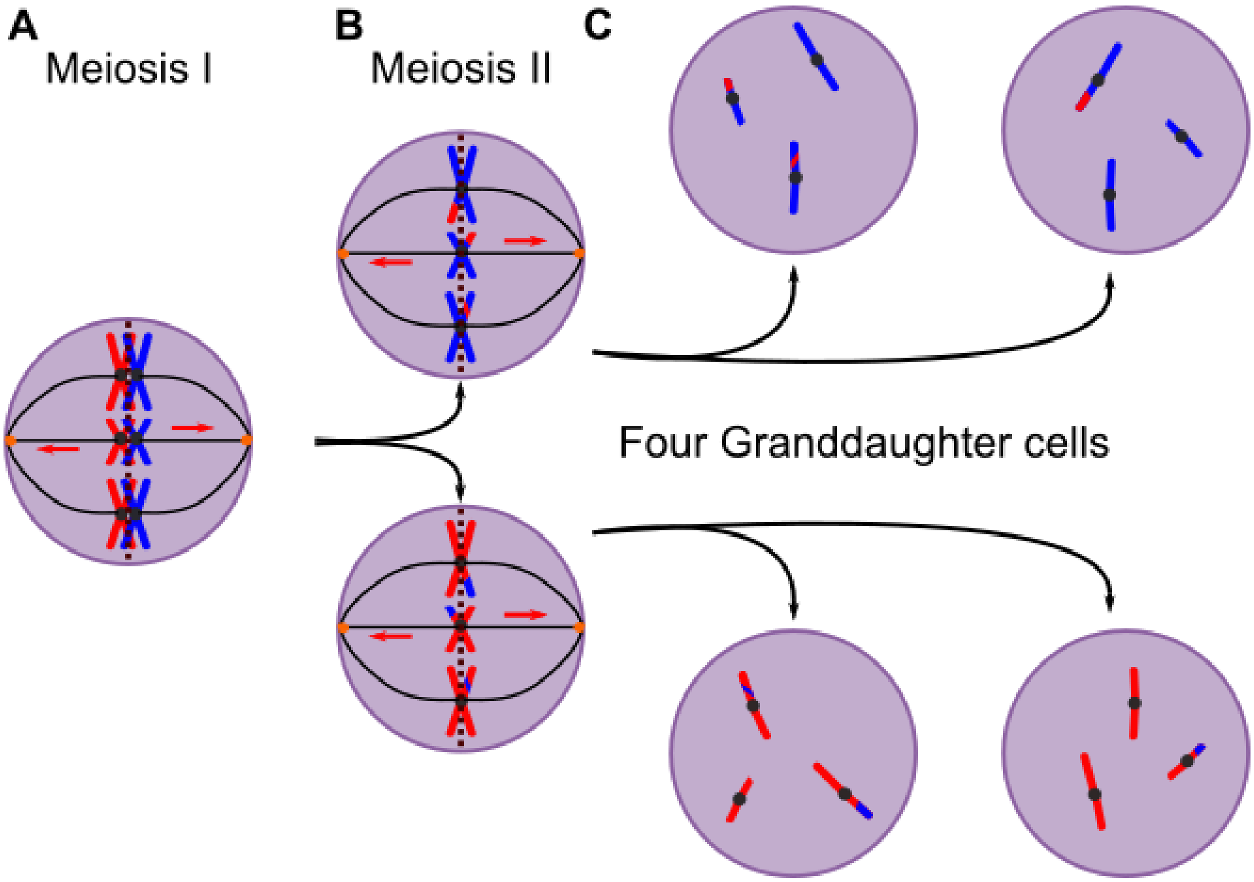


Figure 1.3: DNA Meiosis. A. *Meiosis I begins with chromosome pairs with 2n DNA lining up along the centre attached to the spindle by the centromere. The spindle fibres retract, separating chromosome crossovers which can result in DNA sharing between the pairs or dissolution.* ***B.*** *Meiosis II occurs in the same way, but with single chromosomes which have 2n DNA. This process separates the chromatids to create four granddaughter cells from the original progenitor cell.* ***C.*** *Granddaughter cells are produced which have half the genetic material of a normal cell, these can later become sex cells. DNA can be shared between chromosomes by this process.*

During meiosis I, after alignment of chromosome pairs, each chromosome can have a single break introduced into their sequence. This leaves two free ends which allow the chromosomes to strand invade each other, leading to a cross over event generating a single Holliday junction (HJ) (figure 1.4), which can branch migrate along the DNA [19]. There are two modes of action for repairing this junction which are dependent of the orientation of the crossover. Horizontal resolution (dissolution) unwinds the chromatids and allows for correct separation of chromosomes, therefore, no recombination of the two strands takes place [20]. Alternatively, if the strands are resolved vertically DNA is exchanged between the chromosomes [21-22]. This process can be performed by process can be performed by gap endonuclease 1 (GEN1). GEN1 acts as a dimer cutting opposing strands 5'-3' one nucleotide into the duplex, which is the same approach FEN1 utilises on 5'-flapped substrates.

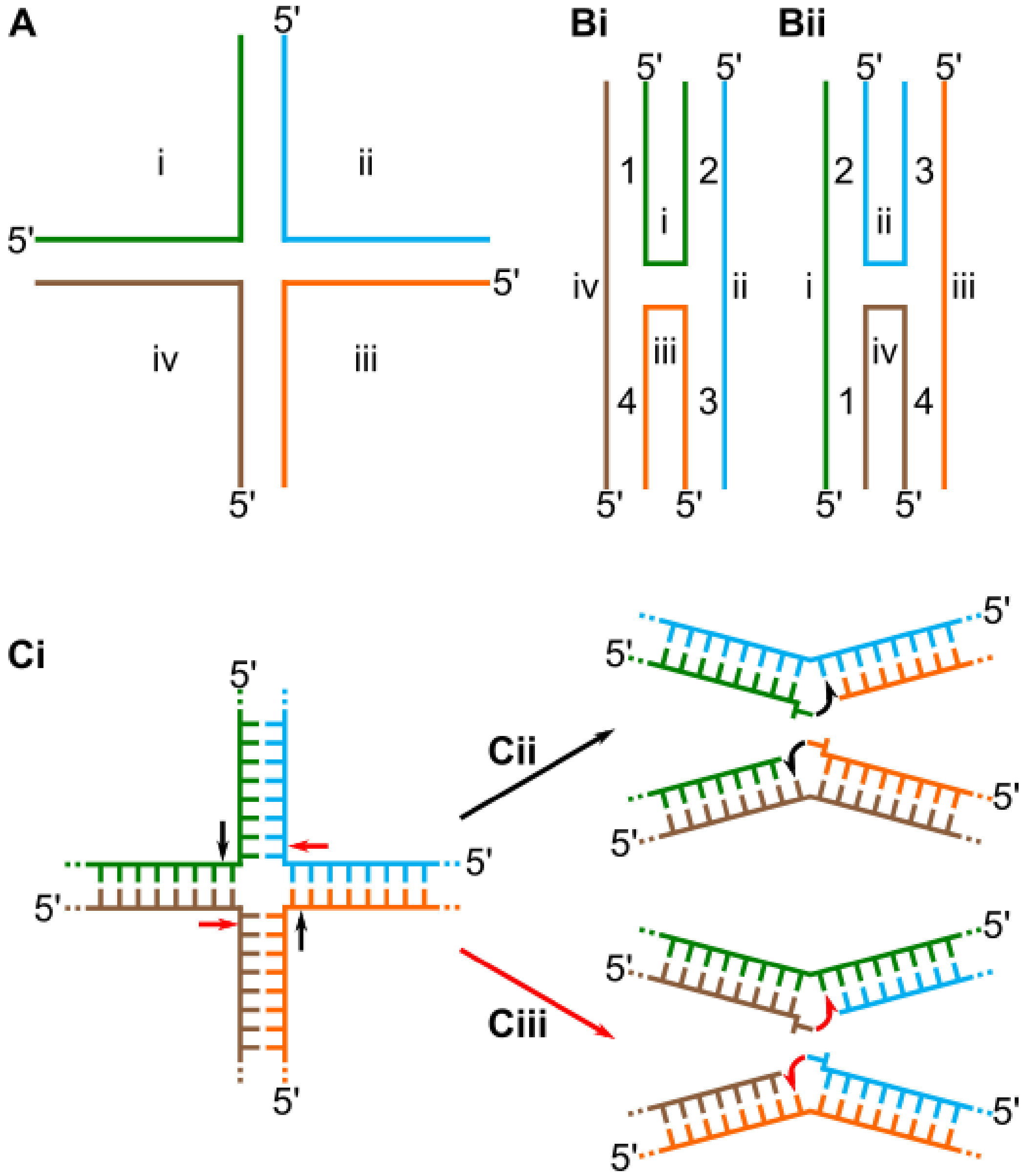


Figure 1.4: Chromosomal crossovers and recombination. A. *A Holliday junction (HJ); each colour represents a separate DNA strand (i-iv), with strand polarity highlighted.****Bi-ii.*** *The two*

conformations the chromatids can form. The strand orientation determines where GEN1 cuts the junction, and impacts which strands are shared. 1-4 highlights the helices formed, to clarify their locations. ***Ci.*** *A HJ showing the cleavage points for GEN1, which acts as a dimer. The arrows highlight the cleavage sites for* ***Cii*** *and* ***Ciii*** *with black and red respectively. The DNA duplexes formed,* ***Cii-iii****, are repaired by DNA ligase. This process leads to successful and efficient resolution of chromatid crossovers.*

Processing HJs by resolution over dissolution is cell cycle dependent. Dissolution occurs during G1, S and G2 phase, whereas resolution takes place primarily in mitosis or meiosis. It is unclear as to why this is observed, but it may provide greater genetic diversity in sex cells [21, 23]. As illustrated above in *figure 1.3*, granddaughter cells consist of one set of chromatids which can have DNA sharing between chromosomes. Whatever the cause, GEN1 is an efficient endonuclease which processes HJs in a similar manner to the way FEN1 processes Okazaki fragments. Another method of chromosomal crossovers is due to double strand break repair (DSBR) which is mentioned in the next section (*section 1.3.2*). The final phase relies on DNA repair by homologous recombination via strand invasion [24]. In DSBR both strands require repair, which leads to the formation of two HJ's, however, they are both treated in the same was as single HJ's, like above [19].

1.3 DNA repair mechanisms

1.3.1 Mismatch repair

As outlined in *section 1.2.1*, eukaryotic DNA replication utilises three major DNA polymerases (Pols) in leading and lagging strand synthesis. Pol α lacks proofreading capabilities, meaning it cannot remove incorrectly incorporated nucleotides by 3' to 5' exonuclease activity and can only process DNA 5' to 3', whereas the other major replicative DNA polymerases (Pol ε and δ) contain proofreading functionality [9]. The ability to remove replicative errors increases fidelity of Pols by 100-fold from 1 mismatch in 10^6 to 1 in 10^8 basepairs [25]. However, the human genome is approximately 3 billion base pairs ($3x10^9$), making it prone to mistakes. DNA mismatch repair (MMR) has evolved to remove any mistakes that are not removed by the Pols in tandem with DNA replication. Defects in MMR can lead to Lynch syndrome, otherwise known as Heredity Nonpolyposis Colorectal Cancer, leaving sufferers susceptible to certain cancers [26-27].

The eukaryotic MMR pathway allows for the hydrolysis of long tracts of DNA to remove any DNA mismatches inserted. Exonuclease 1 (EXO1) was first identified in *Schizosaccharomyces pombe,* and plays a central role in MMR as the primary nuclease, before efficient repair by DNA synthesis with Pol δ and DNA ligase. When a mismatch is inserted into DNA it is first detected by the MutSα complex, a heterodimer of MSH2 and MSH6 [28-30], which clamps around the DNA at the mismatched base pair. This is achieved by a loop in MSH6 (Phe-X-Glu) which is conserved in eukaryotes and is even present in the bacterial mutS protein [31]. MutSα recognises mismatches primarily, but it can also recognise small insertion-deletion loops (IDLs). Upon recognition of a mismatch the protein clamps the DNA by ATP hydrolysis. When the protein is clamped it is free to slide along the DNA in the presence of ATP [32]. Other MutS complexes exist, but 80-90% of MSH2 is present in the MutSα complex [33]. The MutSα heterodimer then forms a larger complex with MutLα, which is a heterodimer of MLH1 and PMS2, in an ATP-dependent interaction. Like MutS complexes, there are multiple MutL heterodimers, but 90% of MLH1 is in the MutLα form [34-35]. The MutSα-MutLα complex can slide along the DNA contour and MutLα is thought to discriminate between a mismatch and a canonical base pair [36]. Together MutSα-MutLα recognise whether a mismatch is present, allowing for stage two of the repair process.

PCNA is loaded onto the DNA by replication factor C (RFC) which interacts with the MutSα-MutLα complex via MSH6 [37-40]. Early studies into mismatch repair suggested that a nick could be introduced 3' and 5' of the mismatch and the error would still be removed. However, EXO1 can only perform exonuclease activity 5' to 3', therefore, a nick 3' of the mismatch would not be effective for EXO1 utilisation [41]. It was later discovered that MutLα introduces a nick into the DNA 5' of the mismatched base pair via discrete endonuclease activity of its PMS2 domain [42]. This process was shown to be PCNA-dependent by p21 inhibition; depletion of PCNA by direct binding with p21 showed a 50% reduction of 5'-directed nicks and abolished 3'-directed nick activity [43-45].

When a nick is introduced upstream of the mismatch EXO1 can remove a tract of DNA, including the mismatched base. EXO1 acting with MutSα removes ~2000 nucleotides, but when replication protein A (RPA) is also present, DNA removal is reduced to ~250 nucleotides and nucleolytic activity is terminated after the mismatch is removed [41]. After the release of EXO1, the DNA can be repaired with a combination of DNA elongation by Pol δ and ligation by DNA ligase (*figure 1.5*). Though EXO1 is the most integral nuclease for MMR, it has been shown that other nucleases can be used as replacements [46]. FEN1 is also thought to act as a back-up for EXO1 by strand displacement, treating the mismatched strand as a large flap. These processes allow for genome conservation during replication should the proofreading ability of Pol δ and ε fail.

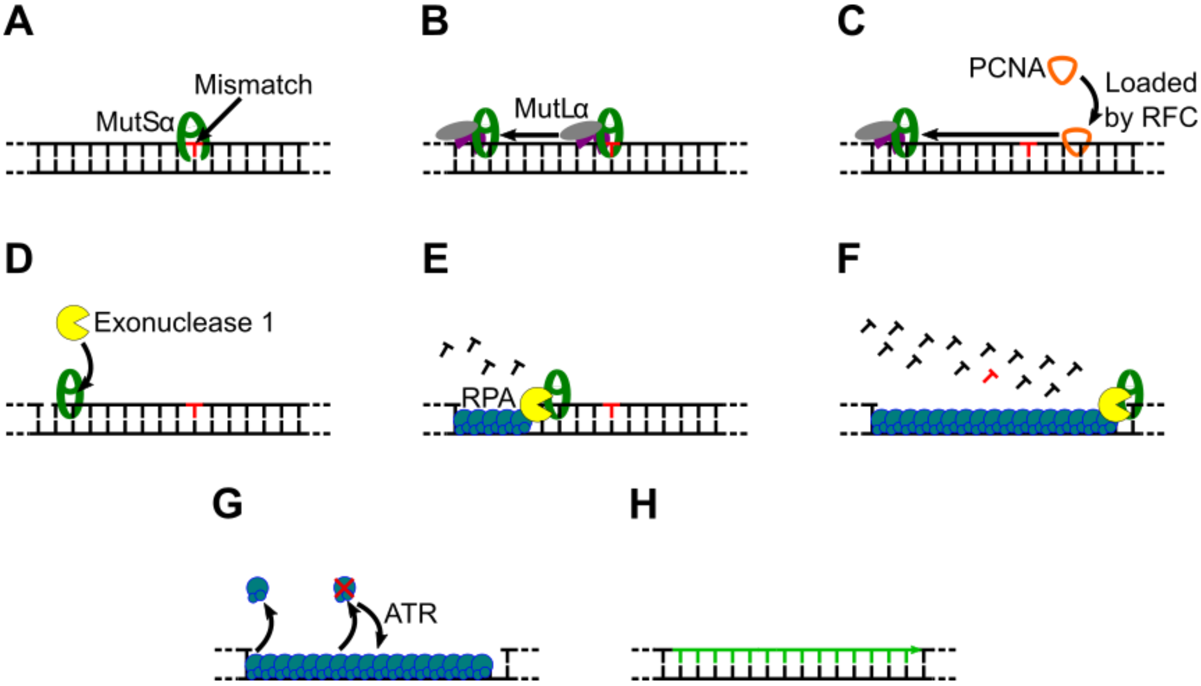


Figure 1.5: Eukaryotic DNA Mismatch Repair. *A mismatch is introduced into the DNA, highlighted by the red nucleotide.* ***A.*** *MutSα first determines if a mismatch is present before clamping the DNA.* ***B.*** *Upon recognition by MutSα it interacts with MutLα and slides along the DNA. MutLα is also predicted to help in mismatch discrimination.* ***C.*** *PCNA is loaded onto the DNA by RFC, allowing for a nick to be introduced by the discrete endonuclease of the PMS2 subunit of MutLα.* ***D.*** *MutSα recruits EXO1 to a nick 5' of the mismatch.* ***E.*** *EXO1, in combination with MutSα, resects around 250 nucleotides, removing the mismatch. RPA binds to the ssDNA.* ***F.*** *Shortly after the removal of the mismatch, MutSα promotes EXO1 to stop resecting DNA.* ***G.*** *RPA can either release from the DNA or activate the DNA damage checkpoint via ATR.* ***H.*** *DNA polymerase δ resynthesises the DNA before DNA ligase fixes the ssDNA gap generated.*

1.3.2 Double strand break repair

Another form of DNA damage considered the most cytotoxic is breaks in the DNA [47]. Single strand breaks are relatively easy to rectify; however, formation of double strand breaks (DSBs) can be difficult to repair, relying on various pathways. The pathway used to repair a DSB is dependent on the phase of the cell cycle; for example, non-homologous end joining (NHEJ) takes place primarily in G1 phase. Therefore, DSB repair (DSBR) is a vastly regulated process consisting of two major pathways. The first mentioned above, NHEJ, directly ligates the DNA ends back together, but can only perform this role on free unmodified (e.g. not oxidised) ends [48]. The second method is homologous recombination (HR), which relies on first resecting the DNA at the free 5'-ends generating overhangs [24]. These overhangs are then repaired by DNA synthesis with DNA polymerase using the sister chromatids as templates for homologous recombination before ligation by DNA ligase. This leaves a repaired DNA junction as its product. A third pathway exists known as microhomology-mediated end joining (MMEJ), which uses aspects of the other two pathways and is active throughout the cell cycle.

A heterodimer of Ku70 and Ku80 form the Ku complex which can determine pathway choice in a cell cycle-dependent manner [49]. If the cell is in G1 phase, Ku will interact with the DSB free ends, preventing DNA resection as the 5'-ends are not accessible by the HR nucleases (*figure 1.6A*) [50]. Therefore, Ku directly inhibits HR and promotes NHEJ. Conversely, when the cell cycle is in G2 or S phase, resection removes long tracts of DNA. The resulting duplex has no free double-stranded ends for Ku to bind, inhibiting NHEJ (*figure 1.6B*). Although both pathways are important, focus will be on HR as EXO1 is central to extensive resection.

In mammalian HR, CtIP and MRN (MRE11-Rad50-NBS1), initiate the process by binding at the DSB; this is a required step for HR [51-52]. MRN recruits the two nucleases, EXO1 and DNA2, which have slightly different mechanisms of resection (*figure 1.6B*). For DNA2 resection, BLM (Bloom syndrome RecQ helicase) interacts with DNA2 and resects in an ATP-dependent manner [24]. CtIP recruits RPA which in turn promotes BLM unwinding of the DNA enforcing a 5' to 3' polarity [53]; MRN accelerates this process by promoting BLM interaction with the DNA ends [24]. The EXO1 resection mechanism starts with BLM increasing EXO1’s affinity for the DNA ends before MRN stimulates and enhances EXO1’s processivity [24, 54]. Again, RPA also helps stimulate this process via BLM, maintaining exonuclease activity at the 5'-end. Due to its role in both the DNA2 and EXO1 resection methods, BLM is considered to be a dominant protein in DNA end resection [24, 54]. However, it has also been shown to have an inhibitory function in homologous recombination by preventing Rad51 interaction with the ssDNA [55]. Homologous recombination forms two HJ's via strand invasion of sister chromatids, which is repaired as described in *section 1.2.2*.

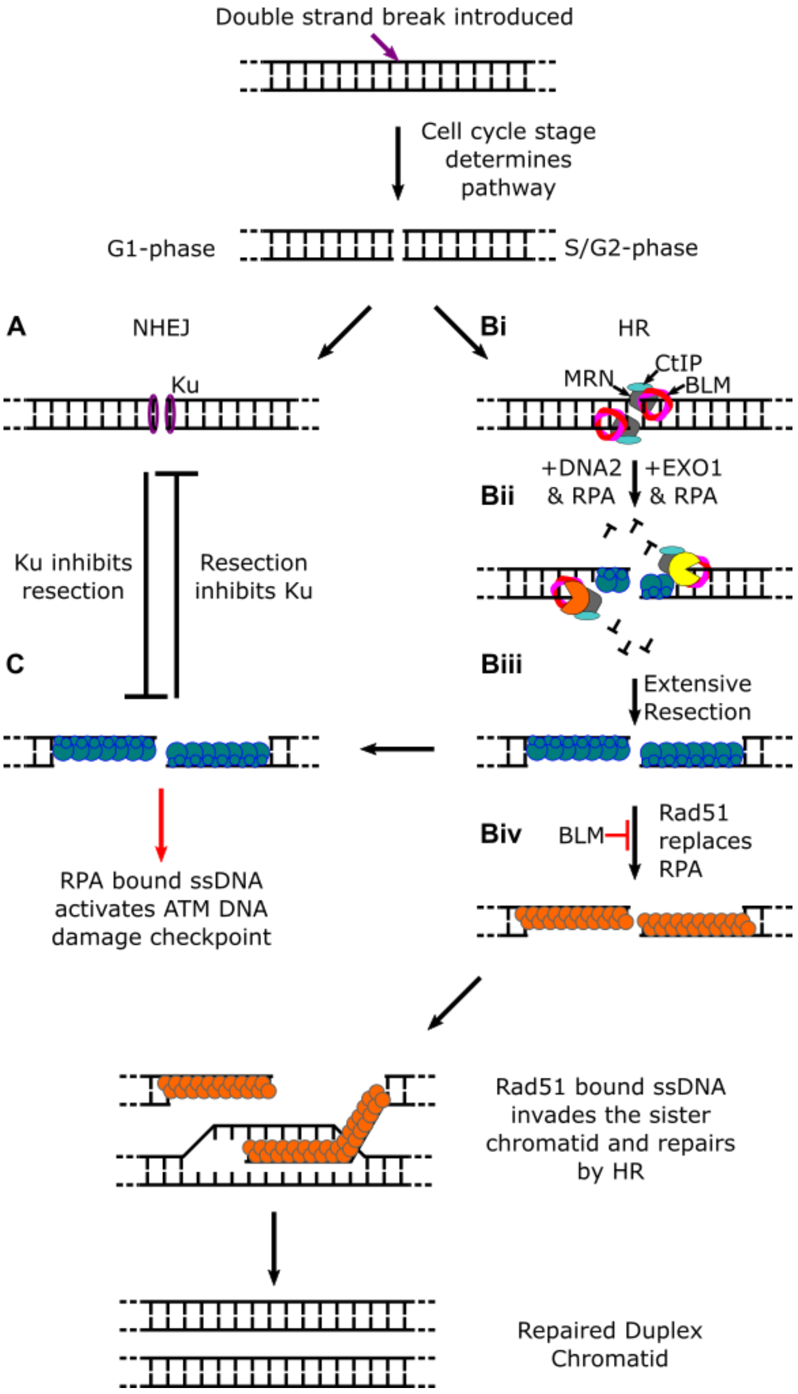


Figure 1.6: Mammalian Double Strand Break Repair. *A double strand break is introduced by endogenous (e.g. DNA polymerase mistakes) or exogenous (e.g. ionising radiation) means. The cell cycle stage determines the pathway choice.* ***A.*** *For cells in G1-phase the DSB is repaired by non-homologous end joining (NHEJ). The Ku heterodimer binds to the DNA ends, inhibiting resection and ensuring NHEJ.* ***Bi.*** *If the cell is in G2 or late S-phase the DSB is processed by homologous recombination (HR). CtIP and MRN interact with the ends and start the process with the help of the BLM helicase. MRN recruits nucleases to the DNA ends.* ***Bii.*** *There are two major nucleases that can process the break, DNA2 and EXO1, with the aid of BLM and RPA for polarity discrimination.* ***Biii.*** *Once extensive resection has occurred, the product duplex has large overhangs coated with RPA. This can trigger two processes:* ***Biv.*** *The RPA can be replaced with Rad51 which encourages HR by strand invasion of the sister chromatids.* ***C.*** *Alternatively the RPA-coated DNA can activate the DNA damage checkpoint by ATR.*

1.3.3 Base excision repair

Oxidation is another form of DNA damage, which can result in oxidised nucleobases. Reactive oxygen species (ROS) generated either endogenously by leakage from the electron transport chain or exogenously by environmental oxidative stressors can lead to oxidation [56-57]. ROS can generate oxidative DNA base lesions which has the potential to cause mutations, abnormal gene transcription or epigenetic instability, if left in disrepair [56]. An example of an oxidised base is 8-oxoguanine (8-oxoG) which forms a non-Watson-Crick base pair with adenine as well as its canonical cytosine pairing [58-60].

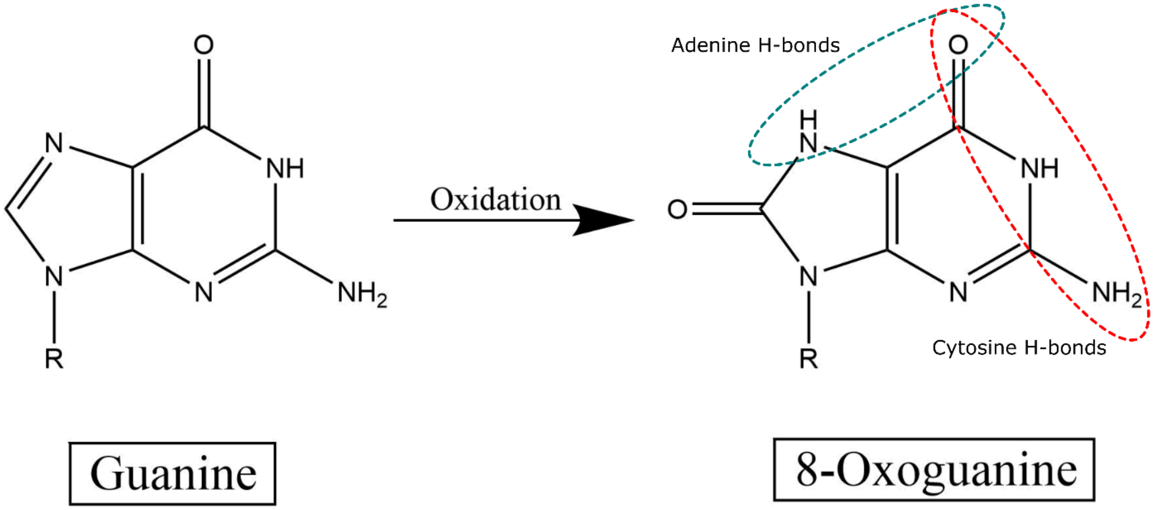


Figure 1.7: *Guanine is oxidised to 8-oxoguanine by exogenous and endogenous sources allowing it to form hydrogen bonds with adenine as well as cytosine.*

Repairing this lesion can be achieved by multiple methods, but it always begins with the removal of the oxidised base. This is achieved by glycosylases; in the case of 8-oxoG the aptly named 8-oxoguanine DNA glycosylase is utilised to remove the base, generating an abasic (AP) site [59]. AP endonuclease 1 (APE1) generates a nick at the AP site allowing for one of three processes to occur [61-62]. For a native deoxyribose sugar which has not been oxidised the process requires a combination dRP lyase to remove the 5' deoxyribose sugar and Pol β to fill the gap [63]. The two strands are then repaired by DNA ligase. This process is known as single-nucleotide base excision repair (sn-BER)(*figure 1.8*).

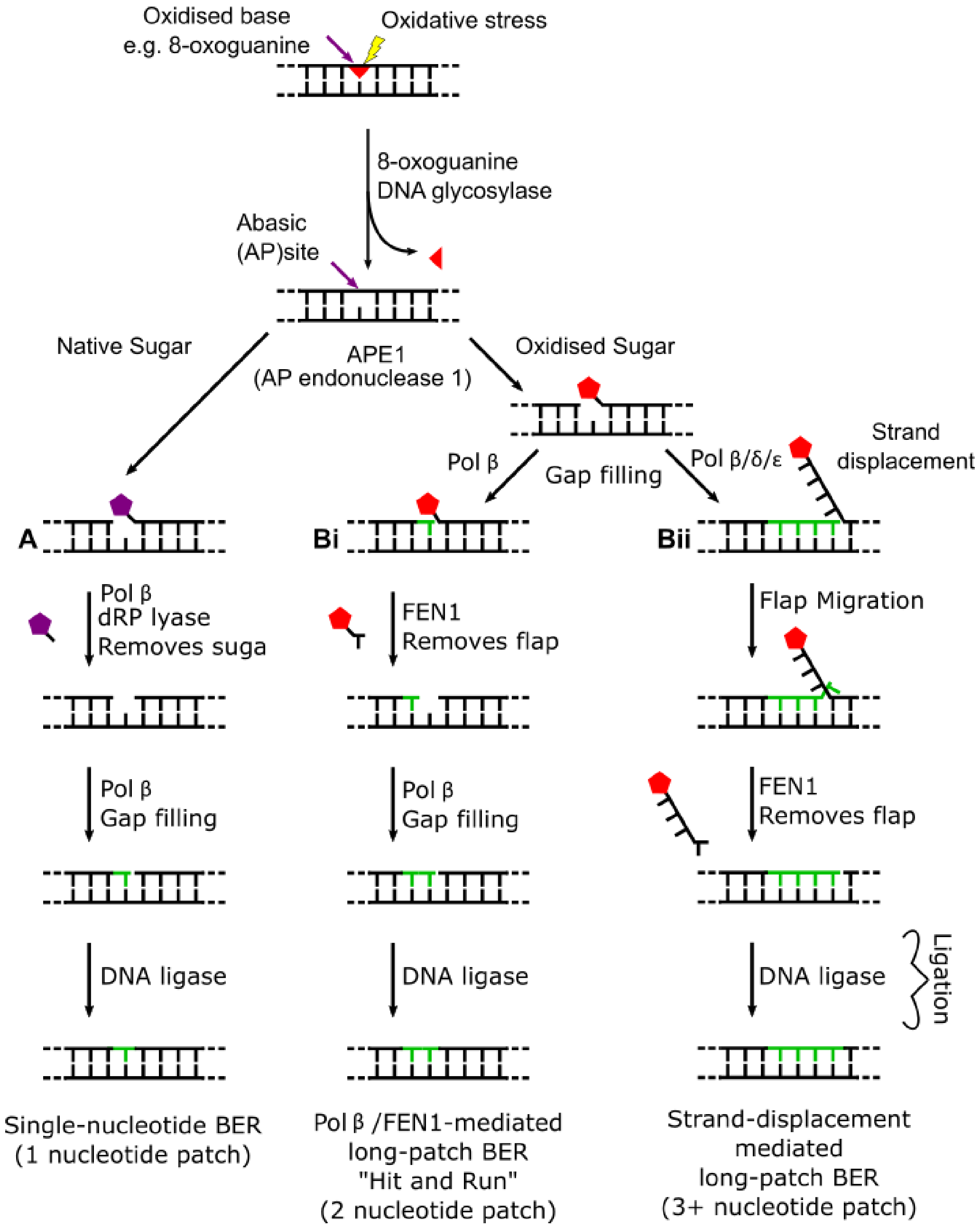


Figure 1.8: The removal of an oxidised base. *A DNA base can be oxidised by exogenous or endogenous means; in this example 8-oxoguanine is formed. This is removed by 8-oxoG DNA glycosylase generating an abasic site which can be excised by APE1.* ***A.*** *In this case the sugar is not oxidised and can be repaired by a combination of DNA Pol β and dRP lyase. Pol β fills in the gap and the nick is repaired by DNA ligase.* ***B.*** *Oxidised sugars must be repaired by alternative mechanisms.* ***i.*** *One method fills the gap with Pol β, removing the sugar by FEN1 cutting into the duplex, before Pol β fills the gap again. Finally DNA ligase repairs the nick.* ***ii.*** *Alternatively, strand displacement synthesis by Pol β/δ/ε forms a 5' flap, before migrating*

to form a more suitable substrate of FEN1. After removal of the flap the DNA simply requires ligation to form a repaired duplex. This figure is based on figure 1 of [64].

The two other forms of repair deal with an oxidised deoxyribose and rely on Pol β (though in some cases other polymerases can also be used), FEN1 and DNA ligase. The second pathway, illustrated in *figure 1.8Bi*, has been referred to as the "Hit and Run" method [65]. This description relates to the fact that Pol β first fills the gap created by APE1 with a single nucleotide. FEN1 then removes the oxidised deoxyribose and the next nucleotide before Pol β fills the second gap generated by the nucleotide's removal. Finally DNA ligase joins the nick. The final method of BER, *figure 1.8Bii*, begins with strand displacement synthesis by Pol β, which generates a 5' flap [66]. Before FEN1 can remove this flap the strand migrates to generate a single nucleotide 3' flap. After FEN1 has removed the flap the nick only requires processing by DNA ligase. This process is known as strand-displacement BER, and fits into long-patch BER with the Hit and Run method. For an oxidised sugar the most efficient process is strand-displacement mediated long-patch BER due to the fact that Pol β only has to process the sequence once, before FEN1 generates a product that requires minimal processing.

Defects in the BER pathways lead to predisposition to cancer formation [67]. BER has also been linked to the expansion of trinucleotide repeats (TNRs), which are associated with various neurodegenerative diseases such as Huntington's chorea [68-69]. In this case if a base is oxidised in a TNR region, BER attempts to repair this error. Upon removal of the oxidised base and APE1 cleavage at the AP site generated, the TNR can form a branched structure due to DNA slippage [64, 70]. Repair continues as normal but leaves a product with the addition of more TNR's as shown in *figure 1.9*. An *in vivo* example implicates the MSH2-MSH3 heterodimer MutSβ, a mismatch repair protein required for removal of longer IDLs, as being essential for expansion of CAG repeats by stabilising the hairpin branches for BER to extend the region [71-73].

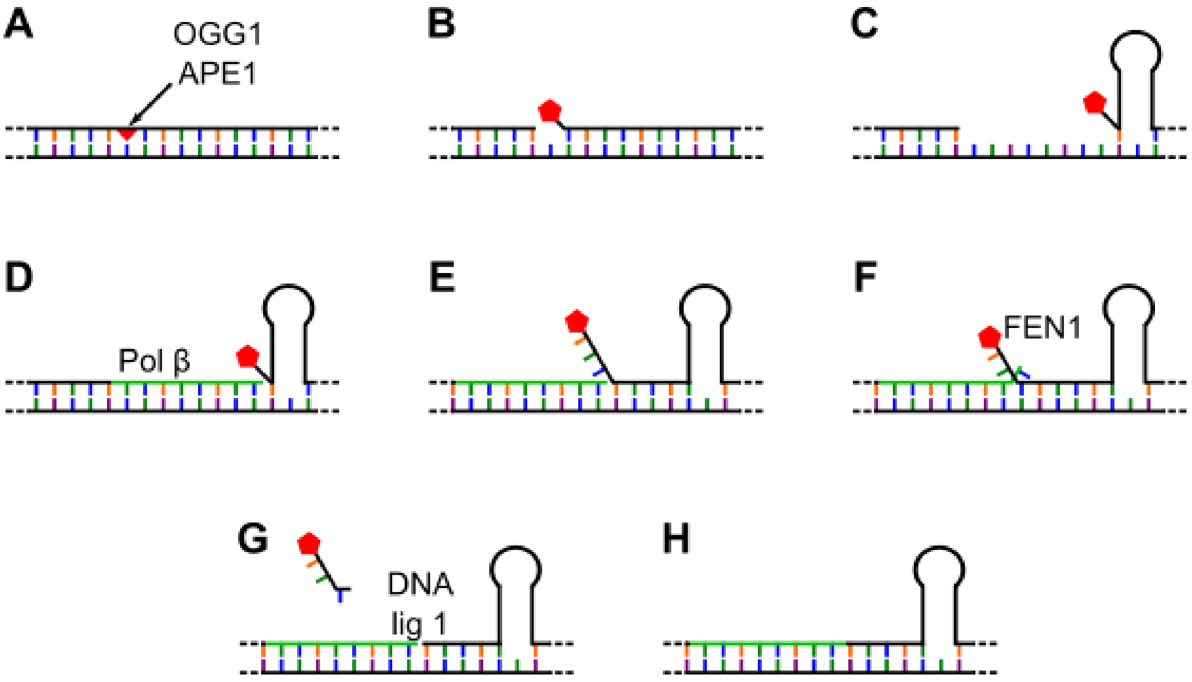


Figure 1.9: Trinucleotide repeat expansion by BER. *If BER is required in a region of TNR's; expansion can occur.* ***A & B.*** *A guanine that has been oxidised is removed by OGG1 and the abasic site is cleaved open by APE1.* ***C.*** *The DNA slips in this region to form a CAG hairpin structure.* ***D.*** *Pol β fills in the gap up to the oxidised sugar.* ***E.*** *FEN1 is unable to remove the sugar so close to the hairpin, therefore, the DNA realigns to allow access.* ***F.*** *The flaps migrate to form the optimal FEN1 substrate.* ***G.*** *FEN1 removes the flapped structure creating a nicked product.* ***H.*** *Finally DNA ligase religates the nick, leaving a repaired duplex with a hairpin, successfully expanding the TNR's on one strand. This figure is based on figure 3 [64].*

Although BER is a very important process that relies on FEN1, there is one instance where it can be hijacked by EXO1 [74]. In antibody maturation an enzyme known as activation-induced cytidine deaminase (AID) oxidises cytosine bases to uracil. Normally, BER takes over, the base is removed and APE1 opens up a nick. Normally Pol β and FEN1 would remove the sugar and resynthesise any removed nucleotides. However, in these regions it is predicted that EXO1 can resect the DNA, and an error-prone polymerase fills the gaps generated. This can cause nucleotide substitutions and alter the sequence in these regions to provide greater antibody diversity which is paramount to recognising different antigens [75-76]. This process is illustrated in *figure 1.10*.

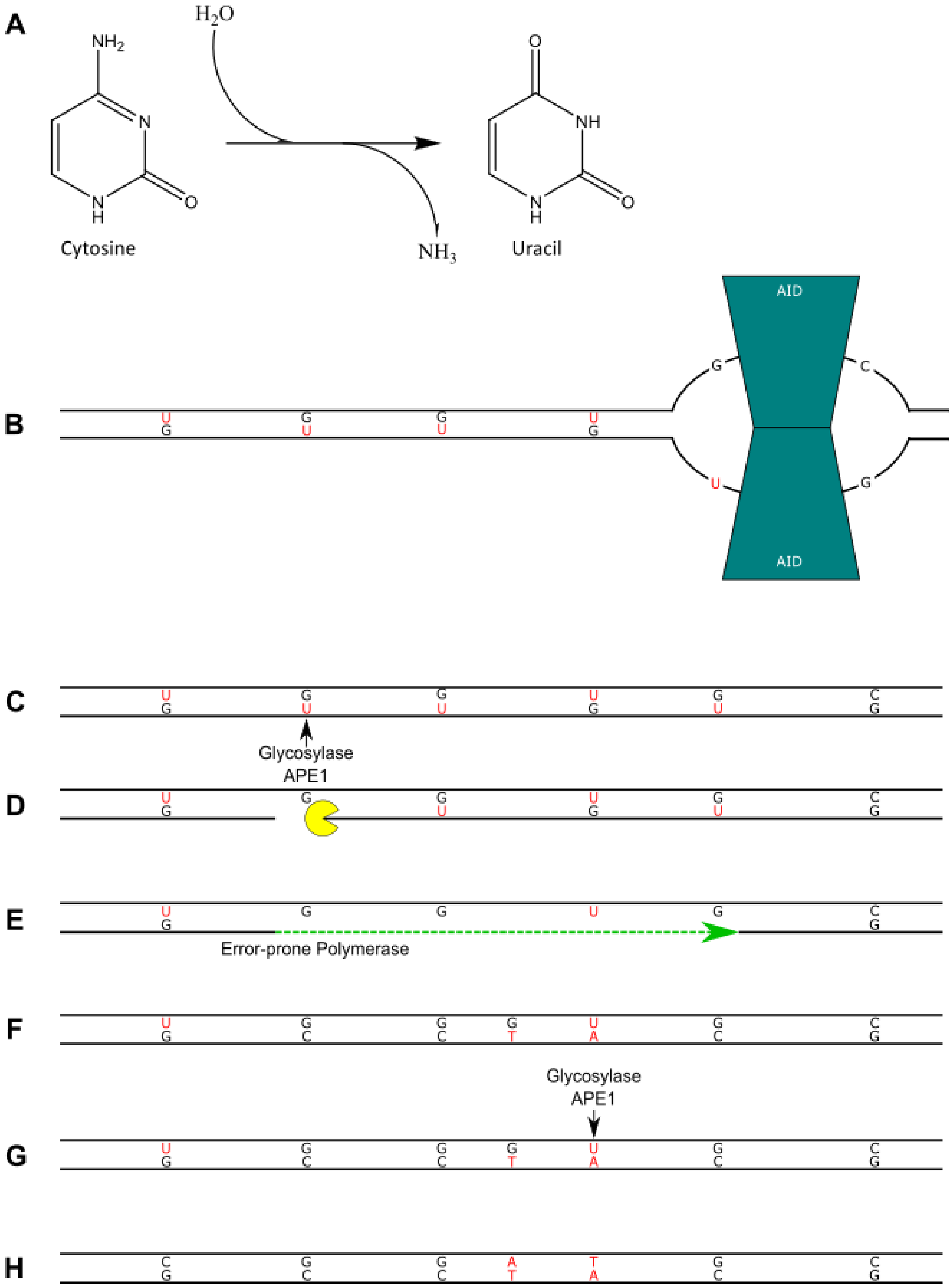


Figure 1.10: Somatic hypermutation in antibody maturation. A. *Cytosine nucleotides are deaminated to uracil with the addition of water and the removal of* NH_3. ***B.*** *Deamination by*

activation-induced cytosine deaminase (AID) as it scans along the DNA. ***C.*** *BER enzymes act, with a glycosylase to remove the base and APE1 to generate a nick.* ***D.*** *EXO1 hijacks the nick and resects tracts of DNA removing multiple uracil nucleotides.* ***E-F.*** *An error-prone DNA polymerase fills in the gap, which can cause nucleotide substitutions due to uracils on the sister strand as well as mistakes caused by a lack of proofreading capability of the polymerase.* ***G.*** *BER attempts repair on the sister strand but is again hijacked by EXO1.* ***H.*** *The repaired strand has various nucleotide substitutions providing greater antibody diversity.*

1.3.4 Nucleotide excision repair

Ultraviolet radiation (UV) causes exogenous damage to DNA, such as the dimerization of two adjacent thymine nucleobases leading to unpairing with the complementary strand due to internal stacking [77-79]. This dimer is recognised by the xeroderma pigmentosum type C (XPC), a member of a pathway of proteins, which initiates repair [80]. The other XP-factors coordinate the removal and repair of the bubble. XPA acts as a scaffold for other proteins [81], RPA coats the single-stranded DNA, and TFIIH, a transcription factor for other XP proteins, forms a complex with various proteins, including XPD and XPB, which act as bidirectional helicases [82]. The nucleases XPF and XPG perform bilateral cleavage of the bubble, starting with XPF generating a flapped structure which is removed by XPG [83-84]. Once the DNA bubble is removed, DNA polymerase and ligase repair the gap.

Alternatively, upon removal of the bubble, EXO1 can resect and extend the gap generated by the bubble’s removal [85]. The gapped DNA becomes saturated with RPA which in turn triggers the ATR pathway [85-86]. ATR refers to a group of ataxia telangiectasia and Rad3 related serine/threonine kinases, that in combination with other protein partners and various phosphorylation events, trigger the DNA damage response, which leads to apoptosis [87].

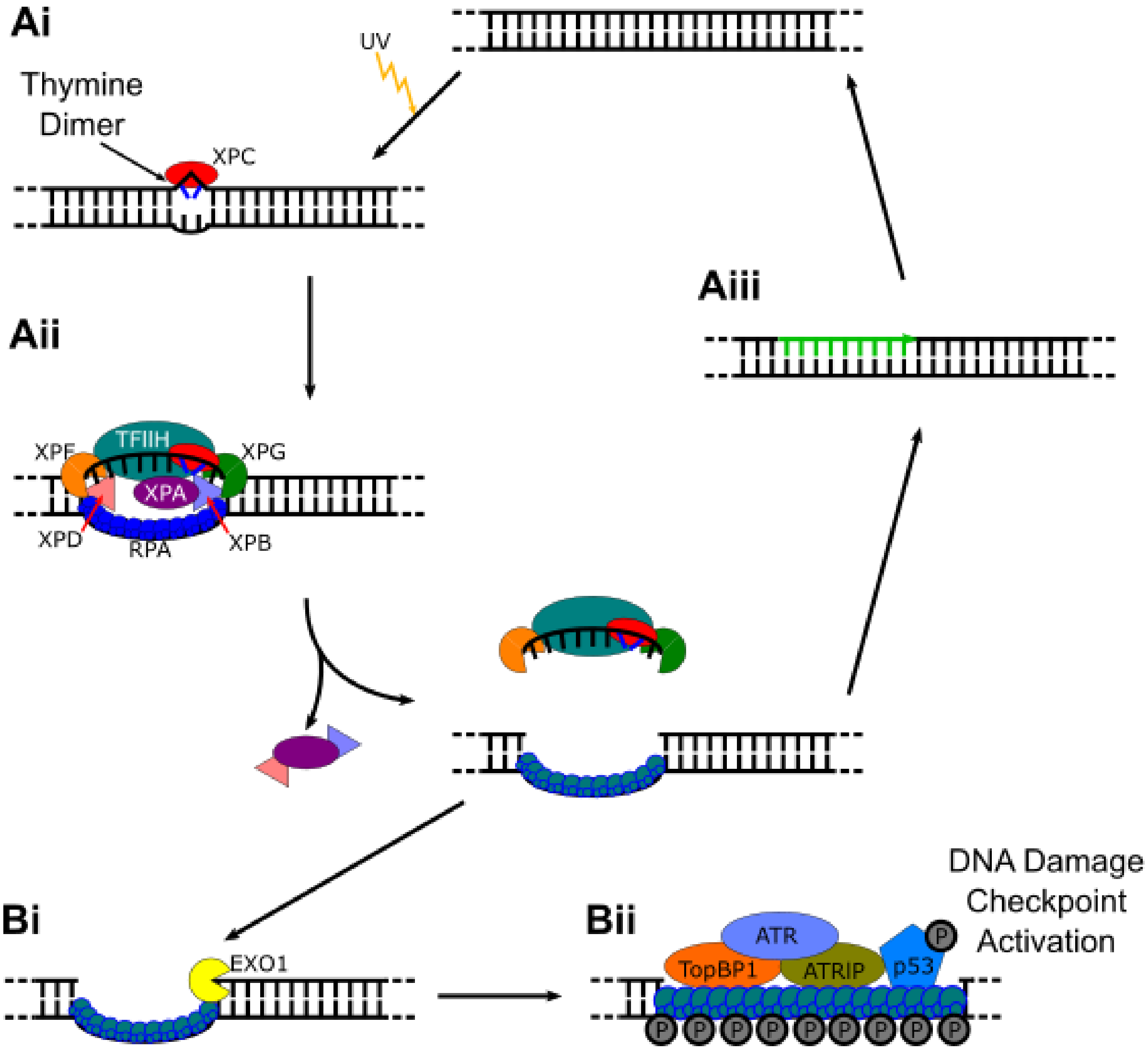


Figure 1.11: Nucleotide excision repair and DNA damage checkpoint activation. Ai. *UV light generates a thymine dimer, which is recognised by XPC.* ***Aii.*** *The other NER pathway proteins coordinate the removal of the DNA bubble.* ***Aiii.*** *The ssDNA gap is repaired by DNA synthesis and ligation.* ***Bi.*** *Alternatively, the gapped region is extended by EXO1.* ***Bii.*** *RPA coats the ssDNA and initiates the ATR pathway which leads to the activation of the DNA damage checkpoint.*

1.4 The flap endonuclease superfamily

It is clear that nucleases are important enzymes in pathways relating to genome repair and stability. The focus of the repair pathways mentioned has been a conserved family of nucleases.

1.4.1 Family members

The Flap endonuclease (FEN) superfamily of structure-specific 5' nucleases is a conserved family of enzymes which can be traced back to viruses [88]. There are four major DNA nuclease members and two ribonuclease members which take part in various DNA/RNA replication, repair or metabolic mechanisms. Flap endonuclease 1 (FEN1) is involved in Okazaki fragment maturation (*section 1.2.1*), where the enzyme removes flapped structures from patches of DNA in lagging strand synthesis (*figure 1.12A-B*) [16]. Exonuclease 1 (EXO1) is a key nuclease in various repair pathways, such as mismatch repair (*section 1.3.1*) and double strand break repair (*section 1.3.2*), where it removes tracts of DNA errors or damage allowing for repair by a DNA polymerase (*figure 1.12C-D*) [89-90]. Gap endonuclease 1 (GEN1) resolves Holliday junctions that form between homologous chromosomes (*figure 1.12E-F*) in meiosis (*section 1.2.2*) [21, 91]. Xeroderma pigmentosum (XPG) is part of a protein pathway charged with removing DNA bubbles generated by ultraviolet light damage (*section 1.3.4*). XPG removes the DNA bubble bilaterally with XPF before the duplex is repaired by a DNA polymerase (*figure 1.12G-H*) [86, 92]. The two RNA members are exoribonuclease (XRN) 1 and 2. XRN1 is known to degrade mRNA sequences, and XRN2 is thought to be involved in transcription termination; these RNA targeting enzymes will not be discussed in detail [93-94].

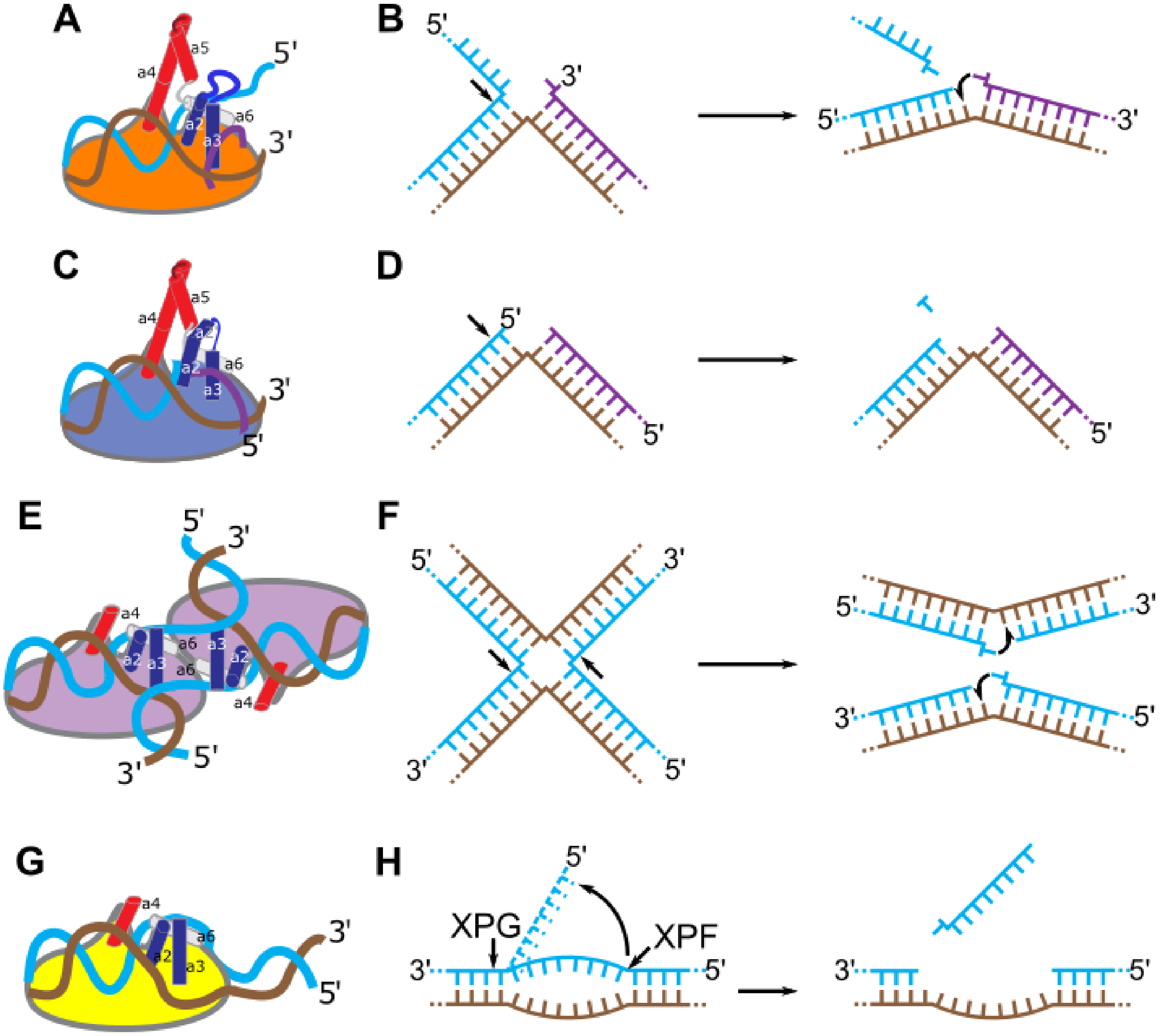
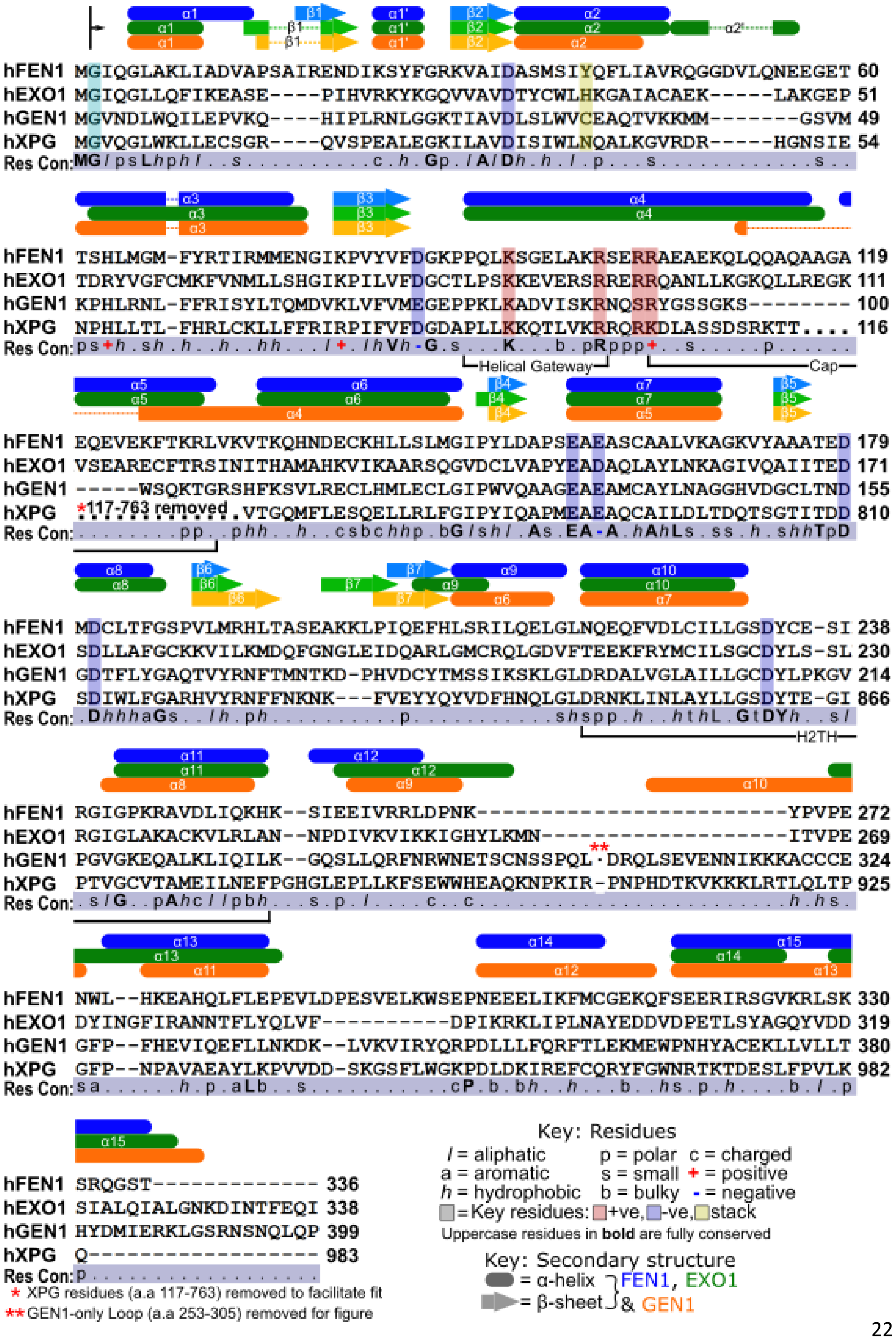


Figure 1.12: The FEN superfamily and their substrates. *Cartoon representations highlight the key helices in each protein, for example the α2-3 wedge domain and the α4-5 which form a helical archway in FEN1 and EXO1.* ***A.*** *Flap endonuclease 1 (FEN1) with a double flapped substrate.* ***B.*** *The preferred substrate of FEN1 with a 5' flap of any length and a 3' single nucleotide flap. FEN1 cuts one nucleotide into the duplex 5' to 3' and the gap is filled by the 3' flap.* ***C.*** *Exonuclease 1 (EXO1) with a nicked duplex.* ***D.*** *A nicked substrate is processed 5' to 3' removing single nucleotides; interestingly, the major product is also a substrate.* ***E.*** *Gap endonuclease 1 (GEN1) as a dimer with a Holliday junction (HJ).* ***F.*** *GEN1 acts as a dimer on HJ's cutting one nucleotide into the junction on both sides; this forms two duplexes that can be repaired by ligation.* ***G.*** *Xeroderma pigmentosum type G (XPG) with a DNA bubble substrate.* ***H.*** *XPG works in tandem with other repair proteins to process a DNA bubble, XPF cuts the opposing side of the bubble before XPG removes the flapped structure generated.*



```
hFEN1 MGIQGLAKLIADVAPSAIRENDIKSYFGRKVAIDASMSIYQFLIAVRQGGDVLQNEEGET 60
hEXO1 MGIQGLLQFIKEASE----PIHVRKYKGQVVAVDTYCWLHKGAIACAEK-----LAKGEP 51
hGEN1 MGVNDLWQILEPVKQ----HIPLRNLGGKTIAVDLSLWVCEAQTVKKMM-------GSVM 49
hXPG  MGVQGLWKLLECSGR----QVSPEALEGKILAVDISIWLNQALKGVRDR-----HGNSIE 54
Res Con:MG/psLhph/..s...........c.h.Gp./A/Dh.h./.p...s.............s..

hFEN1 TSHLMGM-FYRTIRMMENGIKPVYVFDGKPPQLKSGELAKRSERRAEAEKQLQQAQAAGA 119
hEXO1 TDRYVGFCMKFVNMLLSHGIKPILVFDGCTLPSKKEVERSRRERRQANLLKGKQLLREGK 111
hGEN1 KPHLRNL-FFRISYLTQMDVKLVFVMEGEPPKLKADVISKRNQSRYGSSGKS-------- 100
hXPG  NPHLLTL-FHRLCKLLFFRIRPIFVFDGDAPLLKKQTLVKRRQRKDLASSDSRKTT.... 116
Res Con:ps+h.sh.h..h..hh.../+./hVh-G.s...K...b.pRppp+..s.....p......

hFEN1 EQEVEKFTKRLVKVTKQHNDECKHLLSLMGIPYLDAPSEAEASCAALVKAGKVYAAATED 179
hEXO1 VSEARECFTRSINITHAMAHKVIKAARSQGVDCLVAPYEADAQLAYLNKAGIVQAIITED 171
hGEN1 -----WSQKTGRSHFKSVLRECLHMLECLGIPWVQAAGEAEAMCAYLNAGGHVDGCLTND 155
hXPG  *117-763 removed.VTGQMFLESQELLRLFGIPYIQAPMEAEAQCAILDLTDQTSGTITDD 810
Res Con:.........pp..phh..h..csbchhp.bG/sh/.As.EA-A.hAhLs.ss.h.shhTpD

hFEN1 MDCLTFGSPVLMRHLTASEAKKLPIQEFHLSRILQELGLNQEQFVDLCILLGSDYCE-SI 238
hEXO1 SDLLAFGCKKVILKMDQFGNGLEIDQARLGMCRQLGDVFTEEKFRYMCILSGCDYLS-SL 230
hGEN1 GDTFLYGAQTVYRNFTMNTKD-PHVDCYTMSSIKSKLGLDRDALVGLAILLGCDYLPKGV 214
hXPG  SDIWLFGARHVYRNFFNKNK---FVEYYQYVDFHNQLGLDRNKLINLAYLLGSDYTE-GI 866
Res Con:.DhhhaGs../h.ph...........p............shspp.h..hthL.GtDYh..s/

hFEN1 RGIGPKRAVDLIQKHK--SIEEIVRRLDPNK------------------------YPVPE 272
hEXO1 RGIGLAKACKVLRLAN--NPDIVKVIKKIGHYLKMN-------------------ITVPE 269
hGEN1 PGVGKEQALKLIQILK--GQSLLQRFNRWNETSCNSSPQL**·DRQLSEVENNIKKKACCCE 324
hXPG  PTVGCVTAMEILNEFPGHGLEPLLKFSEWWHEAQKNPKIR-PNPHDTKVKKKLRTLQLTP 925
Res Con:.s/G..pAhc//pbh...s.p./....c..c........................h.hs.

hFEN1 NWL--HKEAHQLFLEPEVLDPESVELKWSEPNEEELIKFMCGEKQFSEERIRSGVKRLSK 330
hEXO1 DYINGFIRANNTFLYQLVF----------DPIKRKLIPLNAYEDDVDPETLSYAGQYVDD 319
hGEN1 GFP--FHEVIQEFLLNKDK--LVKVIRYQRPDLLLFQRFTLEKMEWPNHYACEKLLVLLT 380
hXPG  GFP--NPAVAEAYLKPVVDD-SKGSFLWGKPDLDKIREFCQRYFGWNRTKTDESLFPVLK 982
Res Con:sa......h.p.aLb..s............cP.b.bh..h...b..hs.p.h....b./.p

hFEN1 SRQGST-------------- 336
hEXO1 SIALQIALGNKDINTFEQI 338
hGEN1 HYDMIERKLGSRNSNQLQP 399
hXPG  Q------------------ 983
Res Con:p..................
```

* XPG residues (a.a 117-763) removed to facilitate fit

** GEN1-only Loop (a.a 253-305) removed for figure

Figure 1.13: Sequence alignment of Flap endonuclease superfamily members. *The sequences were aligned using PROMALS3D with the secondary structure aligned above based on the crystal structures of hFEN1 (3Q8K.pdb), hEXO1 (3QEB.pdb) and hGEN1 (5T9J.pdb) [16, 89, 91, 95-96]; no crystal structure of hXPG has been assigned at this time. The residue consensus is displayed below the protein sequence based on the key at the bottom and catalytically important residues are highlighted in boxes. Key structural elements are also shown below the protein sequences. Residues 117-763 of hXPG were removed to facilitate the fit, and residues 253-305 of hGEN1 were removed from the figure as this loop is only present in GEN1. The arrow at the beginning of the sequence shows that the methionine residue in position 1 is removed post-translationally.*

All family members have a conserved catalytic core region, which is generally confined to the N-terminus of the protein. The exception is XPG which has a large region in between the two parts that make up the catalytic domain that is excluded from the alignment. The structurally conserved catalytic core is responsible for the nuclease activity of the individual proteins, and suggests a unified mechanism. The remainder of the proteins is made up of differing C-terminal domains (or in XPG, the gap between the nuclease domains and the remainder of the C-terminus) which allow for protein partner interactions that act as controls in their respective repair pathways [88]. Greater examination of specific regions of the superfamily proteins will encompass the remainder of this chapter.

1.4.2 The hydrophobic wedge

A region termed the "hydrophobic wedge" motif is formed by a combination of α2-3 which stabilises DNA bending to ~100^0 in hFEN1 and hEXO1 complexes [16, 89]. This hydrophobic wedge is conserved throughout the family members, with a high density of hydrophobic residues, especially in α3. EXO1 also contains an extra helical motif at the top of α2 termed α2', which potentially forms interactions with α5 and stabilises reaction once a substrate is bound [89]. For example, a possible salt bridge exists between glutamate 44 (E44) and arginine 116 (R116).

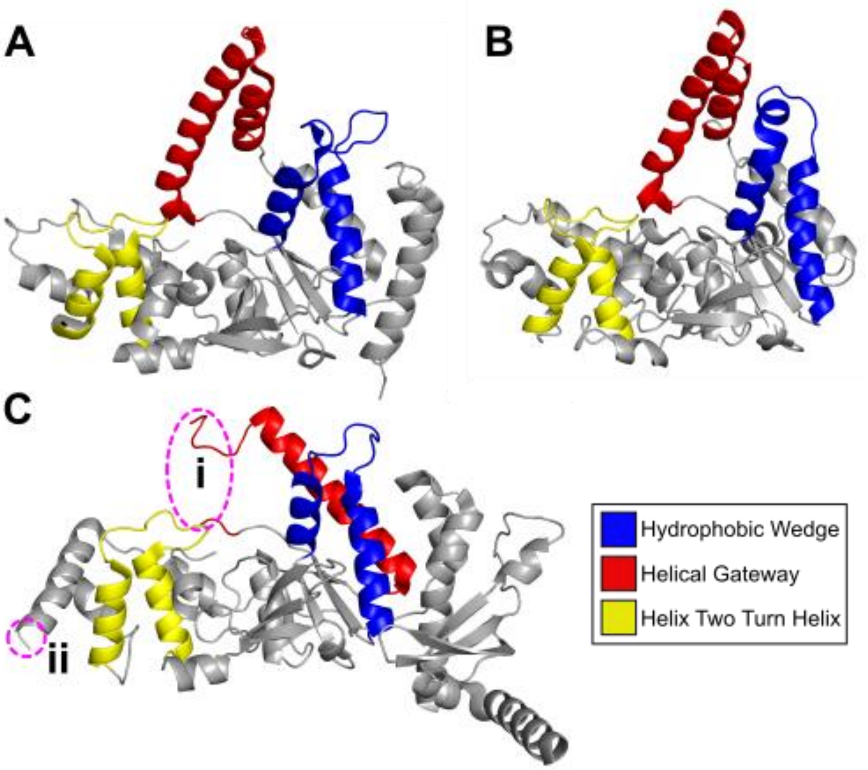


Figure 1.14: Crystal structures highlighting key regions. *The crystal structures of* ***A.*** *FEN1 (3Q8K.pdb),* ***B.*** *EXO1 (3QEB.pdb) and* ***C.*** *GEN1 (5T9J.pdb) highlighting the hydrophobic wedge (blue), helical gateway (red) and H2TH (yellow) of each protein. Structural gaps in GEN1 are displayed by* ***Ci*** *and* ***Cii****.*

1.4.3 Helical gateway and cap

A helical gateway is formed by the lower parts of α2 and α4 as highlighted in the sequence alignment. This region forms the radius around the active site above the two divalent cations which are required for catalysis [88]. The residues that form the gateway are conserved through the superfamily, which is illustrated by *figure 1.12* (A, C, E & G). However, the gateway is missing from the crystal structure of hGEN1, and as a result is not shown within *figure 1.13*, but the amino acid sequence is relatively conserved.

The latter portion of α4 combined with α5 acts as a cap at the top of the arch, forming a hole in the protein. This archway formation was a subject of controversy for many years with respect to how a flapped substrate is processed in hFEN1. However, it has been shown recently in both FEN1 and EXO1 that the gateway is big enough to facilitate threading of discontinuous flapped substrates through the hole before catalysis [97-101]. Interestingly, this cap is not present in GEN1 and XPG, displayed by a gap in the sequence. This is probably related to their characteristic substrate specificities, which involve continuous DNA structures. A Holliday junction is a closed duplex and can therefore not be threaded [21]. Whereas, XPG removes DNA damage bubbles, hydrolysing the DNA in tandem with another nuclease, which could facilitate threading a flapped structure. However, it has been shown that XPG binds to the DNA bubble prior to substrate cleavage, and as a result it seems unlikely to thread its substrates.

1.4.4 DNA sliding

A helix two turn helix (H2TH) motif present at α10-11 in FEN1 and EXO1 or α7-8 in GEN1 has been shown to co-ordinate with a potassium ion, allowing for interaction with the DNA backbone. Crystal structures of superfamily members have not shown the presence of a potassium ion in the absence of DNA, suggesting that it co-ordinates when the DNA binds [16, 89]. Carbonyl groups act as potassium-binding sites with water molecules; also, in FEN1 and EXO1 the hydroxyls of serine 237 and 229 respectively have been implicated. The K+/H2TH motif may facilitate local sliding of the protein on DNA, allowing junction seating to occur; this has been likened to the sliding capability of DNA polymerase β [102].

1.4.5 Less significant helical motifs

Other structural elements act in stabilising the enzymes. After the helical gateway, α6 (part of α4 in GEN1) spans the back of the hydrophobic wedge with various hydrophobic residues facing internally in the protein allowing for protein stability, whilst externally facing residues are either charged or polar. The α7 helix (α5 in GEN1) forms a scaffold through the centre of the enzyme, with cation-stabilising residues at the top of the helix. In FEN1 and EXO1 α8 has a similar role in cation stability. Various carboxylic residues are present around the active site for co-ordination of the metal ions, which will be mentioned below.

1.4.6 Active site residues

The active site co-ordinates divalent cations involved in catalysis of the reaction with various conserved acidic residues. These have been highlighted by blue boxes in *figure 1.13*, and include D34, D86, E159, E160, D179, D181 and D233 (FEN1 numbering) [16]. These residues are conserved through all the family members and emphasise the importance of the divalent metal ions in catalysis (*figure 1.15A*) [103]. Another key feature is the mainchain amino of G2 (as M1 is removed post-translationally), which is conserved in all the enzymes, and is suggested to interact with the nucleophilic water molecule that is added to the phosphate to hydrolyse the phosphodiester bond in FEN1 and EXO1 (*figure 1.15B*) [99].

Basic residues on the gateway interact with the oxygen atoms of the phosphodiester backbone, helping to stabilise the polynucleotide in the active site. For example, R104 is required for interaction with the 5’-phosphate (data collected as part of this thesis) in hFEN1. Also R100 mutants are catalytically inactive relative to wild-type (10,000-fold down). All basic residues highlighted in red are important for FEN1 catalysis, and are conserved through the superfamily (*figure 1.15B*) [99-100].

One last feature present in only FEN1 and EXO1 is a stacking residue present in α2, shown in *figure 1.15B*. The residue (Y40 in FEN1; H36 in EXO1) stacks with the nucleobase, allowing for a smooth transition of the nucleotide into the active site [16, 89, 99-100]. This interaction could be important in threading of substrates or at the very least positioning of the scissile phosphate, and significant drops in activity have been observed upon mutagenesis. No obvious stacking residue is present in GEN1 and XPG, which could represent its importance in discontinuous DNA substrates.

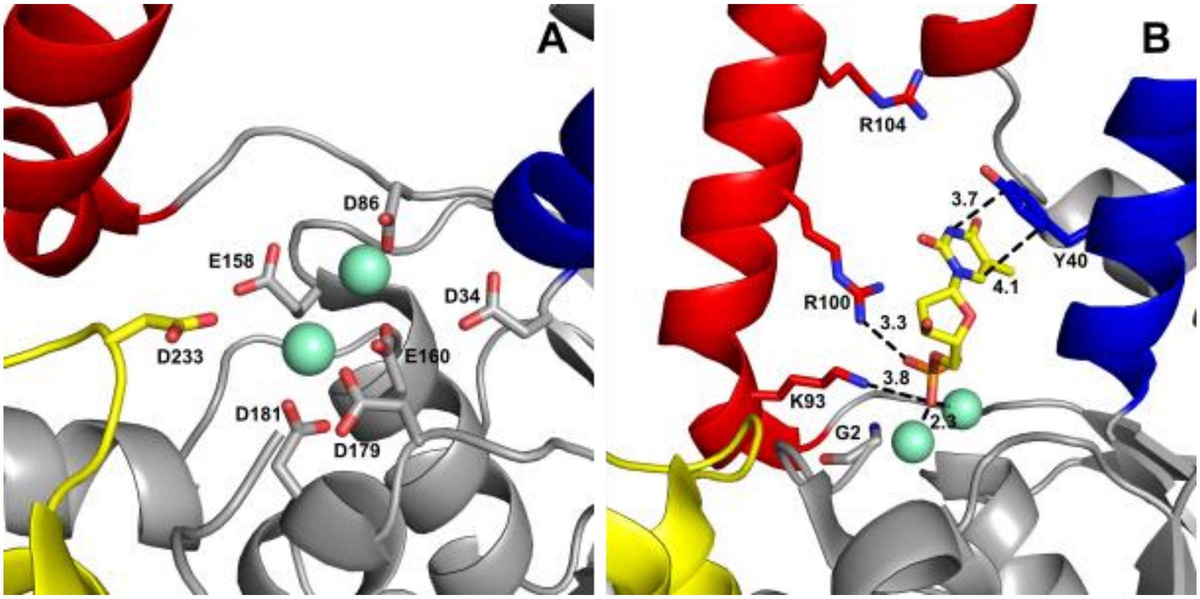


Figure 1.15: The active site residues of hFEN1. *The crystal structure of hFEN1 (3Q8K.pdb) was used to display the key residues that are conserved in the superfamily for catalysis.* ***A.*** *The carboxylate residues that interact with the divalent cations, forming a halo around the active site.* ***B.*** *The active site residues with distances in angstroms; R100 and K93 interact*

with the phosphate in combination with the divalent metal ions. The aromatic residue, Y40, which is only conserved in FEN1 and EXO1 shows stacking interactions with the nucleobase. The G2 amino is positioned into the active site for interaction with the water molecule.

1.4.7 Enzyme specific regions

As discussed throughout *sections 1.2* and *1.3*, each family member takes part in different repair pathways, although in some cases they can act as back-ups for one another. FEN1 has a requirement for single nucleotide 3' flaps, allowing for quick repair by DNA ligase when FEN1 cuts one nucleotide into the 5' duplex as the 3' fills the gap. To do this hFEN1 has developed a specialised pocket to facilitate 3' flap binding, aptly named the 3' flap binding pocket. This binding domain is made up of α14-15 helices and the looped region between α2 and 3. As shown in *figure 1.16A-B*, the crystal structure demonstrates a clear accommodation of the flap into the pocket [16]. The amino acid sequence beyond α15 forms a loop that contains a PIP-box for PCNA binding [103-104].

In the case of hEXO1, which takes part in multiple repair pathways, the enzyme contains an extended C-terminal domain for binding its protein partners from different pathways. Helices α14-15 of EXO1 form along the back of the enzyme, and begin the extended C-terminal region, shown in the crystal structure [89]. In *figure 1.17* the N-terminal core represents the 352 amino acids present in the crystal structure (3QEB.pdb) and displays protein-partner binding sites in the extended C-terminus beyond. The binding domains are for MSH3, MLH1 and MSH2, which are components of MutSβ, MutLα and MutSα respectively [105]. These proteins are all important in mismatch repair and represent a small number of binding partners. EXO1 is also postulated to have a PIP-box [104].

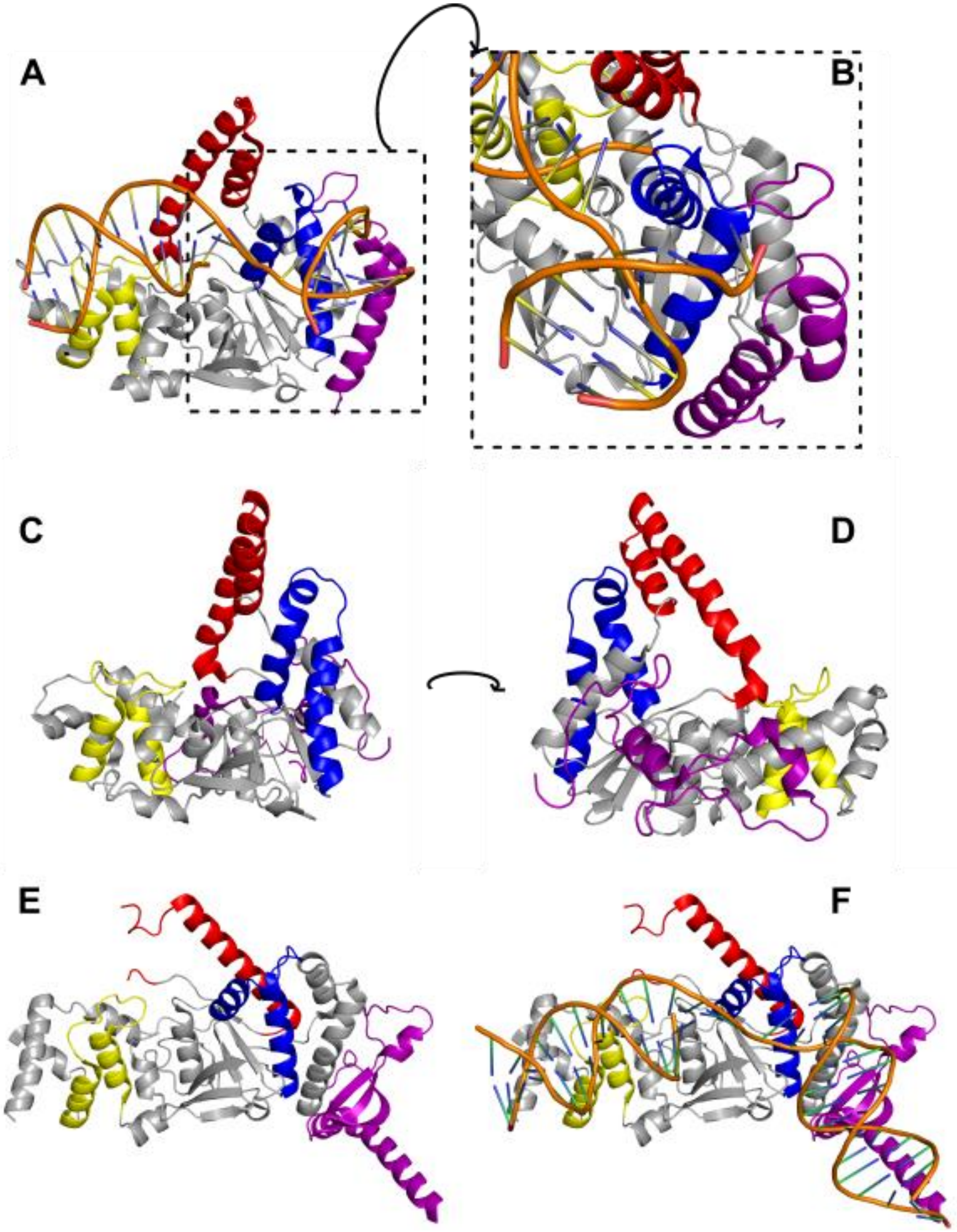


Figure 1.16: The C-terminal regions of family members. *The crystal structures of hFEN1 (3Q8K.pdb), hEXO1 (3QEB.pdb) and hGEN1 (5T9J.pdb) with C-terminal regions in purple. The red, yellow and blue motifs represent the helical archway (in FEN1 and EXO1), H2TH motif and hydrophobic wedge respectively.* ***A.*** *hFEN1 with a double flap DNA substrate.* ***B.***

The 3' flap binding pocket of hFEN1 (α14-15 and regions of the α2-3 loop) highlighted in purple; this region accommodates a single nucleotide 3' flap. ***C.*** *The front of hEXO1 displaying the nucleolytic core* ***D.*** *The back of hEXO1 shows the extended C-terminus that has been crystallised, prior to the enzymes truncation at 352 amino acids.* ***E.*** *The structure of hGEN1 with the chromodomain highlighted in purple.* ***F.*** *The hGEN1 crystal structure modelled with the DNA used to crystallise C. thermophilum GEN1 (5CO8.pdb) illustrating how DNA could interact with that region.*

The GEN1 substrate, the Holliday junction, is a 4-way junction made of continuous DNA strands that forms from crossover events of sister chromatids [21]. GEN1 acts as a dimer and must interact on opposing junctions of the 4-way junctions. As a result the extended C-terminal region contains a chromodomain which allows for extra DNA interaction to stabilise this process (*figure 1.16E-F*) [91]. In the case of hXPG, not a lot is known about the regions outside of the nucleolytic core, which are expected to be unstructured. Some evidence of a TFIIH binding site has been reported; TFIIH is a transcription factor used in the NER pathway [106].

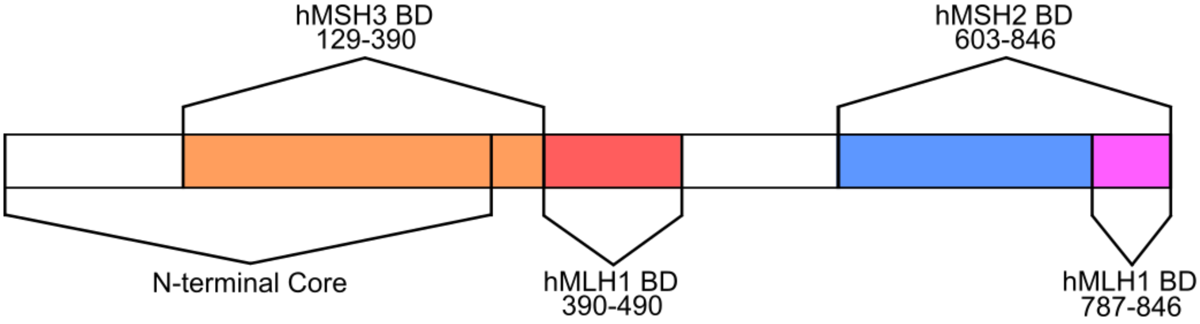


Figure 1.17: EXO1's Protein Partners in Mismatch Repair. *The binding domains for MSH3, MLH1 and MSH2 which are members of heterodimers with other proteins to form important mismatch repair complexes. The magenta region at the end is to illustrate an overlap of the MSH2 and the second MLH1 binding domains.*

1.5 Project aims

Human flap endonuclease 1 (hFEN1) has been extensively studied over the last decade. However, the other members of this family have not been examined in great detail. As the member of the family with the most sequence and structural homology to hFEN1 is hEXO1, this was the focus of this project. By studying the enzyme's substrate specificity, kinetic parameters of its exonuclease and endonuclease capabilities and various elements of substrate dynamics it is hoped that a reaction mechanism can be determined. This may provide clues to whether there is a universal reaction mechanism for the whole superfamily, despite their diverse roles *in vivo*. Testing of hEXO1 with compounds known to inhibit FEN1 will be performed to determine their specificity for FEN1. Finally, further studies into the threading capabilities of hEXO1 and determination of residues important for phosphate steering in hFEN1 and hEXO1 will be performed in the hopes of providing therapeutic targets for the future.

Chapter 2: Materials and Methods

- 2.1 Table of buffer
- 2.2 Purification of hEXO1-352
- 2.3 Oligonucleotide sequences and constructs used herein
- 2.4 Multiple turnover kinetics
- 2.5 Single turnover kinetics
- 2.6 Circular dichroism spectroscopy
- 2.7 Förster resonance energy transfer and fluorescence anisotropy

Chapter 2: Materials and Methods

2.1 Tables of buffers

2.1.1 Media and buffers for expression and purification

Buffer	Ingredients
Luria Bertani Broth (LB) 1L	10 g tryptone, 10 g NaCl , 5 g yeast extract
Terrific Broth (TB) 1L	12 g tryptone, 24 g yeast extract
SOB media 1L	20g tryptone, 0.5 g NaCl, 5 g yeast extract, 10 mL 250 mM KCl, 10 mL 1 M $MgCl_2$
SOC media	SOB media supplemented with 20 mM D-glucose
Inoue buffer	10 mM PIPES, pH 6.7, 55 mM $MnCl_2$, 15 mM $CaCl_2$ and 250 mM KCl
50×5052 (0.5%, 0.05% and 0.2% final - 5052)	25% glycerol, 2.5% glucose, 10% allolactose
20×Phosphate	1 M Na_2HPO_4, 1 M KH_2PO_4, 0.5 M $(NH_4)_2SO_4$
1000× metals	50 mM $FeCl_3$, 20 mM $CaCl_2$, 10 mM $MnCl_2$, 10 mM $ZnSO_4$, 2 mM $CoCl_2$, 2 mM $CuCl_2$, and 2 mM $NiCl_2$
Autoinduction media (AIM) 5 mL	4.64 mL LB, 10 µL 1 M $MgSO_4$, 1 µL 1000× metals, 100 µL 50×5052, 250 µL 20×Phosphate – Supplemented with antibiotics as appropriate
AIM 0.5 L	464 mL TB, 1 mL 1 M $MgSO_4$, 0.1 mL 1000× metals, 10 mL 50×5052, 25 mL 20×Phosphate - Supplemented with antibiotics as appropriate
10×Phosphate Buffer Saline (PBS)	1.37 M NaCl, 27 mM KCl, 100 mM Na_2HPO_4, 18 mM KH_2PO_4, pH 7.4
IMAC A1	20 mM tris, pH 7.0, 1 M NaCl, 5 mM imidazole, 0.02% (w/v) NaN_3, (5% glycerol (v/v)*)
IMAC A2	20 mM tris, pH 7.0, 0.5 M NaCl, 40 mM imidazole, 0.02% NaN_3, 0.1% Tween-20 (v/v), (5% glycerol (v/v)*)
IMAC B1	250 mM imidazole, pH 7.2, 0.5 M NaCl, 0.02% (w/v) NaN_3, (5% glycerol (v/v)*)
Anion A1	20 mM tris, pH 8.0, 1 mM EDTA, 0.02% (w/v) NaN_3, 20 mM β-mercaptoethanol (βME)
Anion B1	20 mM tris, pH 8.0, 1 mM EDTA, 1 M NaCl, 0.02% (w/v) NaN_3, 20 mM βME
Dialysis Buffer	25 mM tris, pH 8.0, 150 mM NaCl, 14 mM βME
Cation A1	100 mM tris, pH 7.0, 100 mM NaCl, 1 mM EDTA, 5% glycerol (v/v), 5 mM dithiothreitol (DTT)
Cation B1	100 mM tris, pH 7.0, 500 mM NaCl, 1 mM EDTA, 5% glycerol (v/v), 5 mM DTT
2× Storage Buffer	100 mM HEPES, pH 7.5, 200 mM KCl, 2 mM $CaCl_2$, 0.04% (w/v) NaN_3, 10 mM DTT, (20% glycerol (v/v)*)
Storage Buffer	50 mM HEPES, pH 7.5, 100 mM KCl, 1 mM $CaCl_2$, 0.02% (w/v) NaN_3, 50 mM DTT, 50% glycerol (v/v)

*glycerol supplemented in optimised purification scheme

2.1.2 Reaction Buffers

Buffer	Ingredients
1× Resolving Buffer	375 mM Tris, pH 8.8, 0.1% (w/v) SDS
1× Stacking Buffer	125 mM Tris, pH 6.8, 0.1% (w/v) SDS
10× Reaction Buffer (RB)	0.5 M HEPES, pH 7.5, 1 M KCl, 80 mM $MgCl_2$, 1 mg/mL bovine serum albumin (BSA)
1× Reduced RB (RRB)	50 mM HEPES, pH 7.5, 100 mM KCl, 8 mM $MgCl_2$, 0.1 mg/mL BSA, 1 mM DTT
10× Folding Buffer (FB)	0.5 M HEPES, pH 7.5, 1 M KCl
1× FB	50 mM HEPES, pH 7.5, 100 mM KCl
Quenchflow (QF) Purge Buffer	50 mM HEPES, pH 7.5, 100 mM KCl, 8 mM $MgCl_2$, 0.1 mg/mL, 0.02% NaN_3
QF single turnover (ST) push buffer	55 mM HEPES, pH 7.5, 110 mM KCl, 8 mM $MgCl_2$, 0.02% NaN_3, 15% glycerol
QF ST quench	1.5 M NaOH, 50 mM EDTA
5×Tris Borate EDTA (5×TBE)	54g Tris base, 27.5g boric acid, 20 mL 0.5 M EDTA, pH 8.0 Dilute down to 1×TBE
Denaturing PAGE gel 0.25L	105.1g urea, 100 mL 5×TBE, 50 mL 19:1 acrylamide. Mix with gentle heat until dissolved; make up to 250 mL with deionised water
Denaturing PAGE loading buffer	80% formamide, 10 mM EDTA, 1 mg/mL bromophenol blue and/or 1 mg/mL xylene cyanol FF. Can be run without dye.
Circular Dichroism (CD) RB	50 mM HEPES, pH 7.5, 100 mM KCl, 10 mM $CaCl_2$, 1 mM DTT
Förster Resonance Energy Transfer (FRET) RB	50 mM HEPES, pH 7.5, 100 mM KCl, 10 mM $CaCl_2$, , 0.1 mg/mL BSA, 1 mM DTT
Calcium SA* RB	25 mM HEPES, pH 7.5, 50 mM KCl. 2 mM $CaCl_2$, 0.1 mg/mL BSA, 0.5% glycerol, 1 mM DTT
Magnesium SA* RB	25 mM HEPES, pH 7.5, 50 mM KCl. 16 mM $MgCl_2$, 0.1 mg/mL BSA, 0.5% glycerol, 1 mM DTT
EDTA SA* RB	25 mM HEPES, pH 7.5, 50 mM KCl. 2 mM EDTA, 0.1 mg/mL BSA, 0.5% glycerol, 1 mM DTT
Calcium SA* push buffer	25 mM HEPES, pH 7.5, 50 mM KCl, 2 mM $CaCl_2$, 0.1 mg/mL BSA, 0.02% NaN_3, 15% glycerol
Magnesium SA* push buffer	25 mM HEPES, pH 7.5, 50 mM KCl, 16 mM $MgCl_2$, 0.1 mg/mL BSA, 0.02% NaN_3, 15% glycerol
SA* Quench	8 M Urea, 300 mM EDTA

*SA = Streptavidin

2.2 Purification of hEXO1-352

2.2.1 Generation of competent cells by the Inoue method

Chemically competent cells were produced via the Inoue methodology [107]. Previously produced BL21(DE3)-RIPL (Cm^R) competent *E coli* cells were plated onto 1.5% agar LB plate supplemented with 34 µg/mL chloramphenicol (Cm) and grown for 16-20 hours at 37 ^oC. A single colony was used to inoculate 25 mL of LB broth containing 34 µg/mL Cm and incubated for 6-8 hours at 250 rpm. Three 1L flasks containing 250 mL of SOB media containing 34 µg/mL Cm were inoculated with 6, 4 and 2 mL before incubation overnight at 18-22 ^oC at 200 rpm. The following morning the OD_{600} of the cultures was monitored, checking every 45 minutes. Once one of the cultures reached an OD_{600} of 0.55 the flask was placed in ice-cold water for 10 minutes and the other cultures were discarded appropriately. The chosen cells were centrifuged for 10 minutes at 2500 xg at 4 ^oC before the removal of the supernatant and the pellet was thoroughly dried. The pellet was resuspended in 80 mL of ice-cold Inoue buffer by swirling. The cells were once again centrifuged for 10 minutes at 2500 xg at 4 ^oC and the pellet was thoroughly dried. The pellet was then resuspended in 20 mL of ice-cold Inoue buffer before the addition of 1.5 mL of DMSO and mixing by swirling. The mixture was stored on ice for 10 minutes, and then 100 µL aliquots were dispensed into 1.5 mL eppendorfs and flash frozen with liquid nitrogen. The cells were stored at -80 ^oC before until required for transformation.

2.2.2 Cloning and expression of hEXO1-352

A codon-optimised (for *E coli*) truncated human EXO1 plasmid was acquired from GeneArt (Thermo Fisher scientific) with a TEV-cleavable-$(His)_6$ tag coded into the sequence (available in *figure A1* in the appendix). The optimised construct was subcloned into the pET21a vector (Novagen) (Amp^R) using the NdeI and NotI restriction sites. This expression vector (graphic representation is available in *figure A2* in the appendix) was transformed into chemically competent BL21(DE3)-RIPL (Cm^R) *E coli* cells (cells were produced by the Inoue method in *section 2.2.1*). Approximately 50 ng of vector was incubated with 50 µL of cells for 30 minutes on ice, before heat shocking at 42^{0}C for 90 seconds. The cells were then returned to the ice for 5 minutes before being rescued by the addition of 1 mL of SOC media and incubation at 37^{0}C for one hour. The cells were plated on LB-Agar plates supplemented with 100 µg/mL ampicillin (Amp) and 34 µg/mL Cm.

The hEXO1-352-TEV-(His)$_6$ protein (the exact protein sequence is available in *figure A3* in the appendix) was produced first by the inoculation of a 25 mL LB starter culture supplemented with 100 μg/mL Amp and 34 μg/mL Cm and allowed to grow for 16-17 hours overnight at 37^{0}C. Autoinduction media [108] was used to induce the plasmid in multiple 2L flasks, each containing 500 mL of media supplemented again with 100 μg/mL Amp and 34 μg/mL Cm, with inoculation with 5 mL of starter culture and grown for 4 hours at 37^{0}C with shaking at 210 rpm. After this fast-growth period of incubation the temperature was reduced to 21^{0}C and the culture was grown for between 14-16 hours to an OD$_{600nm}$ of 14-18. The cultures were pelleted by centrifugation at 6000 ×g for 20 minutes at 4^{0}C, and washed with ice-cold 1×PBS after the removal of the supernatant before further centrifugation at 4000 ×g for 15 minutes at 4^{0}C. The supernatant was discarded once again and the pellet was resuspended in lysis buffer (IMAC A1) plus a final concentration of 5 mM β-mercaptoethanol and incubated at 4^{o}C on a roller mixer for 1-2 hours. Lysis was completed by freezing at -20^{0}C and the lysate was stored until use.

2.2.3 Purification of hEXO1-352

Cell lysates were sonicated using a Vibra-cell VCX-130 ultrasonic liquid processor (Sonics & Materials, inc) at 70% amplitude for 10 seconds on and 20 seconds off until smooth, before the supplementation of 10% Tween-20 to a final w/v of 1%. The resulting solution was centrifuged at 30,000 xg for 20 minutes at 4^{o}C to separate the soluble and insoluble fractions. All purification columns were supplied from GE healthcare and columns were used in conjunction with an Äkta pure FPLC (GE healthcare). Initial purification success of hEXO1-352-TEV-(His)$_6$ was achieved by the following methodology. The supernatant was loaded onto three tandem 5 mL crude Co^{2+}-affinity chromatography column (immobilised metal-affinity chromatography (IMAC)) and elution was achieved with a 10 column volume (CV) gradient of 40-250 mM imidazole (IMAC A2 and B1 with 5 mM β-mercaptoethanol). The fractions containing hEXO1-352-TEV-(His)$_6$ were loaded onto a 5 mL HiTrap Q anion exchange column to remove DNA contamination; due to the fact that hEXO1-352 binds to DNA it flows through the positively charged column.

The 6-His tag was removed by TurboTEV in the ratio described in the manufacturer's instructions (Bioscience, Ltd), 1 unit TurboTEV to 10 μg of target protein; concentrations were approximately determined via the Bradford assay. To do this 50 μL of sample was added to 950 μL of Bradford reagent (Bio-rad; named bio-rad reagent), incubated for 5 minutes at room temperature and absorbance was determined at 595 nm using a Nanodrop microvolume spectrophotometer (ThermoFisher scientific). The protein was dialysed into an appropriate buffer for TurboTEV with low enough NaCl (the buffer was changed once, and dialysis was in 2L's of buffer) for successful loading onto the next column. The dialysed sample was centrifuged at 4,000 rpm for 20 minutes at 4^0C to remove any precipitate and loaded onto a 5 mL HiTrap SP-Cation exchange column pre-equilibrated with Cation buffers A1 and B1, with elution by a 10-column volume gradient of 100-500 mM NaCl. The cation exchange flow-through was re-run through the cation exchange column again several times to extract as much hEXO1-352 as possible from the dialysed solution. Finally the pure protein was concentrated to a volume of 10 mL and desalted into 2× storage buffer by desalting with a 50 mL HiPrep 26/10 desalting column. The desalted sample was concentrated by Vivaspin-20 centrifugal concentrators to 200 μM of enzyme (determined by absorbance at 280 nm using a Nanodrop microvolume spectrophotometer, the molecular weight and extinction coefficient shown below) before dilution 1:1 with 100% glycerol that had been treated with Chelex-100 and stored at -20^0C. This mechanism produced ~10 mL of pure 100 μM hEXO1-352 from 8 litres of cell culture. Human EXO1-352 has a molecular weight of 40427.9 Da and an extinction coefficient of 28475 M^{-1} cm^{-1} (at 280 nm, assuming all cysteines are reduced, which was determined using ExPASy ProtParam tool (https://web.expasy.org/protparam/))

2.2.4 Optimised purification of hEXO1-352

Protein purification was optimised to develop a more efficient scheme. The addition of 5% glycerol stabilised hEXO1-352 against precipitation. The cells were grown, lysed, sonicated and centrifuged above to separate the soluble and insoluble fractions. The supernatant was diluted with gentle mixing 1:1 with deionised water before loading onto a three tandem 5 mL crude Co^{2+}-affinity chromatography column. The standard IMAC buffers were supplemented with 5% glycerol and 2.5 mM β-mercaptoethanol. A column wash was performed with 40 mM imidazole (IMAC A2) before isocratic elution with 250 mM imidazole (IMAC B1). Fractions containing hEXO1-352-TEV-$(His)_6$ were loaded onto a 5 mL HiTrap SP-Cation exchange column. This step combined the removal of DNA contamination with isolation of positively charged enzymes. The protein was eluted using a gradient between 100-500 mM NaCl with Cation buffers A1 and B1. The eluted sample was incubated with TurboTEV as described above overnight at 4^{o}C to remove the 6-His tag. TurboTEV was removed from solution by incubation at 4^{o}C for 1 hour on a roller mixer with MagneGST beads (Promega corporation) per the manufacturer's instructions. Finally the pure protein was concentrated in batches of 10 mL and desalted into 2× storage buffer containing 20% glycerol using a 50 mL HiPrep 26/10 desalting column. Samples were concentrated to 200 μM of protein (determined at 280 nm using a nanodrop microvolume spectrophotometer), and diluted with 80% glycerol to a final w/v of 50%, and stored at -20^{0}C. This optimised scheme produced ~30 mL of pure 100 μM hEXO1-352 from 4 litres of cell culture.

Protein purity was determined by analysis of various concentrations by a 12% Sodium Dodecyl Sulphate – Polyacrylamide Gel Electrophoresis (SDS-PAGE); composed of an 8 cm resolving gel (1× Resolving buffer) and a 2 cm stacking gel (1× Stacking buffer). Gel layers contained 12% acrylamide and were polymerised with 0.1% APS and TEMED. Separation was performed at 250 volts for 30 minutes followed by staining with InstantBlue protein stain (Sigma-Aldrich) for 20 minutes.

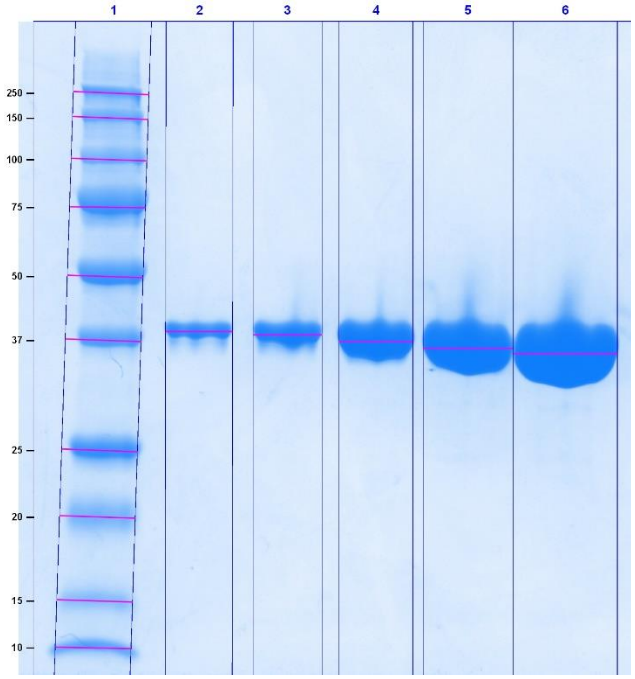


Figure 2.1: SDS-PAGE purity gel of hEXO1-352 after purification. *Lane 1: Protein Ladder (Precision plus protein^{TM} all blue stained protein standard; Bio-rad – weights to the left of gel), lanes 2-6 contain 1, 2, 4, 8 and 16 μg of protein respectively. Enzyme mass load was determined based on 100 μM sample equating to ~4 μg/μL.*

2.3 Oligonucleotide sequences and constructs used herein

2.3.1 Oligonucleotide strand sequences

The oligonucleotides (*Table 2.1*) were produced and HPLC-purified by LGC Biosearch. All the sequences used are detailed below with modifications included; a combination of the oligonucleotides used in various constructs are shown in their respective section.

Table 2.1A: Kinetic oligonucleotide sequences used herein

Oligo	Sequence
E1	5'-**FAM**-ACAAGGACTGCTCGACAC-3'
E2	5'-**PHOS**-ACAAGGACTGCTCGACAC-**FAM**-3'
E3	5'-ACAAGGACTGCTCGACAC-**FAM**-3'
E4	5'-**PHOS**-ACAAGGACTGCTCGACAC-3'
F1	5'-**FAM**-TTTTTACAAGGACTGCTCGACAC-3'
F2	5'-TTTTTACAAGGACTGCTCGACAC-**FAM**-3'
F3	5'-**FAM**-TTTTTTTTTTGAGGCAGAGTAGGACC-3'
T1	5'-GTGTCGAGCAGTCCTTGTGACGACGAAGTCGTCC-3'
T2a	5'-**BioTEG**-GGTCCTACTCTGCCTCAAGAGAGAGACGGTCTGCTGCACTGGATCTGG-3'
T2b	5'-**BioTEG**-CCAGATCCAGTGCAGCAGACCGTCTCTCTCC-3'
T3	5'-GTGTCGAGCAGTCCTTGT-3'
T4	5'-GTGTCGAGCAGTCCTTGTGACGACGAAGTCGTC-3'
T5	5'-GTGTCGAGCAGTCCTTGTGACGAC-3'
Bio-psY	5'-**FAM**-**BioTEG**-GAACACACAGAACACACACCGCTTGCGGTGTGTGTTTCCACAAC-3'

Table 2.1B: ECCD oligonucleotide sequences used herein

Oligo	Sequence
EEC1	5’-**PHOS**-**2-AP**-**2-AP**-GAGGCAGAGTG-3’
EEC2	5’-**PHOS**-G-**2-AP**-**2-AP**-AGGCAGAGTG-3’
FEC1	5’-TTTTT-**2-AP**-**2-AP**-GAGGCAGAGTG-3’
FEC2	5’-TTTTTG-**2-AP**-**2-AP**-AGGCAGAGTG-3’
UEEC1	5'-**PHOS**-**2-AP**-**2-AP**-GAGGCAGAGTGCGTGCACTCTGCCTCTTGACAGCG-3'
UEEC2	5'-**PHOS**-G-**2-AP**-**2-AP**-AGGCAGAGTG CGTG CACTCTGCCTTTCGACAGCG-3'
TC1	5’-CACTCTGCCTCTTGACAGCGAAGCTGTCC-3’
TC2	5’-CACTCTGCCTTTCGACAGCGAAGCTGTCC-3’
TC3	5’-CACTCTGCCTCTTGACAGC-3’
TC4	5’-CACTCTGCCTTTCGACAGCG-3’
TC5	5’-CACTCTGCCTCTTGACAGCGAAGCTGTC-3’
TC6	5’-CACTCTGCCTTTCGACAGCGAAGCTGTC-3’

Table 2.1C: FRET oligonucleotide sequences used herein

Oligo	Sequence
E_P	5'-**PHOS**-TTGAGGCAGAGTAGGACC-3'
F_P	5'-**PHOS**-TTTTTT TTG AGG CAG AGT AGG ACC-3'
T_{na}	5'-GGTCCTACTCTGCCTCAA GACGGTCTGCTGCACTGG-3'
T_A	5'-GGTCC-**TAMRAdT**-ACTCTGCCTCAAGACGGTCTGCTGCACTGG-3'
T_{nd}	5'-CCAGTGCAGCAGACCGTCC-3'
T_{nd2}	5'-CCAGTGCAGCAGACCGTC-3'
T_D	5'-CCAG-**FAMdT**-GCAGCAGACCGTCC-3'
T_{D2}	5'-CCAG-**FAMdT**-GCAGCAGACCGTC-3'

Sequences of individual oligonucleotides used. (**PHOS**): phosphate, (**FAM**): fluorescein, (**BioTEG**): biotin, (**2-AP**): 2-aminopurine and (**TAMRA**): tetramethylrhodamine.

2.3.2 DNA constructs

Table 2.2A: DNA kinetic constructs

Construct	Oligo Combination	Description
BD1	E1 + T3	5' **FAM**, blunt duplex
BD2	E2 + T3	3' **FAM**, 5' **phosphate**, blunt duplex
ND1	E1 + T4	5' **FAM**, nicked duplex
ND2	E2 + T4	3' **FAM**, 5' **phosphate**, nicked duplex
ND3	E2 + T4	3' **FAM**, 5' **hydroxyl**, nicked duplex
OVH1	E1 + T5	5' **FAM**, 3' overhang
OVH2	E2 + T5	3' **FAM**, 5' **phosphate**, 3' overhang
OVH3	E4 + T5	5' **phosphate**, 3' overhang competitor
PY1	F1 + T5	5' **FAM**, pseudo-Y
PY2	F2 + T5	3' **FAM**, pseudo-Y
PY3	Bio-psY	5' **FAM**, 5' **biotin**, unimolecular pseudo-Y
SF1	E1 + T1	5' **FAM**, single flap
SF2	E2 + T1	3' **FAM**, 5' **phosphate**, single flap
SF3	E3 + T1	3' **FAM**, 5' **hydroxyl**, single flap
DF	F1 + T1	5' **FAM**, double flap
DF2	F3 + T2a + T2b	5' **FAM** flap, 5' **biotin** (T2a and T2b), double flap

The oligonucleotides (Table 2.1A) were form the substrate constructs in table 2.2A.

Table 2.2B: DNA ECCD constructs

Construct	Oligo Combination	Description
SF$_{+1-1}$	EEC1 + TC1	5' **phosphate**, **2-AP** +1-1 positions, single flap
SF$_{-1-2}$	EEC2 + TC2	5' **phosphate**, **2-AP** -1-2 positions, single flap
OVH$_{+1-1}$	EEC1 + TC3	5' **phosphate**, **2-AP** +1-1 positions, 3' overhang
OVH$_{-1-2}$	EEC2 + TC4	5' **phosphate**, **2-AP** -1-2 positions, 3' overhang
N$_{+1-1}$	EEC1 + TC5	5' **phosphate**, **2-AP** +1-1 positions, nicked duplex
N$_{-1-2}$	EEC2 + TC6	5' **phosphate**, **2-AP** -1-2 positions, nicked duplex
UOVH$_{+1-1}$	UEEC1	5' **phosphate**, **2-AP** +1-1 positions, unimolecular 3' overhang
UOVH$_{-1-2}$	UEEC2	5' **phosphate**, **2-AP** -1-2 positions, unimolecular 3' overhang
DF$_{+1-1}$	FEC1 + TC1	**2-AP** +1-1 positions, double flap
DF$_{-1-2}$	FEC2 + TC2	**2-AP** -1-2 positions, double flap
PY$_{+1-1}$	FEC1 + TC3	**2-AP** +1-1 positions, pseudo-Y
PY$_{-1-2}$	FEC2 + TC4	**2-AP** -1-2 positions, pseudo-Y

The oligonucleotides (Table 2.1B) were form the substrate constructs in table 2.2B.

Table 2.2C: DNA FRET constructs

Construct	Oligo Combination	Description
ND$_{NL}$	E$_P$ + T$_{na}$ + T$_{nd2}$	Non-labelled, 5' **phosphate**, nicked duplex
ND$_{DOL}$	E$_P$ + T$_{na}$ + T$_{D2}$	Internal **FAM**, 5' **phosphate**, nicked duplex
ND$_{AOL}$	E$_P$ + T$_A$ + T$_{nd2}$	Internal **TAMRA**, 5' **phosphate**, nicked duplex
ND$_{DAL}$	E$_P$ + T$_A$ + T$_{D2}$	Internal **FAM** and **TAMRA**, 5' **phosphate**, nicked duplex
DF$_{NL}$	F$_P$ + T$_{na}$ + T$_{nd}$	Non-labelled, 5' **phosphate**, double flap
DF$_{DOL}$	F$_P$ + T$_{na}$ + T$_D$	Internal **FAM**, 5' **phosphate**, double flap
DF$_{AOL}$	F$_P$ + T$_A$ + T$_{nd}$	Internal **TAMRA**, 5' **phosphate**, double flap
DF$_{DAL}$	F$_P$ + T$_A$ + T$_D$	Internal **FAM** and **TAMRA**, 5' **phosphate**, double flap

The oligonucleotides (Table 2.1C) were form the substrate constructs in table 2.2C. They were first incubated at 95°C for 5 minutes in 1×FB before cooling at room temperature to anneal.

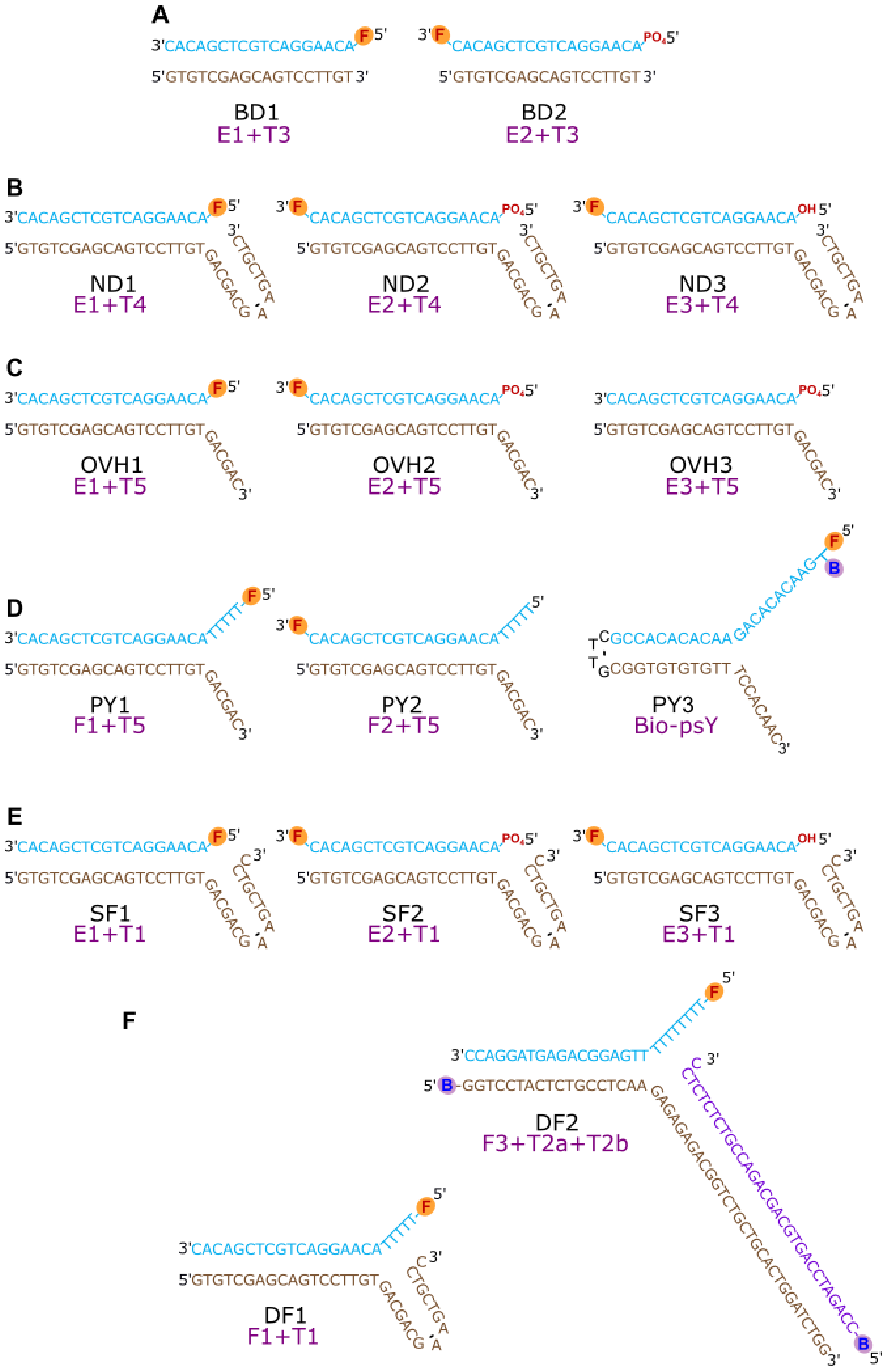

A
3'CACAGCTCGTCAGGAACA F 5'
5'GTGTCGAGCAGTCCTTGT 3'
BD1
E1+T3
3' F CACAGCTCGTCAGGAACA PO4 5'
5'GTGTCGAGCAGTCCTTGT 3'
BD2
E2+T3
B
ND1
E1+T4
ND2
E2+T4
ND3
E3+T4
C
OVH1
E1+T5
OVH2
E2+T5
OVH3
E3+T5
D
PY1
F1+T5
PY2
F2+T5
PY3
Bio-psY
E
SF1
E1+T1
SF2
E2+T1
SF3
E3+T1
F
3'CCAGGATGAGACGGAGTT
5' B -GGTCCTACTCTGCCTCAA
DF2
F3+T2a+T2b
DF1
F1+T1

Figure 2.2: Constructs used in kinetic experiments. ***A****. Blunt duplex (BD)* ***B****. Nicked duplex (ND)* ***C****. 3' overhang (OVH)* ***D****. Pseudo-Y (PY)* ***E****. Biotinylated pseudo-Y (BioPY)* ***F****. single 3' flapped duplex (SF)* ***G****. double flapped duplex (DF)* ***H****. extended double flapped duplex (DF2). Reacting strands are highlighted in cyan and template strands are highlighted in brown (and purple for DF2). Ends highlighted in red (e.g. 3') indicate that multiple modifications are present for this constructs (see table 2.3.1b). For unique constructs the modifications for fluorescein and biotin are highlighted (F and B respectively).*

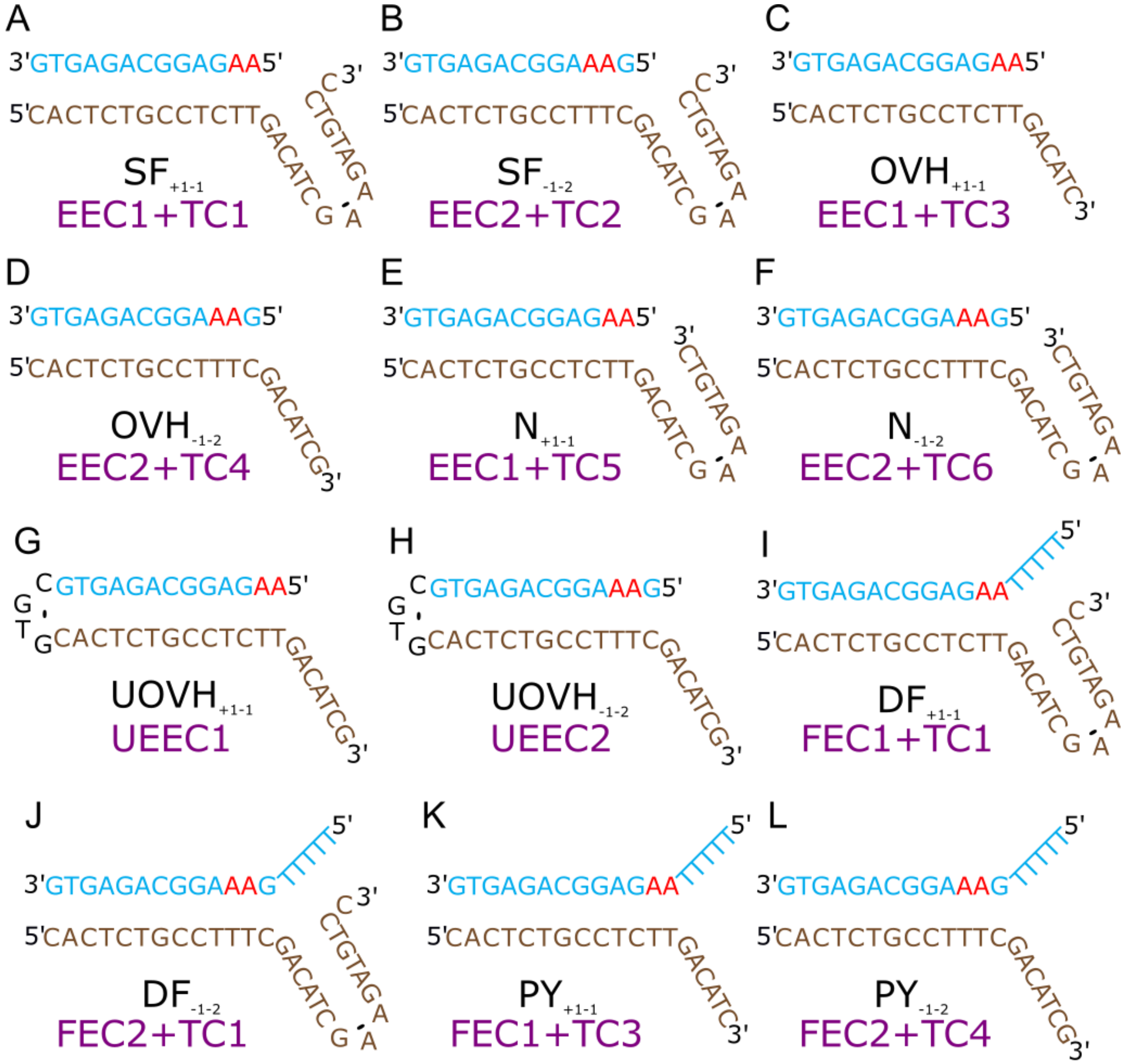


Figure 2.3: Constructs used for ECCD. *Single 3' flapped duplex (SF), 3' overhang (OVH), nicked duplex (ND), unimolecular 3' overhang (UOVH), double flapped duplex (DF) and pseudo-Y (PY) constructs with tandem 2-aminopurine's (in red) in positions +1-1* ***(A****,* ***C****,* ***E****,*

G, ***I***, ***K****) and -1-2 (****B***, ***D***, ***F***, ***H***, ***J***, ***L****) respectively. The reacting strands are highlighted in cyan and the template strands are coloured in brown.*

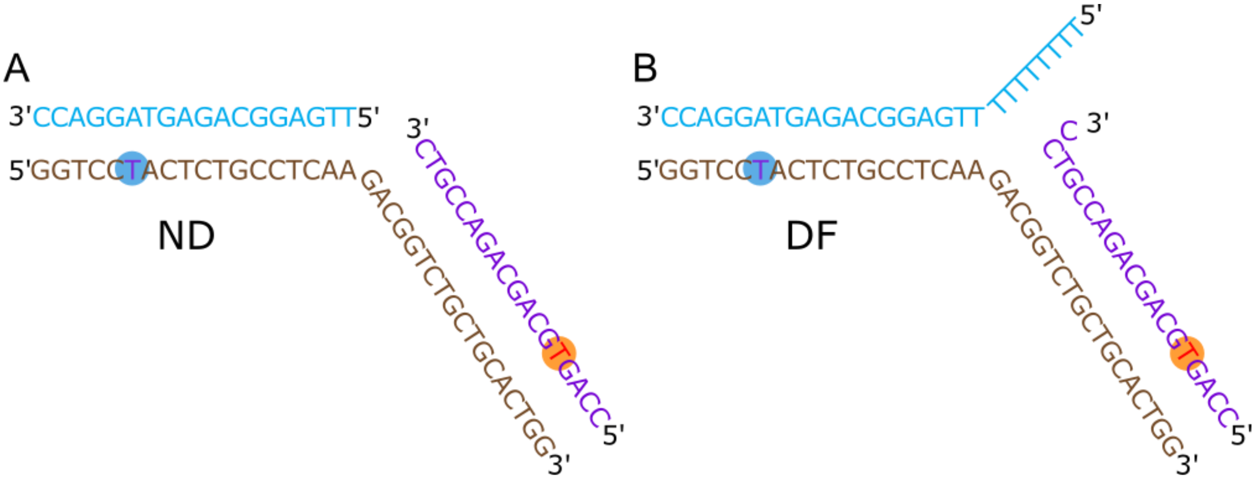


Figure 2.4: Constructs used for FRET. *The reacting strands are highlighted in blue, with the donor and acceptor strands highlighted purple and brown respectively. The nicked duplex (****A****) and double flapped duplex (****B****) with the site for the internal tetramethylrhodamine (purple) on the acceptor and internal fluorescein (red) on the donor. The labels are linked directly to carbon-5 of the thymine base. Presence of the labels is determined by the name; non-labelled (NL) has neither, donor-only labelled (DOL) and acceptor-only labelled (AOL) have the single label represented in the name and donor acceptor labelled (DAL), which contains both. In the absence of a label, the thymine remains.*

2.3.3 Purification of oligonucleotides

Some 3'-fluorescein labelled oligonucleotides required further purification before they could be analysed by kinetics. This step was performed for oligonucleotides E1 and E2 to remove any traces of synthesis intermediates. Initial purification required isolation of the major peak by high pressure liquid chromatography (HPLC) which was run at 50°C to denature the DNA (therefore, dHPLC). Production of 1 M triethylammonium acetate was achieved by gently adjusting a solution of triethylamine to pH 7.0 on ice using acetic acid. This can be used to make the following buffers:

Table 2.3: Oligo HPLC purification buffers

HPLC Buffers	Ingredients
Oligo Purification Buffer A	100 mM triethylammonium acetate pH 7.0, 0.025% acetonitrile
Oligo Purification Buffer B	100 mM triethylammonium acetate pH 7.0, 25% acetonitrile

Retention time (mins)	%B
0.0	5
0.5	5
18.0	50
19.0	100
21.0	100
21.5	5
25.0	5

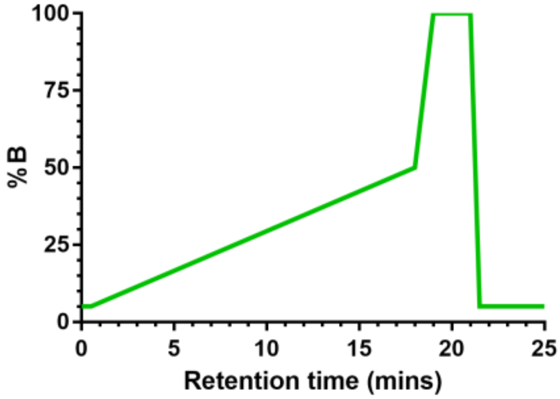


Figure 2.5: HPLC gradient used for oligonucleotide purification. *The gradient used for the dHPLC to isolate the major oligonucleotide product.*

Further purification was performed with a 5 mL DEAE HiTrap fast flow (FF) anion exchange column using the following buffers:

Table 2.4: Oligo Desalting buffers

DEAE Buffers	Ingredients
DEAE Buffer A	10 mM Tris, pH 7.5, 100 mM NaCl, 1 mM EDTA, 0.02% NaN_3
DEAE Buffer B	10 mM Tris, pH 7.5, 1 M NaCl, 1 mM EDTA, 0.02% NaN_3

Oligonucleotides were purified by an isocratic elution with 3 column volumes (CVs) of DEAE buffer B after loading of the sample onto a 5 mL HiTrap DEAE column (GE healthcare) pre-equilibrated with DEAE buffer A. The fractions containing DNA were pooled before desalting into deionised water using NAP-25 columns (Illustra, GE Healthcare) per the manufacturer’s instructions. Purity was determined by capillary electrophoresis.

2.4 Multiple turnover kinetics

2.4.1 Steady-state kinetics

Steady state kinetics were performed with various DNA constructs and an appropriate enzyme concentration to produce between 10-20% product formation over a 20-minute timescale. Reactions were performed to a final concentration of 1×RRB with time points taken manually at 2, 4, 6, 8, 12, 16 & 20 minutes before being quenched in 250 mM EDTA. Samples were analysed by ion-paired reverse-phase denatured HPLC equipped with a fluorescence detector (Wave system, Transgenomic, UK). It should be noted that only experiments using 5' fluorescein-labelled substrates were examined by dHPLC. The chromatograms were integrated to determine the concentration of product formed at each time point (*equation 2.1*).

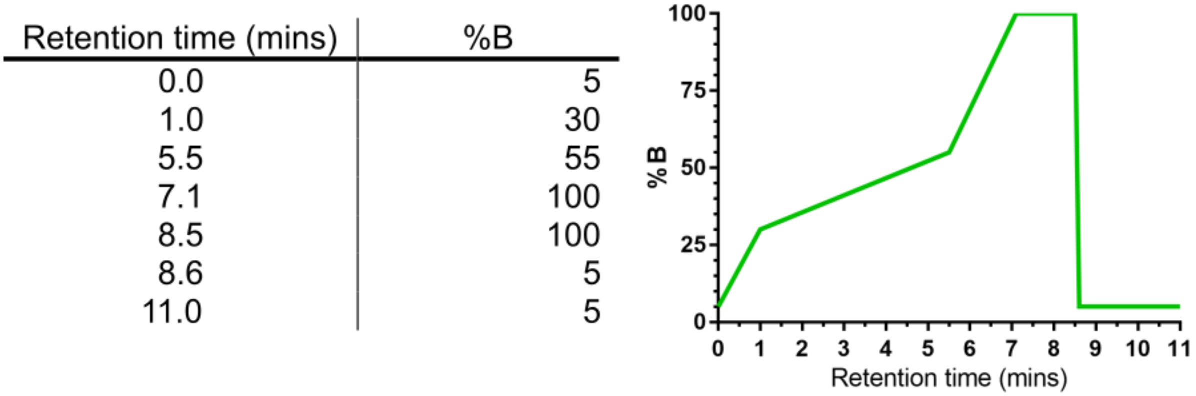
Initial rates of reaction (v, nM min^{-1}) were then obtained using linear regression before conversion into the normalized rates (v/[E], min^{-1}).

Equation 2.1

$$[P] = [S]_0 x \frac{\int P}{\int P + \int S}$$

Table 2.5: Wave dHPLC buffers

dHPLC Buffers	Ingredients
Wave Buffer A	0.1% Acetonitrile, 1 mM EDTA, 2.5 mM tetrabutyl ammonium bromide
Wave Buffer B	70% Acetonitrile, 1 mM EDTA, 2.5 mM tetrabutyl ammonium bromide



Retention time (mins)	%B
0.0	5
1.0	30
5.5	55
7.1	100
8.5	100
8.6	5
11.0	5

Figure 2.6: HPLC wave gradient for kinetic substrates. *The gradient used in dHPLC to separate products and substrate in kinetic reactions, in terms of the percent of wave buffer B.*

2.4.2 Steady state kinetics by capillary electrophoresis

Experiments using constructs labelled with a 3' fluorescein cannot be examined by dHPLC due to the nature of exonucleolytic cleavage, which produces multiple products of similar size. Capillary electrophoresis (CE) is apt at separating nucleotides with single nucleotide resolution and was utilised for this role. Steady state kinetics were performed as before but samples were quenched in a solution of 98% formamide and 20 mM EDTA. Capillary electrophoresis was performed with the P/ACE MDQ Plus system (Beckman Coulter) using the ssDNA 100-R Kit (AB SciEx UK Limited) according to the manufacturer's instructions.

The supplied gel was injected into a 30 cm capillary (internal diameter (ID) of 100 µm) with a 20 cm length to detection window (L_D) using 70 psi for 5 minutes. The gel was then equilibrated between two buffer vials containing the standard Tris-Borate-Urea buffer provided at 3, 5, and 9.3 kV with a ramp time of 0.17 minutes for 2, 2 and 10 minutes respectively. Samples were electrokinetically injected for between 4-8 s, preceded by a 1 s water plug injection of deionised water. Separation of reaction intermediates was performed over 20 minutes at 9.3 kV, applied between the two buffer vials. Samples were analysed at 50 ^{0}C with constant pressure of 40 psi applied to both sides of the capillary. The gel was replaced every 5-7 runs and buffer vials were replaced frequently (the more regularly the replacement, the greater the consistency in retention times). 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 electropherograms were integrated to determine the product concentrations (*equation 2.1*) and normalised rates were determined as above.

2.4.3 Determination of Michaelis-Menten Parameters

Michaelis-Menten parameters were determined by steady state kinetics and analysed by either dHPLC or CE depending on the construct (ND1 and ND2 respectively, see *section 2.3*).

Table 2.6: Concentrations for use in Michaelis-Menten kinetics analysis

[S], nM	ND1 [E], pM	ND2 [E], pM
1000	280	250
750	210	210
500	140	140
250	70	70
150	40	40
100	28	28
75	21	25
50	14	20
25	10	10
20	-	10
15	-	8
12.5	-	7
10	5	5
5	3.5	4
2.5	2	-
1	0.8	-

The normalised rates were determined as mentioned previously and fit to the following equation:

$$v = \frac{V_{max}[S]}{K_m + [S]}$$

$$When\ [S] = K_m\ then\ \frac{v}{[E]} = \frac{k_{cat}}{2}$$

$$v = \frac{V_{max}}{2}$$

$$V_{max} = k_{cat}[E]$$

$$\frac{v}{[E]_0} = \frac{k_{cat}[S]}{K_m + [S]}$$

2.4.4 Observation of reactions by denaturing polyacrylamide gel electrophoresis (PAGE)

Denaturing PAGE gels were constructed in glass plates by the polymerisation of a buffer containing 7 M urea, 1×TBE and 20% acrylamide (19:1). The buffer is initially degassed before the addition of 10% ammonium persulphate (APS), to a final concentration ~0.05%, and N-, N-, N'-, N'- tetramethylethylenediamine (TEMED) to 1/10 the volume of APS, and ~20% acrylamide. Enzyme reactions were quenched 1:1 with denaturing PAGE loading buffer, and heated for 2 minutes at 95°C immediately before loading. Gels were pre-run with 1×TBE, until the gel is at 65°C. Samples were loaded onto the gel and run until the bromophenol blue (the first dye) reaches the bottom of the gel; these gels run optimally at 55°C. The gels were imaged by a Biorad Chemidoc imager using fluorescence from the FAM labels.

2.5 Single turnover kinetics

Single turnover kinetics were performed with excess enzyme concentrations to substrate, and quenched over short time scales. Experiments were performed using a RQF-63 quench flow device (Hi-Tech Sci Ltd., Salisbury, UK).

2.5.1 Determination of k_{ST} for hEXO1-352

Experiments with hEXO1-352 were carried out once a K_m was determined by steady state kinetics for ND1. Experiments were performed with 6.4 nM substrate concentration and either 160 or 680 nM enzyme concentrations in 1×RRB. Samples were initially mixed 1:1 with the substrate and enzyme to a final concentration of 3.2 nM substrate and 80 or 340 nM enzyme (10×K_m and 40×K_m respectively). Mixtures were quenched with QF ST quench in the ratio of 2:1 ([ES]:quench) and analysed by dHPLC equipped with a fluorescence detector. Time points were taken at 9.1, 12.1, 19.4, 27.6, 30.6, 40.8, 41.8, 57.5, 82.1, 124.2, 140.8, 240.8, 440.8, 840.8, 1640.8, 3240.8, 6440.8, 12840.8, 25640.8 & 51240.8 ms and chromatograms were integrated to determine the percentage of product formation. Data was fit to a single exponential non-linear regression in Graphpad Prism using the following equation:

Equation 2.2

$$P_t = P_{\infty}(1 - e^{-kt})$$

Product formation (P_t) per unit time (t) are experimental values, and P_∞ is the amount of product at the end point.

2.5.2 Determination of threading by trapping and blocking with streptavidin

Trapping and blocking experiments were performed using the PY3 substrate. Three types of reaction were performed; a premixed state acting as a streptavidin (SA)-free control, a trapped state where SA traps the substrate onto the enzyme, and a blocked state with the substrate blocked from enzymatic reaction. Reactions were performed with 1-8 μM hEXO1-352 and 10 nM substrate, which were pre-equilibrated in SA reaction buffer containing either 2 mM calcium ions or EDTA.

EXO1 and the substrate were pre-incubated for two minutes at 20°C, for the premixed and trapped states, before the addition of 1×SA-RB (for the premixed state) or five equivalents of SA (for the trapped state) and a further incubation period of five minutes. In the case of the blocked state, the substrate and five equivalents of SA were pre-incubated for five minutes at 20°C before the addition of enzyme and a further 2 minute incubation period. After pre-incubation, reactions were heated to 37°C before initiation with the addition of SA-RB containing 16 mM magnesium ions in the ratio of 1:1.

The reactions were sampled manually (between 15 s and 30 hrs) or by using a RQF-63 quenched flow device (between 9.1 ms and 60 s) (Hi-Tech Sci Ltd., Salisbury, UK). Reactions were quenched 2:1 with 8 M urea containing 300 mM EDTA, pH 8.0. A true zero time point was taken to determine whether reaction occurred before Mg^{2+} addition, which showed only ~1% product formation from pre-incubation. Samples were analysed by dHPLC equipped with a fluorescence detector and chromatograms were integrated to determine the product formation. Data were fit to either a single (*equation 2.2*) or double (*equation 2.3*) exponential non-linear regression in Graphpad Prism.

Equation 2.3

$$P_t = A_1 P_A (1 - e^{-k_1 t}) + A_2 (100 - P_A)(1 - e^{-k_2 t})$$

Product formation (P_t) per unit time (t) are experimental values and P_A is the amount of product at the end point of the first phase. A_1 and A_2 represent the amplitudes of the two equations respectively.

The larger flap of PY3 (see *section 2.3*) required an extended gradient for effective elution with samples containing SA. The following gradient was used:

Retention time (mins)	%B
0.0	5
1.0	30
5.5	55
7.1	100
10.0	100
10.1	5
12.5	5

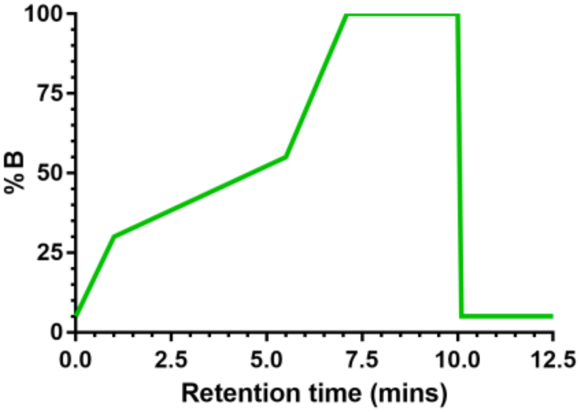


Figure 2.7: Extended dHPLC gradient used for the biotinylated substrate. *The gradient used for the dHPLC to elute all traces of SA, in terms of the percent of wave buffer B.*

2.6 Circular dichroism spectroscopy

Samples were prepared in CDRB with 10 µM substrate and 12.5 µM enzyme where appropriate. Once assembled, CD spectra (300-480 nm) were acquired at 20°C using a Chiralscan Plus spectrophotometer (Applied photophysics). Spectra were taken using a quartz cuvette with a 5 mm pathlength. After spectra of samples containing 10 mM $CaCl_2$ were taken (the standard CDRB), they were supplemented with EDTA to a final concentration of 23.8 mM and another spectra was taken.

CD spectra traces were recorded in 0.75 nm steps (for between 380-480 nm) and 0.5 nm steps (for between 300-380 nm), with 0.5 seconds per step and two spectra being taken. Blanks were taken in the absence of substrate and enzyme to determine a baseline. Spectra were baseline-substracted before gentle correction by smoothing using the Savitsky-Golay filter to a window size of 10. Spectra were converted from CD (mdeg) into molar ellipticity, and entered into an Excel spreadsheet. Data were normalised to account for baseline-shift between 400-480 nm, where no signal from the 2-AP is present. Normalised data were plotted in Graphpad Prism as $\Delta\varepsilon$ per mol 2-AP residue versus the wavelength.

2.7 Förster resonance energy transfer and fluorescence anistropy

2.7.1 Förster resonance energy transfer (FRET)

Measuring the enhancement of fluorescence of the acceptor dye at 37°C allowed for determination of FRET efficiencies by the $(\text{ratio})_A$ method [109]. Samples containing 10 nM of trimolecular NL, DOL or DAL substrates were titrated with increasing concentrations of enzyme and traces were recorded using a Horiba Jobin Yvon FluoroMax-3 fluorometer (Horiba, Ltd). The donor label (FAM) was excited at 490 nm (with a 2 nm slit width) and emission was measured over the range of 515-650 nm (with a 5 nm slit width). For the acceptor label (TAMRA), the sample was excited at 560 nm (with a 2 nm slit width) before measurements of the emission were taken between 575-650 nm (with a 5 nm slit width). 10 nM substrate was made up in FRET RB supplemented with 1 mM DTT; samples were then equilibrated to 37°C by a 10 minute incubation period. The first emission was taken prior to the supplementation of enzyme; enzyme was then added in increasing concentrations with spectra being taken with each step, and corrections made for substrate dilution. The NL sample spectra were subtracted from other spectra, acting as a background for the changing concentrations of buffer and enzyme.

Transfer efficiencies were determined using the following equations:

Equation 2.4

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

Where

Equation 2.5

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

And

Equation 2.6

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

Eqt 2.7 $E = E_{min} + \frac{(E_{max} - E_{min})}{2[S]}\left[([S] + [P] + K_{bend}) - \sqrt{([S] + [P] + K_{bend})^2 - 4[S][P]}\right]$

Where F_{DA} and F_D represent the emission of the DAL and DOL respectively, at their given wavelengths (e.g. $F_{DA}(\lambda^D_{EX}, \lambda^A_{EM})$ represents the emission of the acceptor when the donor label is excited; therefore named DAL). The molar absorption coefficients of the donor and acceptor at the given wavelengths are represented by ε^D and ε^A. The absorbance spectra of the

DAL and AOL are taken to determine the $\varepsilon^{D}(490)/\varepsilon^{A}(560)$ and $\varepsilon^{A}(490)/\varepsilon^{A}(560)$ respectively. The energy transfer efficiency (E) was fitted in the Kaleidagraph program (Synergy software) using a nonlinear regression (*equation 2.7*). The minimum and maximum energy transfer values, E_{min} and E_{max}; [S] and [P] are the substrate and protein concentrations respectively. The bending equilibrium dissociation constant of the protein substrate [PS] complex is represented by K_{bend}.

2.7.2 Fluorescence anisotropy (FA)

The dissociation constant, K_D, was also determined by the Horiba Jobin Yvon FluoroMax-3 fluorometer using fluorescence anisotropy with automatic polarisers. Experiments were performed using the $\mathrm{DOL_{ND}}$ and the 3' ND constructs mentioned previously. The experiments were performed in a similar way to FRET; initially samples with 10 nM substrate in FRET RB at 37°C, with enzyme being titrated into the cuvette and points being taken with increasing concentrations of enzyme. Six measurements of I_{VH}, I_{VV}, I_{HV} and I_{HH} were recorded in 1 minute time courses for each concentration of enzyme, with a control being taken prior to enzyme supplementation. The measurements were used to calculate anisotropy (r) in the following equation:

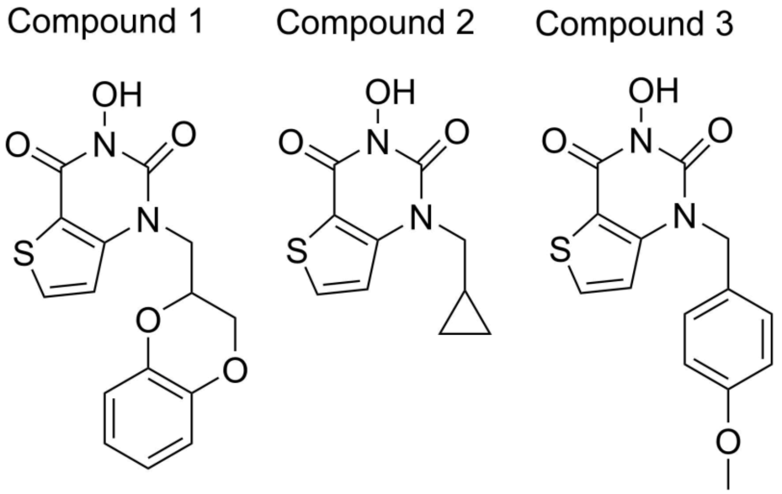
Equation 2.8

$$r = \frac{I_{VV} - G I_{VH}}{I_{VV} - 2G I_{VH}}$$

The intensities of the vertical and horizontal components of the emitted light using vertically polarised excitation is represented by I_{VV} and I_{VH} respectively. G is a correction factor using the inverse intensities of the vertical and horizontal components of the emitted light that is horizontally polarised, I_{HV} and I_{HH}, for the difference in response of the detector to vertical and horizontal polarised light and is expressed by the equation: $G = I_{HV}/I_{HH}$. The total fluorescence, I_{TOT}, in the absence of polarisers was also measured. Data was fit in Kaleidagraph using a similar nonlinear regression to *equation 2.7*. The FRET efficiency (E) is replaced by anisotropy (r), and the K_{bend} constant is replaced by the K_D.

2.8 Inhibitor studies

The compounds used in this section were provided by Astra Zeneca to determine their ability to inhibit FEN1 and their specificity for the enzyme were the project of another student in the Grasby group [110]. The compounds were derivatives of previously identified inhibitors of hFEN1 [111-112] and are shown below:



Compound 1 Compound 2 Compound 3

Figure 2.8: N-hydroxyurea compounds. *The athersys, cyclopropylmethyl and 4-methoxybenzyl N-hydroxyurea compounds, 1, 2 and 3 respectively.*

Studies were performed on hEXO1-352 to determine whether compounds 1 and 2 were specific to hFEN1 or if they affected other family members. The ability for hFEN1 to process the single flap was also tested with compounds 1 and 2. Other kinetic experiments tested the effects of compounds 1 and 3 on proliferating cell nuclear antigen (PCNA) stimulation of hFEN1 with a double flapped substrate.

2.8.1 Multiple turnover studies

Studies were performed as explained in *section 2.4.1* but with the addition of the compounds. The compounds were dissolved in 100% dimethyl sulphoxide (DMSO) and supplemented into the final enzyme dilutions and the reaction mixture in the ratio of 1:100, to a final concentration of 1% DMSO. The compound-free control is supplemented in the same way with 100% DMSO to a final concentration of 1% DMSO. Reactions were performed with 150 nM of SF1 and OVH1 (see *section 2.3*) with hFEN1 and hEXO1-352 respectively; with the concentrations of enzyme and the compounds (1 and 2) shown in *table 2.7*.

Table 2.7: Concentrations of compounds and enzymes used in inhibition experiments

Compound concentration (nM)	hFEN1 [E] (pM) - SF1	hEXO1 [E] (pM)	hFEN1 [E] (pM) - DF2
10000	2500	20000	150
5000	-	-	100
1000	750	2000	70
500	400	-	40
100	150	200	25
50	100	-	20
10	40	-	15
5	-	-	10
0	30	200	8

Experiments to show the effects of compounds 1 and 3 on FEN1 stimulation with substrate DF2 (see *section 2.3*). These experiments were performed by supplementing the reaction mixture with increasing concentrations of PCNA, with a PCNA-free control. The concentrations tested were 0, 0.1, 1 and 10 μM PCNA (it should be noted that this is the trimer concentration). In each case the concentrations of the substrate, compound and enzyme were 50 nM, 10 μM and 150 pM respectively. Samples were run as described in *section 2.4.1*. An IC_{50} was determined for the DF2 with compound 1 prior, to determine enzyme concentrations.

2.8.2 Differential scanning fluorimetry (DSF)

To determine whether an interaction is present between the compounds and the enzymes, differential scanning fluorimetry (DSF) was performed. Thermal shifts were acquired with 2.5 μM of hFEN1 or hEXO1-352 in 1×RRB supplemented with 25 mM EDTA or NaCl with increasing concentrations of compounds 1 or 2. A temperature ramp between 25-95°C was performed, with steps of 1°C taken, and changes in the stability were detected by use of Sypro Orange at 610 nm (excitation 492 nm). The data was converted into the average change in melting temperature (ΔT_m) and plotted against the concentration of the compounds using Graphpad Prism. If the melting temperature increases in the presence of a compound it suggests an interaction is occurring.

Chapter 3: Studies into the Specificity and Kinetic Parameters of Exonuclease 1 as an Exonuclease

3.1 Evaluation of the substrate specificity of hEXO1

3.2 Determination of steady state parameters

3.3 The determination of the single turnover parameters

3.4 The implications of processivity in the mechanism of hEXO1

3.5 Summary of exonuclease studies

Chapter 3: Studies into the Specificity and Kinetic Parameters of Exonuclease 1 as an Exonuclease

Human EXO1-352 (truncated at 352 amino acids) is termed hEXO1 from this point onwards and was expressed via the *lac* operon using autoinduction media [108] in *Escherichia coli*. The bacterial cells were harvested and lysed by freezing at -20 °C. Once defrosted on ice, the lysate was sonicated and centrifuged to separate the soluble and insoluble fractions. The isolated soluble fraction, or supernatant, was first loaded onto a cobalt metal-affinity column to selectively capture the protein by its 6-His tag. The eluate sample was further purified by cation exchange column, isolating the positively charged protein and removing the negatively charged DNA. After removal of the 6-His tag with the TurboTEV protease (Bioscience, Ltd), EXO1 was isolated by removal of TEV from the solution by MagneGST beads (Promega Corporation). Finally, the sample was desalted into 2× storage buffer followed by concentration of the protein to an appropriate concentration and dilution to 1× storage buffer with 50% glycerol and storage at -20 °C, ready for kinetic studies.

3.1 Evaluation of the substrate specificity of hEXO1

In the repair pathways discussed in chapter 1, great diversity in the substrate of exonuclease 1 (EXO1) was observed compared to other superfamily members with both *exo*- and *endo*-nucleolytic capabilities. Briefly, in mismatch repair, the key substrate is a nicked duplex, whereas in double strand break repair, EXO1 cleaves 3' overhang substrates generated by treatment with the MRN complex; or, in the absence of this complex, overexpression of EXO1 has also been shown to cleave blunt-ended duplexes [54]. Therefore, a screen of DNA constructs was performed to determine the preferred exonucleolytic substrate of EXO1. The substrates used in this study are shown below in *figure 3.1*.

Initial tests of hEXO1 were performed using denaturing PAGE to analyse how substrates are processed. Assays with 3'- (OVH2) or 5'-labelled (OVH1) substrates were performed with 150 nM substrate and 1 or 5 nM hEXO1 respectively. Samples were quenched at 0-20 minutes, and run on a 20% gel (*figure 3.2A*). Reactions of the same constructs with 200 pM hEXO1 were analysed by the dHPLC and shown in *figure 3.2B-C*.

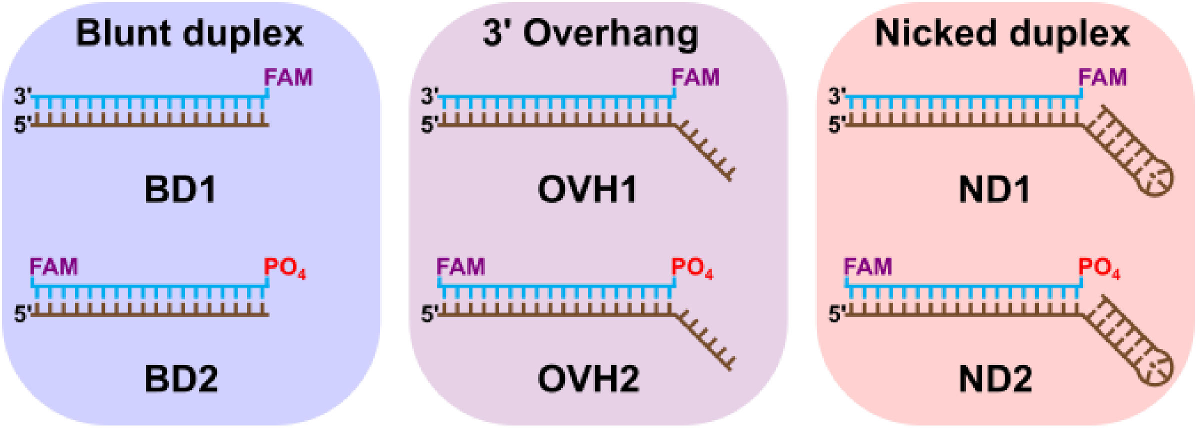


Figure 3.1: DNA constructs. *The various substrates screened with hEXO1 to determine its specificity. With 5' (top) and 3' (bottom) FAM labelled constructs.*

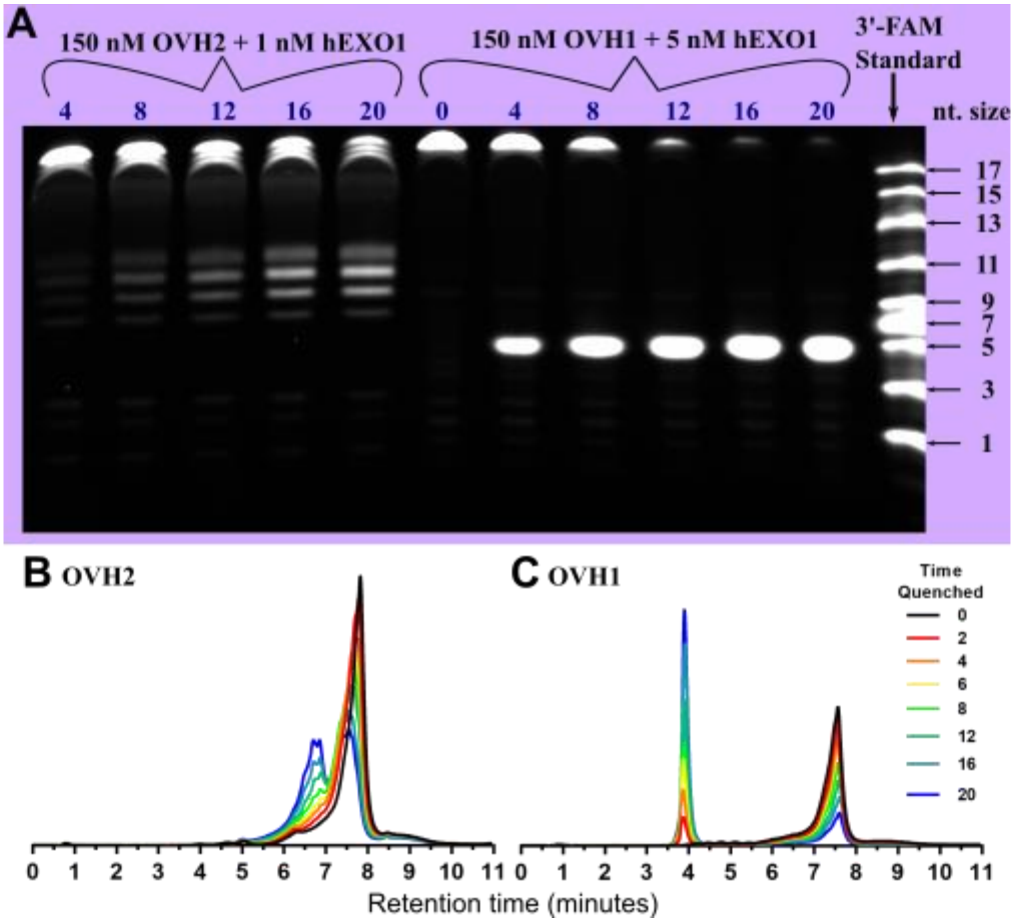
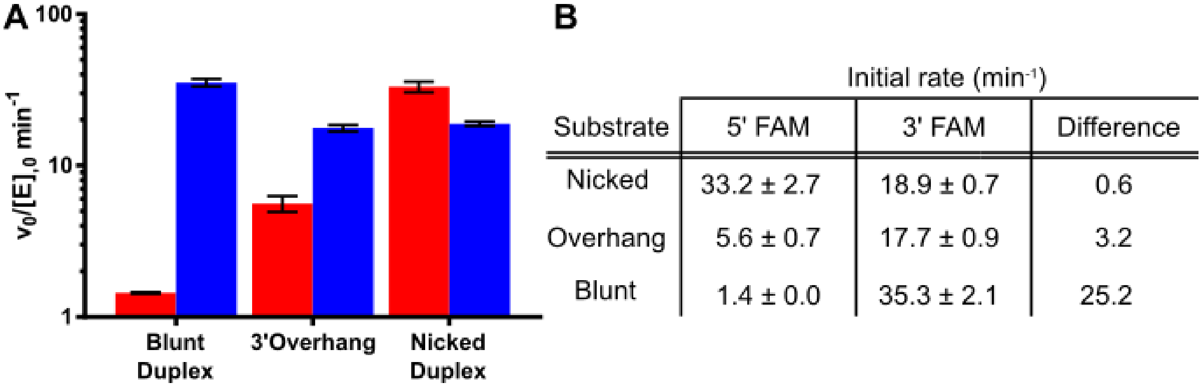


Figure 3.2: Initial kinetic tests of hEXO1 with 5'- and 3'-FAM labelled constructs. *Kinetic assays of a 3' overhang substrates analysed by 20% denaturing PAGE (**A**) or dHPLC equipped with a fluorescence detector (**B-C**). **A.** Lanes 1-5 contain an assay of 1 nM hEXO1 with 150 nM OVH2, and correspond to time points 4, 8, 12, 16 and 20 minutes (blue*

numbering above the lanes). Lanes 6-11 are an assay of 5 nM hEXO1 with 150 nM OVH1, and correspond to time points 0, 4, 8, 12, 16 and 20 minutes (blue numbering above the lanes). The end lane is a 3'-fluorescein (FAM) labelled nucleotide standard containing 1, 3, 5, 7, 9, 11, 13, 15 and 17 nucleotides. It should be noted that 5'-FAM labelled nucleotides travel slower than their respective 3'-FAM labelled alternatives. ***B-C.*** *200 pM hEXO1 with 150 nM of OVH2 (**B**) and OVH1 (**C**), with time points 0, 2, 4, 6, 8, 12, 16 and 20 minutes which are colour coded in the figure. For details on the DNA substrates used see section 2.3.*

Experiments with 150 nM of the blunt duplex (BD1), 3' overhang (OVH1) and nicked duplex (ND1) substrates were assayed with 1000, 200 and 40 pM enzyme respectively. Samples were analysed by dHPLC equipped with a fluorescence detector, and the chromatograms were integrated manually to calculate the initial rate of reaction. However, to determine whether the 5'-fluorescein label acts innocently, meaning the FAM has no effect on the enzymes catalysis, DNA constructs with 3'-fluorescein labels were tested. As before, 150 nM of BD2, OVH2 and ND2 were assayed with 40 pM of enzyme (for details on substrates *section 2.3*). Quenched 3'-labelled samples were analysed by capillary electrophoresis, and integrated manually like with the 5' samples. The results are shown in *figure 3.3*.



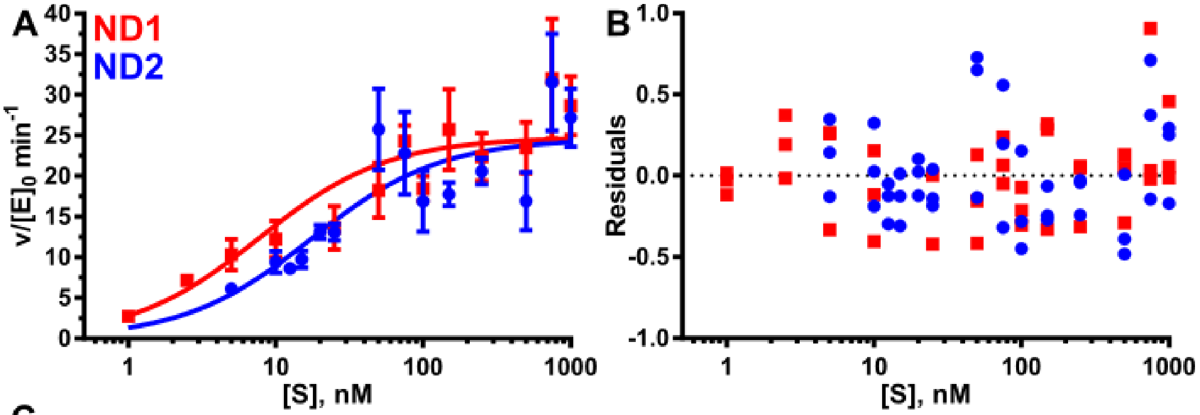
	Initial rate (min^{-1})		
Substrate	5' FAM	3' FAM	Difference
Nicked	33.2 ± 2.7	18.9 ± 0.7	0.6
Overhang	5.6 ± 0.7	17.7 ± 0.9	3.2
Blunt	1.4 ± 0.0	35.3 ± 2.1	25.2

Figure 3.3: Exonucleolytic multiple turnover data. A. *A graph of the multiple turnover data, with both 5' FAM (red) and 3' FAM (blue) respectively.* ***B.*** *The exact numbers plotted in the graph. The fold difference against the 5' FAM is also shown.*

The data show that the nicked duplex displays a similar rate between the two labels (~0.6×), with the overhang substrates demonstrating a 3-fold difference and a 25-fold difference for the blunt duplex. The differences in the nicked duplex and overhang are not significant. However, the observed differences with the blunt duplex labelled substrates may be due to the lack of upstream region in this construct; lack of contacts with the α2-3 hydrophobic wedge may reduce the ability to thread the 5'-FAM into the active site. This hypothesis is only postulated. Due to the consistency of rates between the 3' and 5' FAM labelled nicked duplex substrates, this construct was selected as the preferred substrate and the Michaelis-Menten parameters were determined.

3.2 Determination of the steady state parameters

The steady-state parameters were determined using the ND1 and ND2, over a range of substrate concentrations. Samples were assayed with an appropriate enzyme concentration to produce ~10% product by 10 minutes (exact concentrations are listed in *section 2.4.3*). Constructs with 5'- (ND1) and 3'-FAM (ND2) were analysed by dHPLC and CE respectively. Chromatograms and electropherograms were integrated and converted into normalised rates before fitting by non-linear regression to the Michaelis-Menten equation in Graphpad Prism (GraphPad Software, Inc) (*figure 3.4A*). The residual plot shows no correlation, adding to the validity of the fits (*figure 3.4B*).



C

Enzyme	K_m, nM	k_{cat}, min^{-1}	K_m, M	k_{cat}, s^{-1}	k_{cat} / K_m, M^{-1} s^{-1}
EXO1, ND1	7.8 ± 1.2	24.8 ± 1.5	7.8×10^{-9}	0.41 ± 0.025	5.25×10^{7}
EXO1, ND2	17.5 ± 3.3	24.6 ± 1.8	1.8×10^{-8}	0.41 ± 0.030	2.28×10^{7}
FEN1	19.0 ± 5.0	163.0 ± 9.0	1.9×10^{-8}	2.71 ± 0.083	1.43×10^{9}

Figure 3.4: Determination of the steady state parameters. A. *The Michaelis-Menten fit of the ND1 (red) and ND2 (blue) substrates with hEXO1.* ***B.*** *The residual plot of the two fits, colour scheme as before.* ***C.*** *The parameters determined by the Michaelis-Menten fit (K_m and k_{cat}) and the k_{cat}/K_m, with FEN1 data from a colleagues thesis for comparison [110].*

The normalised rate ($v/[E]_0$) of Michaelis-Menten redefines the equation as follows:

$$\frac{v}{[E]_0} = \frac{k_{cat}[S]}{K_m + [S]}$$

Where the K_m is the substrate concentration needed for half the enzyme to bind the substrate, and therefore, half the enzyme's max rate. The maximal rate constant, otherwise known as the turnover number, defined as k_{cat}, is the reciprocal time (e.g. min^{-1}), and characterises the number of catalytic cycles (or "turnovers") the enzyme can undergo in one unit of time (e.g. one minute). When the K_{m} is equal to the [S], the $v/[E]_0$ becomes equal to half the k_{cat}. Whereas, when the [S] is much higher than the K_{m}, the K_{m} becomes negligible. This results in the $v/[E]_0$ being approximately equal to the k_{cat}. However, the k_{cat} can not truly be reached due to its nature. The $v/[E]_0$ then represents the plateau of the fitted curve.

The K_{m} values for EXO1 (with the 5' and 3' FAM substrates) and also for FEN1 are relatively low and suggest that both enzymes reach their maximal rates at low concentrations of substrate. The turnover number, the k_{cat}, for the 5' and 3' FAM labelled nick constructs with hEXO1 is around 25 min^{-1} , which is 6 times lower than hFEN1 with the double flap substrate. This means that the enzyme does not process its substrate as quickly as hFEN1 and suggests that hFEN1's mechanism is more efficient due to both enzymes having similar K_{m} values.

The specificity constant, k_{A}, which can be represented by $k_{\mathrm{A}} \equiv k_{\mathrm{cat}}/K_{\mathrm{m}}$, represents the reaction at low substrate concentrations. An enzyme is considered to be reaching catalytic perfection if its specificity constant is above 10^8 M^{-1} s^{-1}, and the rate-limiting step becomes substrate diffusion into the active site. In the case of hFEN1 the specificity constant is 1.43×10^9 M^{-1} s^{-1}, which is above this marker for catalytic perfection. This result has been backed up by Brønsted-Lowry analyses with differing groups at 2' of the ribose sugar, which alters the *pKa* of the 3' oxygen which is part of the phosphodiester bond. In the case of hFEN1 Brønsted-Lowry experiments showed no change in the reaction rate with differences in the pKa, which suggests that the chemistry step is not rate limiting, which is also observed for T5FEN [113-114]. Whereas, viscogen studies using sucrose and glycerol showed a reduction in the rate of reaction, which could be due to disruption of the diffusion of substrate into the active site by the presence of a viscous environment. More work may be required, but it provides good evidence that hFEN1 is diffusion-limited [115].

On the other hand, hEXO1 has specificity constants below 10^8 M^{-1} s^{-1}, and is unlikely to be diffusion-limited, although no Brønsted-Lowry analyses or viscogen studies have been performed on the enzyme as of yet. The specificity constant is limited by the K_m, which has evolved to be similar to the substrate concentration that is naturally encountered in the enzymes system. The value of catalytic perfection is used in defining single-substrate reactions; hFEN1 primarily works efficiently on specific flapped substrates and nothing more. Whereas hEXO1 is present in multiple systems and targets a range of substrates, therefore ideas of catalytic perfection may not be broad enough to encompass this enzyme. There is also the possibility of processivity, which will be addressed later in this chapter.

3.3 The mechanism of reaction and determination of the single turnover parameters

The turnover number is a broad number that represents a large part of the reaction mechanism from the enzyme-substrate complex to the product release step. Single turnover kinetics relies on assaying low concentrations of substrate with excess enzyme. The principle of the experiment is for each substrate to be cleaved by a unique enzyme molecule, this means that the single turnover rate (k_{ST}) does not correspond to the product release step. The King-Altman diagram of the reaction mechanism is shown in figure 3.5.

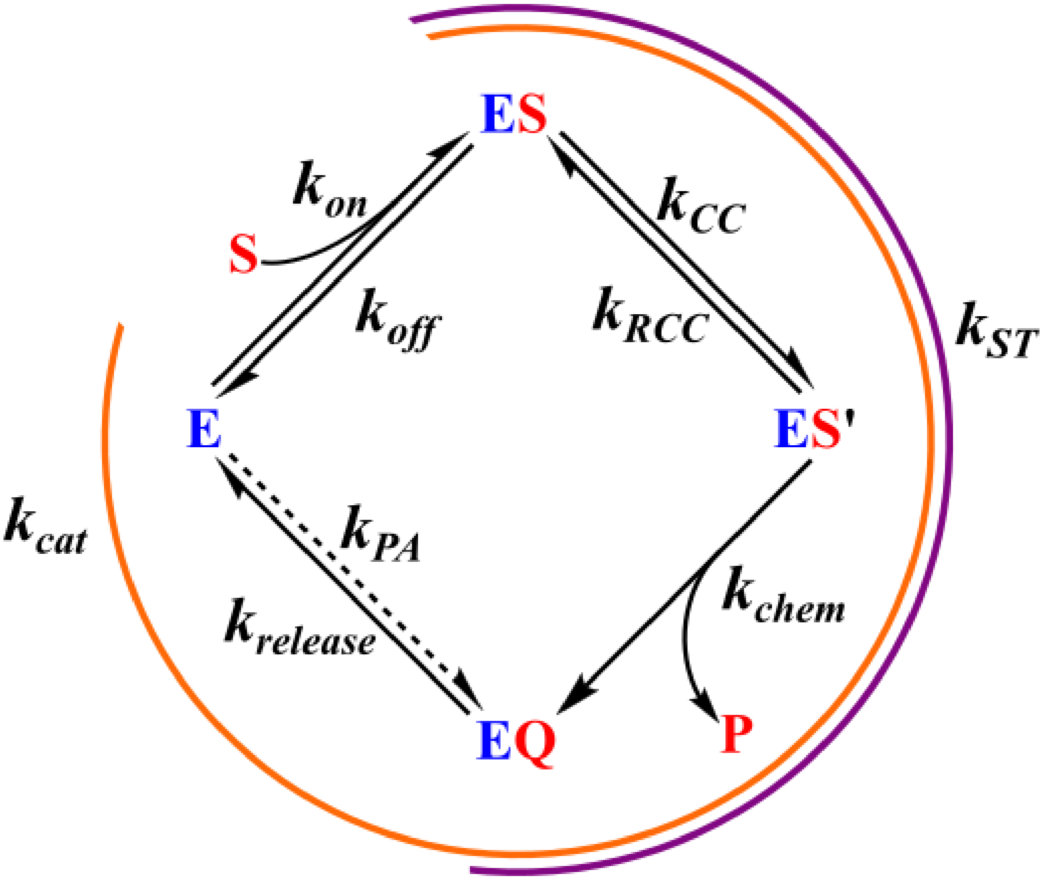
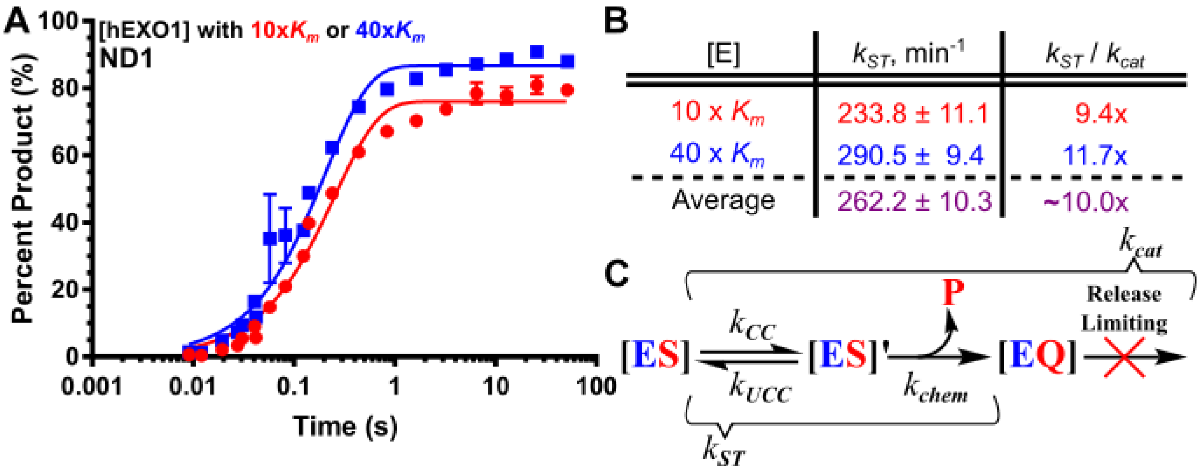


***Figure 3.5: Reaction scheme expected for FEN superfamily members.** The reaction scheme of FEN enzymes starting with enzyme (E) binding to the substrate (S) to form the enzyme-substrate complex (ES) with the substrate binding and release rates (k_{on} and k_{off}). A conformational change is postulated to occur forming ES', with the rate of this change and its reverse (k_{CC} and k_{RCC}), before the substrate is cleaved producing product (P) and creating the enzyme-duplex complex (EQ) with the chemistry rate (k_{chem}) which is unlikely to be reversible. Finally, the rate of product release ($k_{release}$) with the reverse rate corresponding to the product association (k_{PA}) which is more likely to occur with hEXO1 as some products are also substrates. The orange line represents the parts of the mechanism that account for the k_{cat} and the purple line corresponds to the parts of the mechanism that account for the k_{ST}.*

Therefore, single turnover kinetics represents the conformational change and chemistry steps of the reaction scheme. Parameters were determined for 3.2 nM ND1 substrate with either 10× or 40× K_m enzyme concentration (80 and 340 nM) and assayed using quench flow apparatus over multiple time points. Two enzyme concentrations were used to ensure that the substrate was completely saturated. Experiments performed in *figure 3.6A* used the 5' fluorescein-labelled nicked duplex, and were analysed by dHPLC. The resulting data were fit to a single exponential and plotted in *figure 3.6* using equation 3.1. However, the data also fits to a double exponential which may be representative of the conformational change postulated for hEXO1, or due to some other factor.

Equation 3.1 $$P_t = P_\infty(1 - e^{-kt})$$



[E]	k_{ST}, min^{-1}	k_{ST} / k_{cat}
10 x K_m	233.8 ± 11.1	9.4x
40 x K_m	290.5 ± 9.4	11.7x
Average	262.2 ± 10.3	~10.0x

Figure 3.6: The single turnover results. A. *The single exponential fits for single turnover experiments with 10×K$_m$ (red) and 40×K$_m$ (blue) enzyme concentrations with 3.2 nM ND1 (see* section 2.3*).* ***B.*** *The parameters from the logarithmic fits, highlighting that there is a 10-fold difference between the k$_{ST}$ and k$_{cat}$.* ***C.*** *This result suggests that the release step is rate limiting, as it is the only difference between these two rates.*

As highlighted by *figure 3.6B*, the k_{cat} is 10-fold lower than the k_{ST}, therefore the product release step is rate limiting. Due to the products of hEXO1 also being substrates the slower rate of reaction under multiple turnover conditions may in part be a result of the product repositioning into the active site before the next round of cleavage. An enzyme of this nature is considered processive, which will be discussed in the next section.

3.4 The implications of processivity in the mechanism of hEXO1

Processivity is a phenomenon that is observed quite commonly in DNA polymerases, where they synthesise the DNA strand in one continuous process. The definition in terms of a nuclease, is for further available phosphodiester bonds to be processed before the product (which is also a substrate) is dissociated. A reaction mechanism describing this concept is shown in *figure 3.7*.

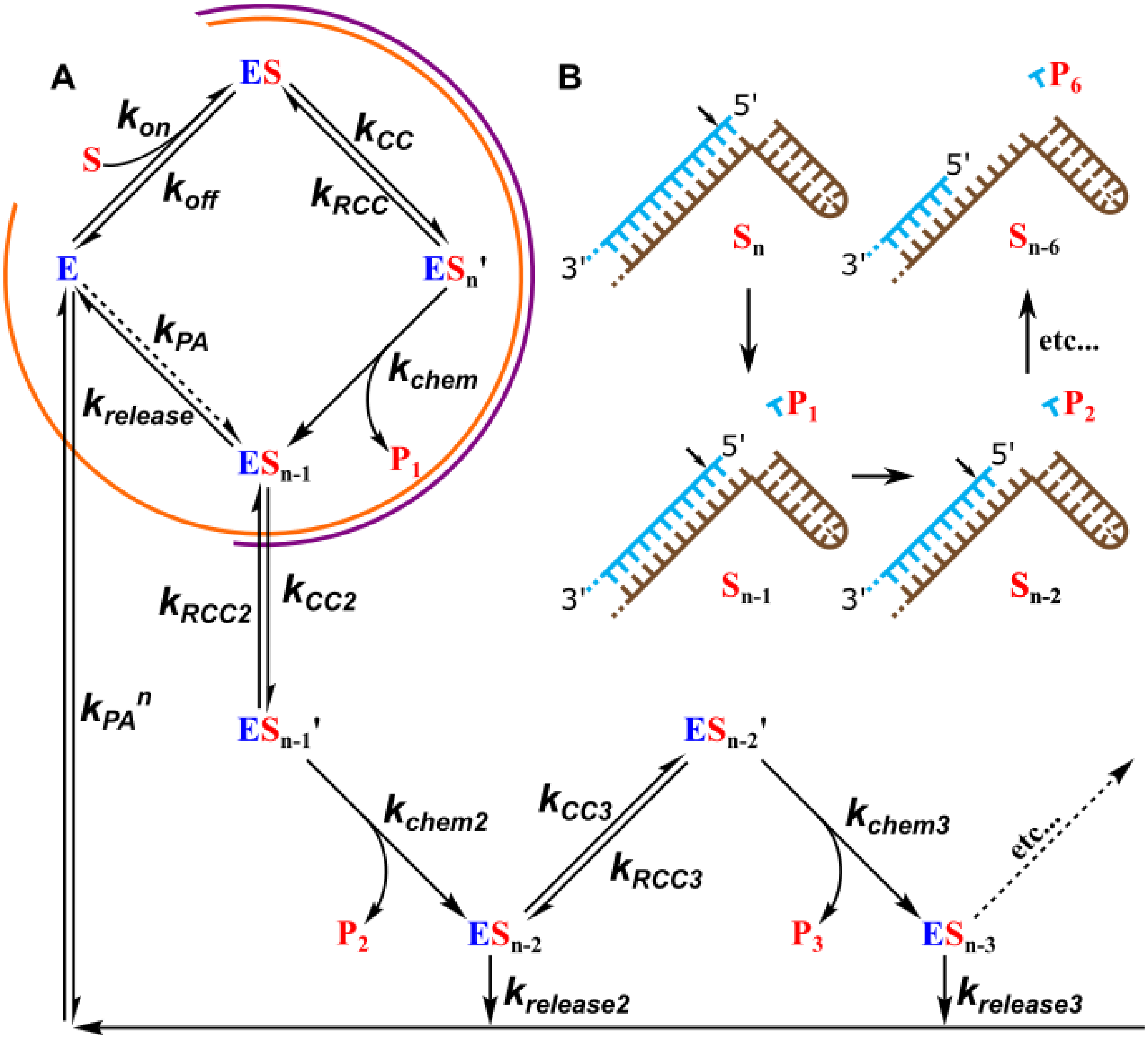


Figure 3.7: An extended reaction scheme to account for processivity. A. *The reaction mechanism diagram with* k_{cat} *and* k_{ST} *highlighted with orange and purple line respectively; this is the standard reaction mechanism. The extended scheme allows for the duplex products to remain bound and continue to be processed, with the ability to be released after every cleavage event depending on the strength of the processive nature of hEXO1. As above, the enzyme is defined by E. The substrate is represented by S with the number of nucleotides that have been*

removed by n (where n is 0 nucleotides removed). The product is simply P with the nucleotide number added (same for the rates). ***B.*** *A substrate schematic representing a nicked duplex after 1, 2 and 6 cleavage events (represented by n for the number of nucleotides). For both* ***A*** *&* ***B*** *this process can be continuous until all cleavable nucleotides are removed.*

In vivo, hEXO1 is a processive enzyme, which can be facilitated by its protein partners [116]. The ladder nature of the 3'-FAM labelled samples in *figure 3.2* suggests some processivity may take place in the absence of protein partners. An efficient processive enzyme would show only two bands on a gel, corresponding to the substrate and product; with every nucleotide being removed before the substrate is released. Samples labelled with the 3'-FAM tested by capillary electrophoresis show multiple bands forming at once, but it is unclear whether the products are preferred over the original substrate. The enzyme is considered to be mildly processive *in vitro*.

The best way to test the hypothesis would be to perform single turnover with the 3' fluorescein labelled nicked duplex. However, this has not been successful for a combination of reasons. The capillary electrophoresis apparatus required to analyse 3' FAM labelled exonucleolytic reactions injects samples electrokinetically. Samples with small amounts of substrate are heavily affected by salt concentrations, which prevent efficient injection. The large enzyme concentrations also have an effect on this injection. Attempts to desalt samples with spin columns showed it to be a viable technique; however, it is not economical with the amount of samples required.

Further experiments with ethanol precipitation have shown some success but it is unclear whether this technique is quantitative of all nucleotide sizes or whether smaller products are lost. It has been shown that the addition of glycogen to ethanol precipitation is effective at providing a quantitative range of products [117]. However, glycogen has an effect on the injection. Another approach where samples were treated with proteinase K has been successful and could provide an effective strategy in the future with further optimisation. No other experiments have been performed to study whether hEXO1 is processive, but they have been considered for future work.

3.5 Summary of exonuclease studies

The preferred substrate of hEXO1 was determined to be the nicked duplex, which gave consistent catalytic parameters with 3' and 5' fluorescein labels. The 3' overhang substrate is also relatively consistent between the two labels, whereas the blunt duplex has an approximately 25-fold difference depending on the position of the label. The nicked duplex was selected and Michaelis-Menten parameters were determined for both the 5' and 3' FAM constructs. The K_m was only 2-fold different between the 5'- (ND1) and 3'-FAM (ND2) (7.9 and 17.5 nM respectively). The k_{cat} is the same for the two labels (~25 min^{-1}).

Studies with the nicked duplex using quenched flow kinetics to determine the single turnover parameters demonstrated that the average k_{ST} is roughly 10-fold higher than the k_{cat}. This suggests that product release is rate limiting, due to single turnover kinetics relying on each substrate being processed by a unique enzyme molecule. Another hypothesis for this difference in rate is the idea of processivity, whereby other phosphodiester bonds are broken prior to product release. This is not measured by the determined k_{ST} as the 5'-FAM labelled construct (ND1) tested only represents the first cleavage. Therefore, continual cleavage of the substrate after the first nucleotide is removed could also be limiting. In any case, the difference between single and multiple turnover reactions implies high affinity for the product despite the question of whether the product then undergoes further reaction being incompletely resolved.

Chapter 4: Mechanistic Studies into how Flapped Substrates are Processed by EXO1 and FEN1

- 4.1 Studies of the endonuclease capability of hEXO1
- 4.2 Determination of hEXO1’s requirement to thread flapped substrates
- 4.3 The importance of conserved basic residues in +1 phosphate interaction in hFEN1
- 4.4 Summary of endonuclease studies

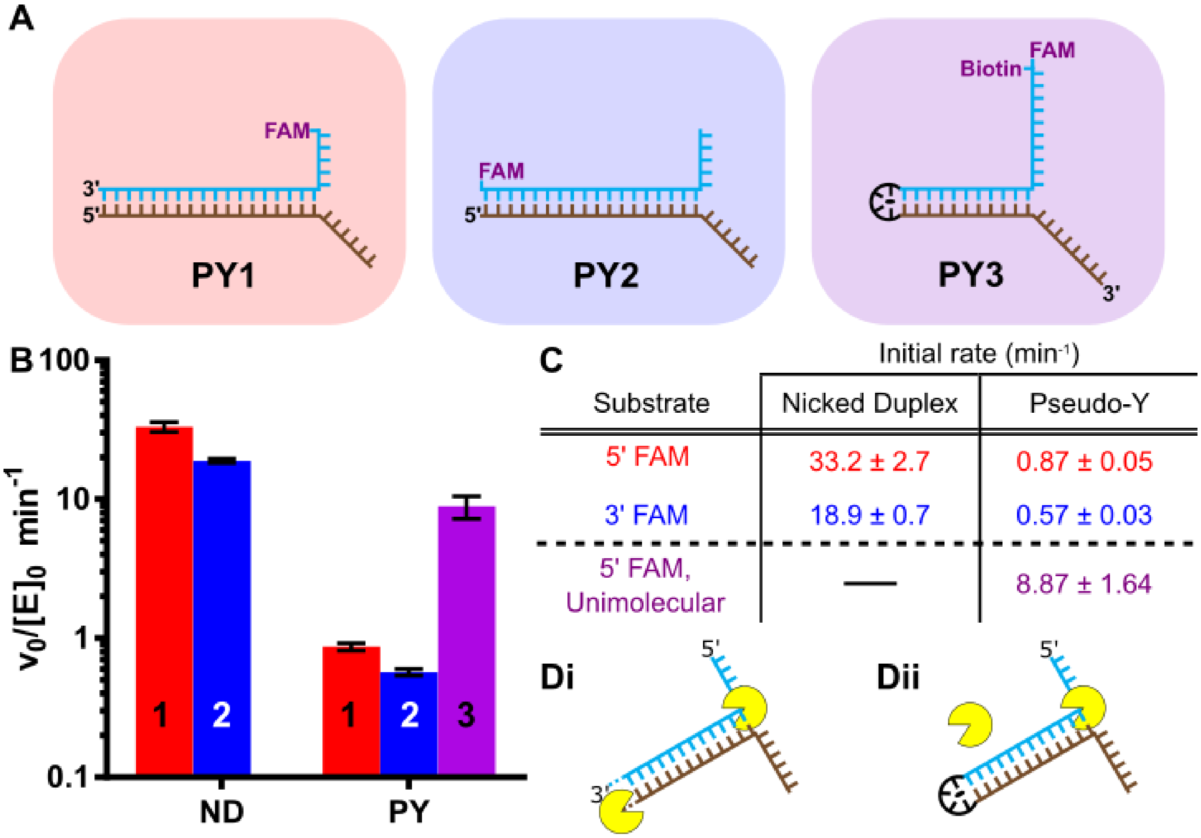
Chapter 4: Mechanistic Studies into how Flapped Substrates are Processed by EXO1 and FEN1

4.1 Studies of the endonuclease capability of hEXO1

The endonucleolytic (*endo*) capability of hEXO1 has been suggested to act as a back-up for hFEN1 in Okazaki fragment maturation, acting as an endonuclease to remove the flapped junctions generated in lagging strand synthesis. The nucleolytic cores of the two enzymes are similar with the major difference between them being a 3' flap binding pocket in FEN1. Without the binding pocket EXO1 may not bind substrates as efficiently; this may explain why FEN1 knockouts are lethal, suggesting that hEXO1 is not perfectly redundant.

To determine whether the *endo* ability of hEXO1 is significant, tests were performed with a flapped pseudo-Y (PY) substrate. Experiments were performed as with the *exo* substrates at 150 nM of the PY1 (5' FAM labelled), PY2 (3' FAM labelled) and PY3 (referred to as the 5' FAM unimolecular substrate in *figure 4.1*) substrates with 0.2, 0.2 and 0.5 nM enzyme concentrations respectively. Due to the size of the endonucleolytic products, all samples could be separated by dHPLC, however, similar rates were obtained when analysing samples by CE. Initial rates were determined by integration of separated peaks before normalisation by the enzyme concentrations. The results are plotted in *figure 4.1* alongside the nicked duplex data.

An interesting observation of the pseudo-Y experiments is that multiple product peaks are formed when assayed with EXO1 (*figure 4.2B*). The nucleotide sizes of these peaks were determined by the addition of nucleotide standards of known sizes in dHPLC. The chromatograms of hEXO1 with the 5'-FAM labelled ND1 and PY1 substrates are shown below. In the case of the nicked duplex, a single product peak forms that localises as a single nucleotide product. Whereas, for the flapped PY, three product peaks form that localise to 4, 5 and 6 nucleotide products. These sizes were determined with 5' FAM labelled nucleotide standards being directly spiked into enzyme reactions. The chromatograms are displayed in *figure 4.2*.



	Initial rate (min^{-1})	
Substrate	Nicked Duplex	Pseudo-Y
5' FAM	33.2 ± 2.7	0.87 ± 0.05
3' FAM	18.9 ± 0.7	0.57 ± 0.03
5' FAM, Unimolecular	—	8.87 ± 1.64

Figure 4.1: Examination of the endonucleolytic capability of hEXO1. A. *A graph of the nicked duplex data previously shown in figure 3.1 alongside the pseudo-Y data. With 5' FAM labelled (red), 3' FAM labelled (blue) and PY3 (purple). No unimolecular nicked duplex has been tested. The graph demonstrates the significant difference between the two constructs, which is removed by use of a unimolecular construct, where only one reaction site is available.* ***B.*** *The initial rates of the nicked duplex and pseudo-Y with the two labels and the unimolecular form (for PY). A 10-fold difference is observed between the 5' FAM labelled unimolecular and bimolecular PY constructs.* ***C.*** *A schematic of the reaction sites between the bimolecular* (***i***) *and unimolecular* (***ii***) *constructs, highlighting that the enzyme can cleave the template strand in the bimolecular substrate.*

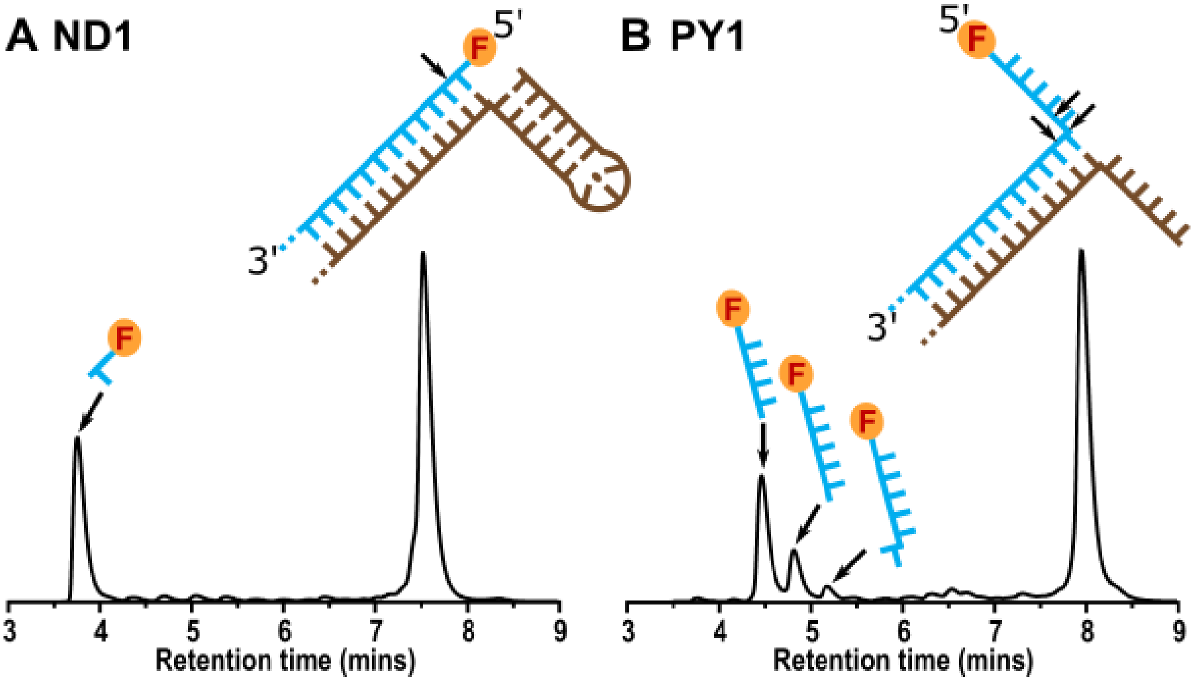


Figure 4.2: The chromatograms of hEXO1 with the ND1 and PY1. A. *ND1 forms a single 18 nucleotide substrate peak and a single nucleotide product peak. The substrate with its possible cleavage site and the product peak are shown.* ***B.*** *PY1 forms three products: 4, 5 and 6 nucleotides from left to right, with the large substrate peak on the right. The substrate therefore has three cleavage sites shown with arrows. These experiments were performed by multiple turnover and analysed by dHPLC.*

Multiple product peaks are expected for *exo* substrates (substrates processed from the end of the DNA strand) with a 3'-FAM label, as each product is a substrate. Similarly, multiple products are observed for *endo* 3'-FAM labelled substrates (substrates cleaved within the DNA strand); for example, multiple products form for the 3'-FAM labelled PY constructs, which are probably due to products being treated exonucleolytically after initial endonucleolytic reaction. However, it is not expected to be visible for the 5'-FAM labelled PY construct. As EXO1 can only cut within the flap it suggests that enzyme does not precisely cleave flapped substrates. A comparison of the exonucleolyic and endonucleolytic data for EXO1 and FEN1 is shown in *figure 4.3*.

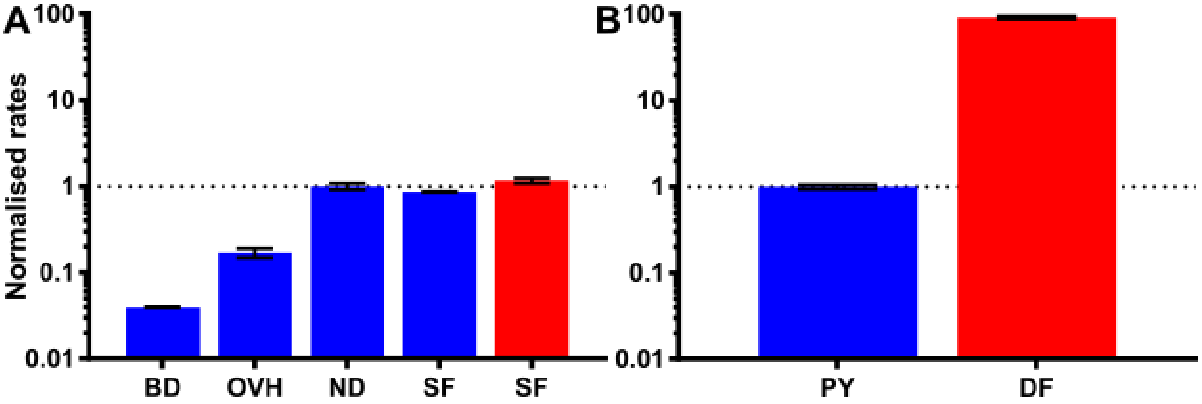


Figure 4.3: A comparison of the reaction mechanisms of hEXO1 and hFEN1. *Rates of the exonuclease (**A**) and endonuclease (**B**) capabilities of hEXO1 (blue) and hFEN1 (red) normalised (normalisation was performed by dividing the rates by the ND1 EXO1 data (figure 3.2) for* exo *constructs or by the PY1 EXO1 data (figure 4.1) for the* endo *constructs). The constructs examined are BD1, OVH1, ND1, SF1, PY1 and DF1 (see section 2.3). The graphs demonstrate that the exonucleolytic capabilities of the two enzymes are similar, but the endonucleolytic capabilities of hFEN1 are much greater than that of hEXO1 (around 90-fold).*

The exonucleolytic and endonucleolytic data for hEXO1 and hFEN1 were normalised by the rate of the nicked duplex data (for the *exo* data – *figure 3.2*) and the pseudo-Y data (for the *endo* data – *figure 4.1*) in *figure 4.3*. Surprisingly the *exo* capability of the two enzymes is relatively equal, with the single 3'-flap being processed by FEN1 with a similar rate to EXO1. Alternatively, the *endo* capability of FEN1 is much higher than that of EXO1, with roughly a 90-fold drop in rate. However, the PY data used in this representation is made up of two separate strands (PY1), and it has been shown in *figure 4.1* that a unimolecular pseudo-Y (PY3) has a 10-fold higher rate. Still, the hFEN1 *endo* rate would be roughly 10-fold higher than that of hEXO1, which can be explained by its preference for *endo* activity. However, the *exo* rate of the two enzymes are relatively equal. Further studies to determine whether the *endo* activity of the two enzymes follows a similar mechanism were then performed.

4.2 Determination of hEXO1's requirement to thread flapped substrates

The endonucleolytic capability of these enzymes are aimed at targeting flapped structures; for example a 5'-flap for EXO1 or a double flapped junction in the case of FEN1. These enzymes endeavour to remove the flap efficiently; for FEN1 this is a precise process whereby a single product is produced one nucleotide into the duplex. However, for EXO1 this is not a specific process as discussed previously. How these enzymes process flapped substrates was formerly an area of controversy, with suggestions of a clamping or threading mechanism of flap alignment. However, as mentioned in the introduction, the true nature of positioning the flap in the active site has been well studied in hFEN1 over the years and the mechanism of threading has been identified as the likely candidate.

Past studies relied on directly trapping a DNA substrate through the helical archway discussed in *section 1.4*. Substrates with a biotin modification at the end of a 5' flap were utilised for this experiment. A tetramer of streptavidin forms a high affinity stable interaction with the biotin molecule; in this case the streptavidin could either block the enzyme from processing the flap, or trap the substrate onto the enzyme depending on the order of addition.

The results showed that for both human and bacteriophage T5FEN homologs bound biotinylated substrate can effectively be trapped on the enzyme by streptavidin if the streptavidin is added after the formation of the enzyme substrate complex; if reaction is initiated this substrate decays very quickly analogously to the reaction of an unmodified substrate. Substrate can also be successfully blocked from threading by adding streptavidin prior to interaction with the FEN protein; this produces a slow reaction rate [101]. This study was backed up with competition experiments, supplementing with an excess of unlabelled substrate after the enzyme is trapped with streptavidin or simply premixed (in the absence of streptavidin). Addition of competitor produced a clear drop in product formation in the premixed state but no significant change in product formation for the trapped state. This proves that the trapped state is in fact trapped, as the labelled substrate cannot dissociate and bind the unlabelled competitor substrate, as a result the original substrate is processed. This study demonstrates that hFEN1 must first thread a flap before it can process the substrate biochemically, however, whether this could be visualised by structural studies was needed to definitively prove this [101].

A crystal structure of hFEN1 threading the 5'-flap of a double flapped substrate through the helical gateway of the enzyme has been obtained more recently [99] (*figure 4.4*). The structure shows that hFEN1 is capable of threading its substrates; together with the threading and blocking experiments this proves that hFEN1 must thread its flapped substrates before catalysis. Whether hEXO1 must also thread its substrates in the same manner is an interesting question. The nucleolytic cores of the two enzymes were shown to be conserved in *section 1.4*. However, given the dramatic difference in the endonucleolytic rates between FEN1 and EXO1 it is unclear whether this mechanism is conserved. Based on previous crystal structures it was not thought possible for hEXO1 to accommodate a flapped structure through its helical gateway [89] (*figure 4.4F*). Therefore, studies into the threading capability of hEXO1 were performed.

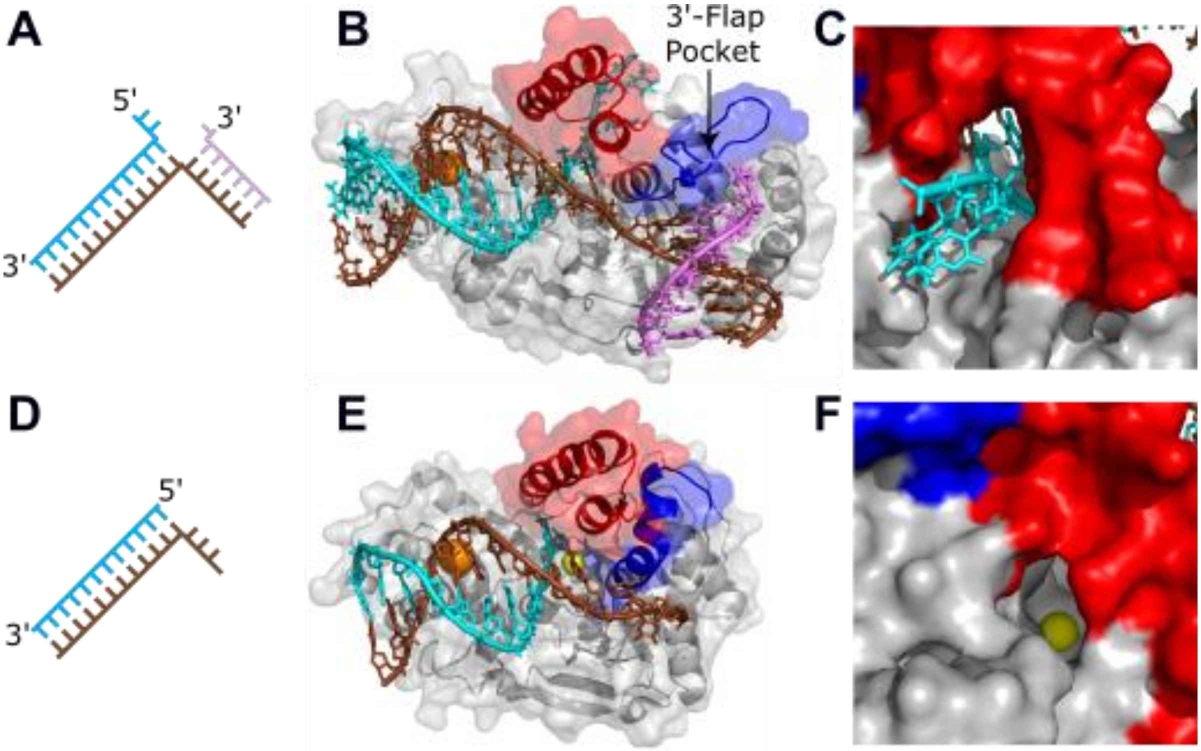
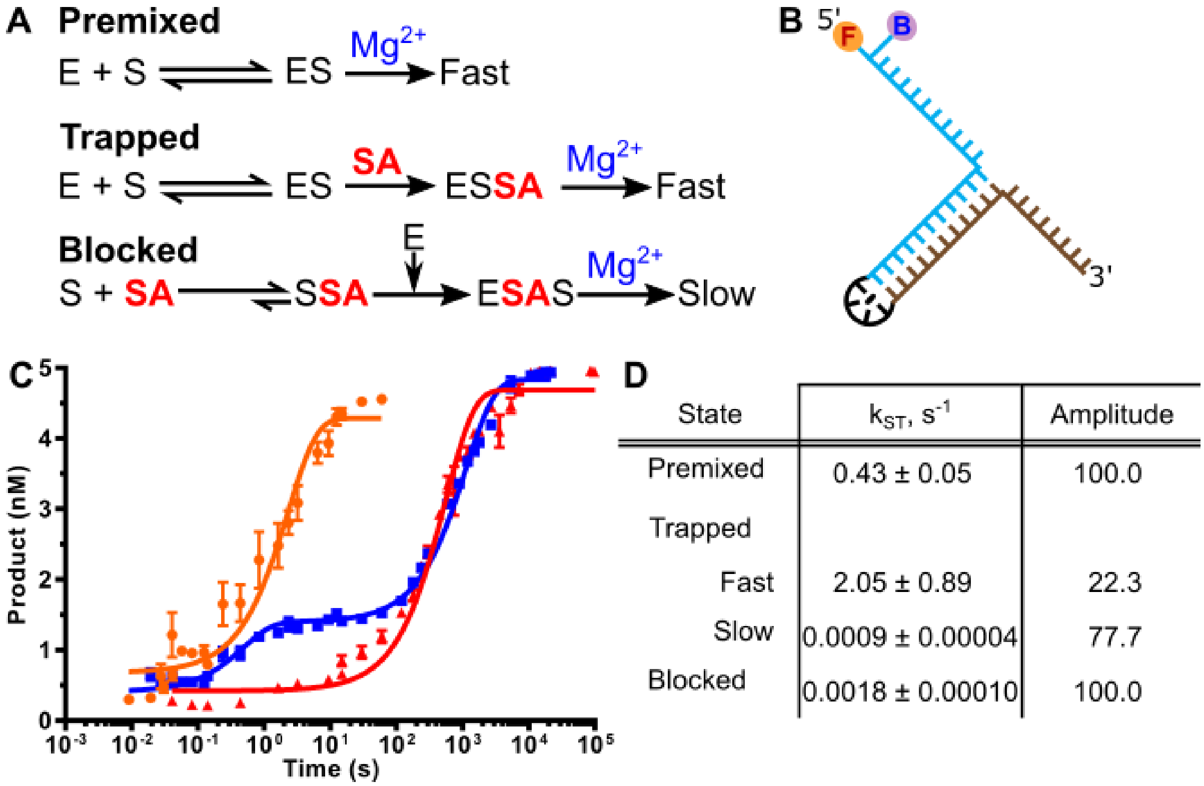


Figure 4.4: Crystal structure of hFEN1 with a threaded substrate and hEXO1 for comparison. *Schematic of the double flapped construct used to crystallise FEN1 (**A**) and EXO1 (**D**), with the colours corresponding to the strands in the crystal. **B.** Crystal structure of hFEN1 (5UM9.pdb) with the substrate in complex with the enzyme, showing the 5' flap threaded below the arch and the 3' flap positioned into the binding pocket. **C.** The back of the archway with the 5' flap threaded through the back of the enzyme without clashes or steric hindrance. **E.** Crystal structure of hEXO1 (3QEB.pdb) with the substrate in complex with the*

enzyme. ***F.*** *The back of the archway of hEXO1 shows a hole that may be too small to accommodate a flap through the helical archway*.

The streptavidin trapping experiment briefly outlined before was utilised to study whether hEXO1 is capable of threading a substrate, and if yes, whether this step is a necessity or a hindrance. There are three states for this experiment; trapped, blocked and premixed (streptavidin-free control). For the trapping and premixed experiments, the substrate was mixed with enzyme and incubated for 2 minutes at room temperature in a buffer containing calcium ions. Calcium ions mimic the catalytic metal ions present *in vivo* (usually magnesium ions) but cannot themselves support catalysis. Depending on the experiment, buffer (for the premixed state) or streptavidin (for the trapped state) was added and a further 5 minute incubation was performed. Alternatively for the blocked state the substrate is first conjugated with streptavidin and incubated for 5 minutes prior to the addition of enzyme and a second incubation of 2 minutes at room temperature. Once pre-incubation had been performed samples were heated to 37°C and initiated with the addition of excess magnesium ions followed by quenching at appropriate time points. The reaction scheme of the three states is shown below in *figure 4.5A.*

Experiments were performed with 10 nM of the PY3 (5' FAM labelled biotinylated unimolecular pseudo-Y – *see section 2.3*) and mixed with 4 μM enzyme during the pre-incubation period, conditions where it is expected that all substrate will be bound by the enzyme. Magnesium ions were added to initiate reaction. The final concentrations after the addition of the magnesium buffer was 5 nM substrate and 2 μM enzyme. Samples were taken at appropriate time points based on the state being examined, for example the premixed state was on a primarily ms timescale as experiments were performed using a quenched flow apparatus, whereas the blocked state experiments were manually run until 30 hours. Samples were analysed to dHPLC as described in *sections 2.4.1* and *2.5.2* with processing via *equation 2.1*). The data were fit to either a single or double exponential equation (*equation 2.2 and 2.3*) and plotted in *figure 4.5C-D* with the exact rates listed.



State	k_{ST}, s^{-1}	Amplitude
Premixed	0.43 ± 0.05	100.0
Trapped		
Fast	2.05 ± 0.89	22.3
Slow	0.0009 ± 0.00004	77.7
Blocked	0.0018 ± 0.00010	100.0

Figure 4.5: The expected trapping scheme and data. A. *The premixed (streptavidin-free), trapped and blocked states outlined based on the supplementation order. The streptavidin, SA, is highlighted in red, and the magnesium buffer is indicated by* Mg^{2+} *in blue. This is the expected mechanism for trapping and blocking. The ESSA complex represents the trapped state and ESAS complex is the blocked state.* ***B.*** *The PY3 construct used for the streptavidin threading experiments with hEXO1 (see* section 2.3*).* ***C.*** *The plot of the threading data, with premixed (orange), trapped (blue) and blocked (red), data fit to single or double exponential fits (equations 2.1 and 2.2).* ***D.*** *The rates and amplitudes of the fits; the trapped fit is biphasic, hence the two separate rates.*

Upon addition of magnesium, the premixed (streptavidin-free) complex decayed relatively quickly with a reaction rate of 0.43 s^{-1}. The data for the blocked state, where streptavidin would prevent threading, are slow as expected, with a biologically irrelevant rate of 0.0018 s^{-1}. However, the trapped data produced an unexpected result. Roughly 20% of the product formed with a fast rate which is around 5-fold faster than for the premixed rate. Whereas, the other 80% of product formation was a much slower rate which resembled the blocked data (a 2-fold difference). The trapped fit illustrates that around 20% of the substrate is successfully threaded and the other 80% is effectively blocked, hence the slow second phase.

To determine whether a lack of saturation of the substrate with enzyme was the cause of this fast phase, various concentrations of enzyme were tested in the trapped state (*figure 4.6A*). However, the concentrations tested showed no significant difference in proportion of fast decaying species with increasing enzyme concentration, suggesting that the substrate is saturated by the enzyme in all cases. This suggests an alternative threading mechanism, with an on-enzyme equilibrium, whereby the substrate can be bound to the enzyme but not threaded. The new mechanism explains the trapped state; it may also be significant in the premixed state, as these data fit a two phase exponential (*figure 4.6C*). A control to determine whether the streptavidin concentration of 5 equivalents is enough to proficiently trap the substrate was performed using 15 equivalents (*figure 4.6D*).

A biphasic fit of the premixed data in *figure 4.5* fits the plotted data points with greater accuracy. This double exponential decay could be due to the threading equilibrium hypothesised from the saturation experiment. If the enzyme must first bind and then thread a substrate, this could account for the biphasic nature of the premixed and trapped states. In the premixed state more substrate would be continuously threaded as the reaction is taking place. However, upon addition of the streptavidin the substrate can become blocked before it is rebound and threaded, hence the small amount of trapped substrate in the experiment. All these experiments suggest that hEXO1 must first thread a flapped substrate to cleave it on a biologically relevant timescale.

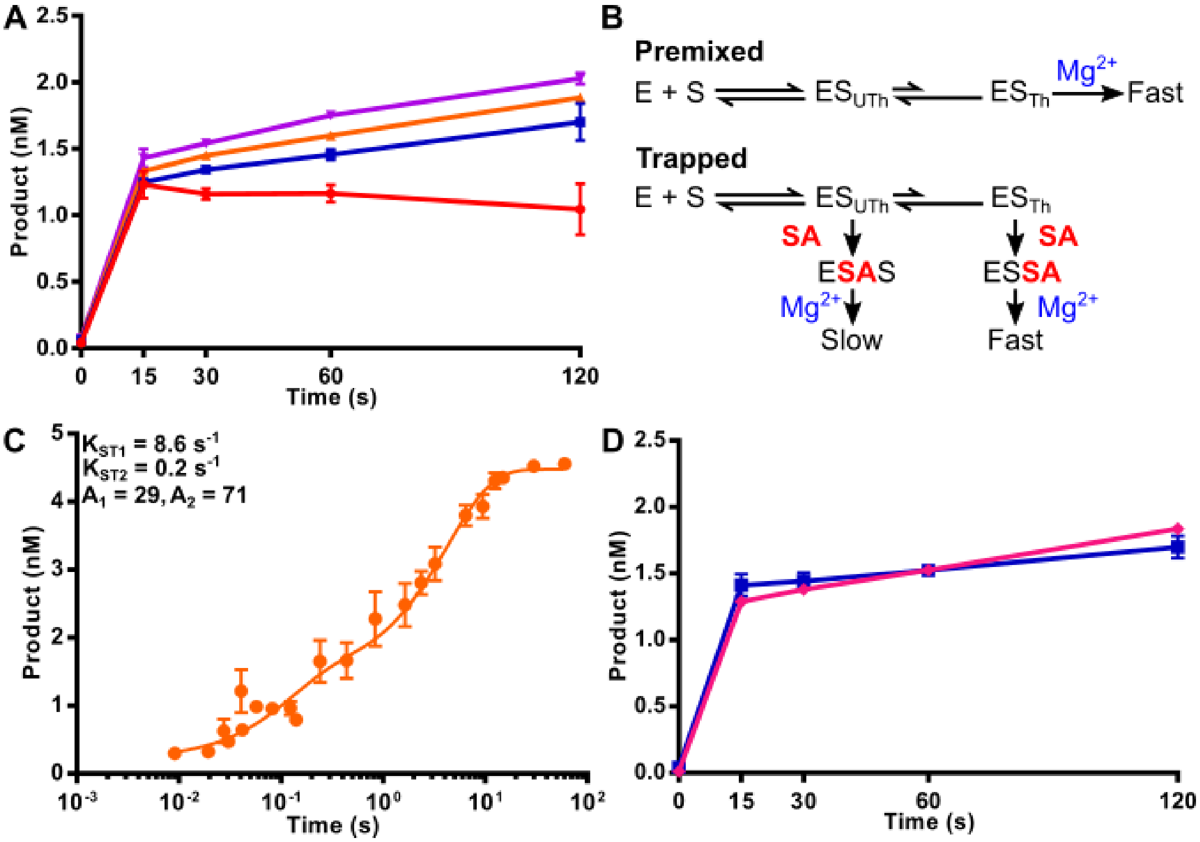


Figure 4.6: Saturation and streptavidin control tests. A. *Substrate saturation test with increasing enzyme concentrations where 0.5, 1, 2 & 4 µM correspond to the red, blue, yellow and magenta lines respectively (these are the final enzyme concentrations after initiation). No significant difference in the trapped state is observed by increasing the enzyme concentrations.* ***B.*** *The modified reaction scheme for the premixed and trapped states, where EXO1 must first thread the flap by an on enzyme equilibrium. This equilibrium is between the unthreaded (UTh) and threaded (Th) state. In the trapped state this leads to the substrate being blocked if the flap is not threaded.* ***C.*** *The premixed data fit to a double exponential equation, which may be more representative of this on-enzyme equilibrium, as more substrate becomes correctly positioned over time.* ***D.*** *A test with streptavidin to determine whether increasing the streptavidin concentration can increase the amount of substrate trapped. The standard concentration used, 5 equivalents, in blue and 15 equivalents in magenta show no difference.*

A series of competition experiments were performed to complement the premixed and trapped data. Supplementation with a large excess of unlabelled 3' overhang competitor (OVH3) after pre-incubation with enzyme and buffer (premixed) or streptavidin (trapped) was followed by a further 10 minute incubation. After the addition of magnesium ions, the reaction was sampled five times at points corresponding to the fast phase of trapped and pre-mixed states using quenched flow apparatus. The experiment reveals whether the competitor can outcompete the substrate; the non-labelled competitor substrate was 1000-fold more concentrated than the labelled biotinylated substrate. In the premixed state, the substrate was expected to be outcompeted. Whereas a substrate that has been successfully trapped on the enzyme cannot dissociate, preventing the competitor from binding (*figure 4.7A*). Therefore, the rate of product formation of the premixed state should be drastically reduced, whereas the trapped experiment would remain unchanged in the fast phase.

The competition data agreed with the conclusions of the earlier streptavidin experiments. The streptavidin-free control product formation rate is reduced by the addition of the competitor (*figure 4.7B*). Whereas, the trapped state is unaffected by the addition of the competitor, which demonstrated that the substrate was indeed trapped on the enzyme, and could not dissociate (*figure 4.7C*). No competition experiment was performed with the blocked state as the rate is already not biologically relevant, and the competitor would just slow this rate further. This effect may also occur with the trapped slow phase, but time points taken were specifically tailored for the fast phase of the experiment. Therefore, the fast decaying substrate portion is being threaded through the enzyme.

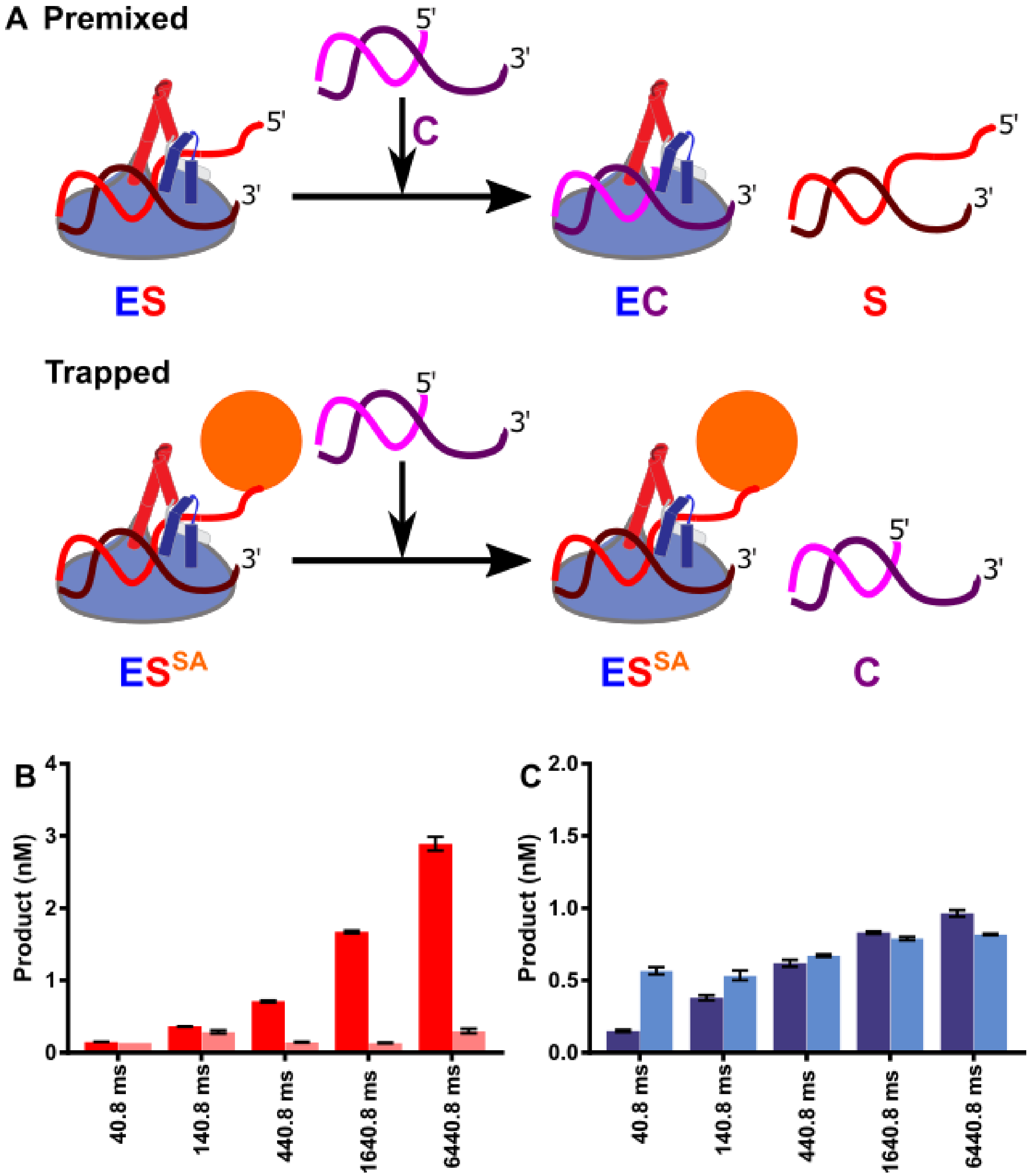


Figure 4.7: Threading competition experiments. A. *A cartoon of the predicted competitor scheme. The addition of the competitor is expected to fill the enzymes active sites before catalysis initiation with the addition of Mg*$^{2+}$*, unless the substrate is successfully trapped. For the premixed (**B**) and trapped (**C**) experiments, 5 time points were selected over the course of the trapped fast phase (41, 141, 441, 1641 & 6441 ms). The premixed data are represented in red, where the lighter pink colour represents the competition experiment, and the red is the competitor-free control. The trapped data is depicted in blue, where the lighter shade is the*

competitor data, and the darker shade is the competitor free-control. The premixed state has reduced product formation in the presence of the competitor, whereas the trapped phase remains the same over the fast phase.

All the data presented in this section suggest that hEXO1 does indeed thread flapped substrates in a mechanism analogous to hFEN1, but with a lower efficiency. Another recent study using time-lapsed crystallography successfully crystallised stepwise snapshots of the enzyme in complex with a flapped substrate [100]. The enzyme was crystallised with a pseudo-Y construct which was threaded below the helical arch. It should be noted that these crystal snapshots were not successful threaded structures the majority of the time, representative of the on-enzyme threading equilibrium. The DNA construct used in both the threaded and cleaved crystal structures are visualised in *figure 4.8.*

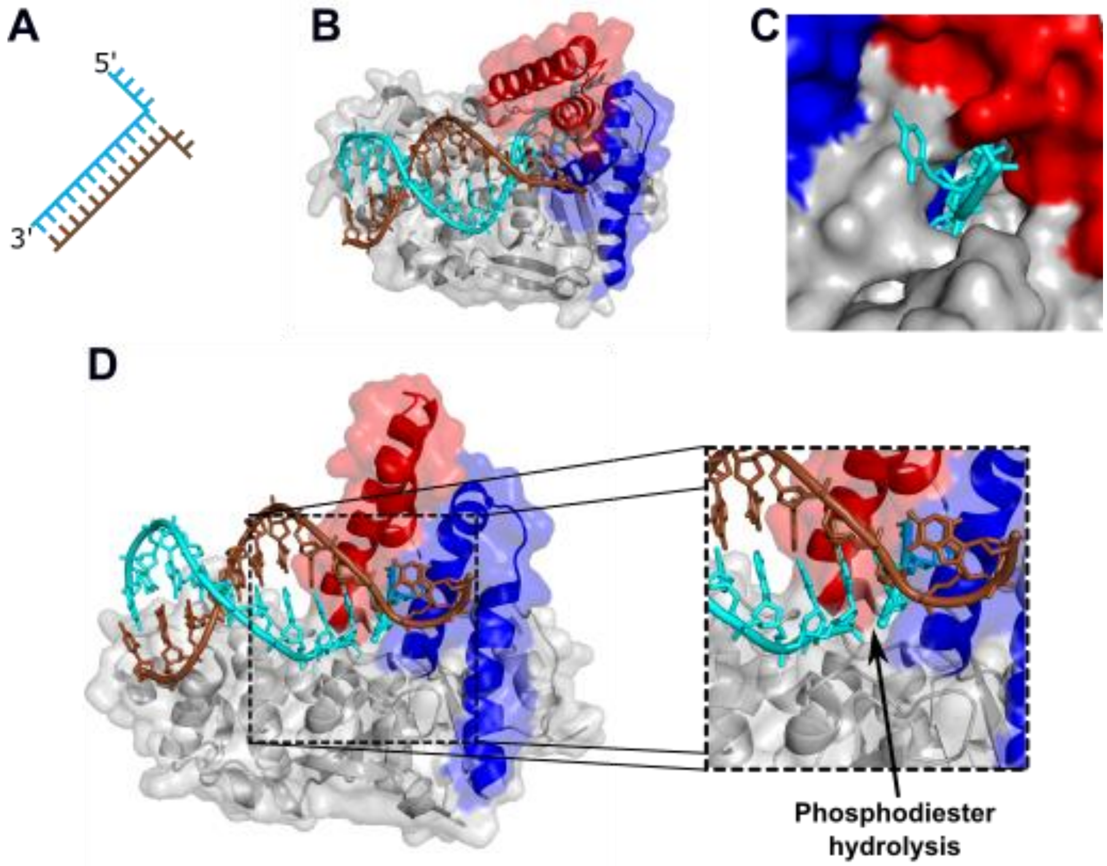


Figure 4.8: EXO1 Threading crystal structures. *Time-resolved crystal snapshots of hEXO1 in the presence of flapped substrates show hEXO1 threading its substrate, like hFEN1 in figure 4.4.* ***A.*** *The construct used to crystallise the threaded state is a small pseudo-Y construct.* ***B.*** *The threaded crystal from the front shows that the DNA substrate is accommodated below the*

helical cap. ***C.*** *The back of the enzyme lacks the steric clashes hypothesised to block a flapped substrate.* ***D.*** *A crystal showing the threaded state after phosphodiester hydrolysis, which is highlighted on the right. For all crystal representations, the helical gateway is highlighted in red and the hydrophobic wedge in blue. The crystals represented in* ***B-C****. 5V0E.pdb and* ***D.*** *5V0D.pdb.*

Therefore, both biochemical and structural studies show that hEXO1 is not only capable of threading flapped substrates, but must do so to process these DNA constructs. This is comparable to the mechanism by which hFEN1 processes flapped substrates. However, the reduced specificity of hEXO1 threading its substrate may account for the lower rate observed with a flapped substrate compared to its exonucleolytic substrates. It is unclear whether differences in the active site reduce the accommodation of a flap, or if this is due to the exonucleolytic processive nature of hEXO1. This will be further examined in the next section.

4.3 The importance of conserved basic residues in +1 phosphate interaction in hFEN1

As discussed previously, both hFEN1 and hEXO1 must first thread flapped structures to process them on a biologically relevant time scale. The next area of research was aimed at studying how the enzyme interacts with the +1 phosphate group and its importance in active site positioning of the scissile phosphate. A simple diagram of the DNA is shown below in *figure 4.9* which highlights the +1 phosphate and sugar, with the cut site represented by a blue dotted line.

Residues of possible importance for +1 phosphate interaction were selected by their crystal contacts with the DNA bound to hFEN1 (3Q8K.pdb). The following work focused on hFEN1 mutants of various basic residues to alanine as single or double mutants in an attempt to prevent phosphate interactions. As phosphates are negatively charged, the positively charged basic residues form a complementary electrostatic environment surrounding them. The basic residues were also selected due to good sequence conservation (*figure 4.10*) across species and they are close to various phosphate oxygens.

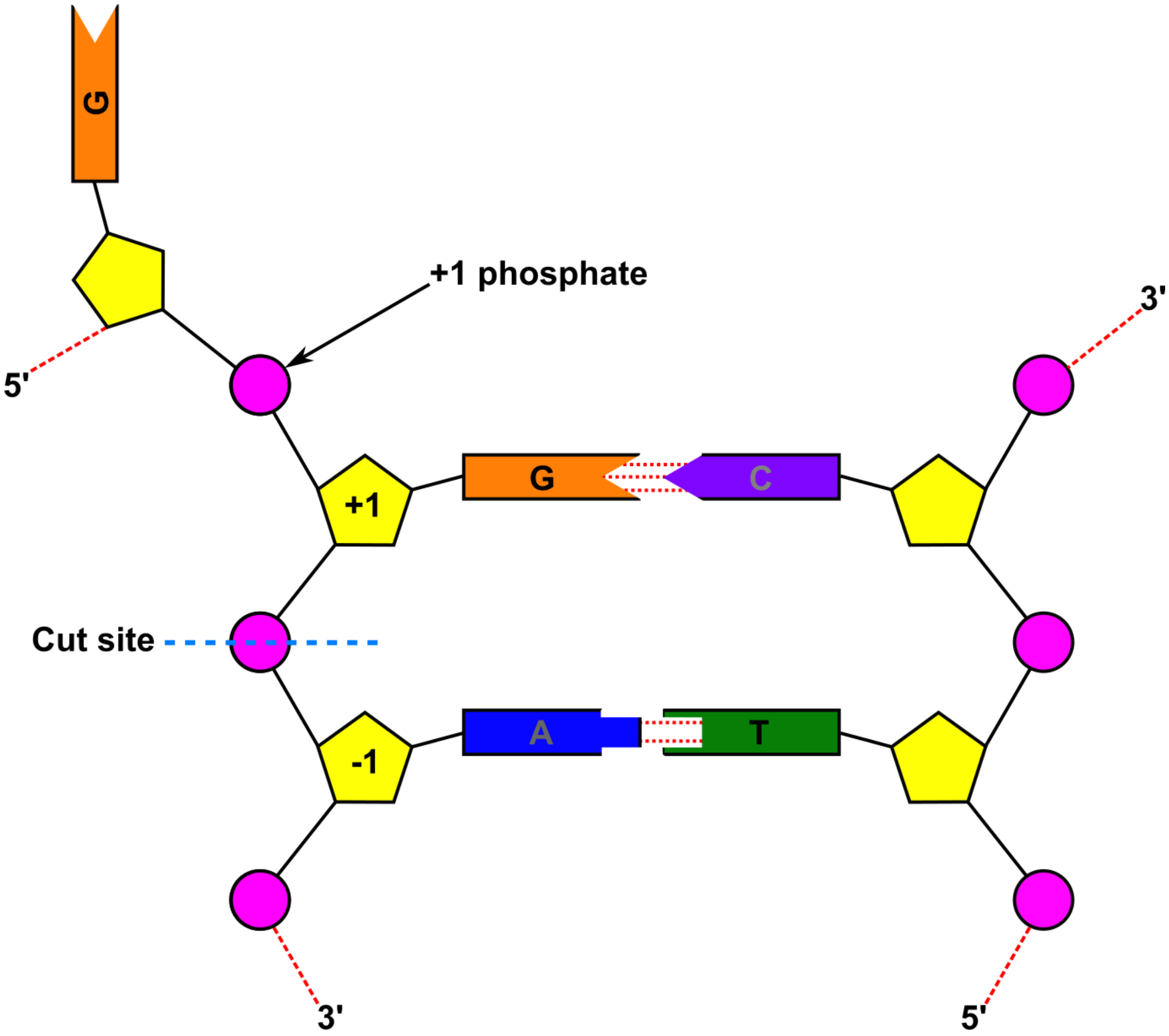


Figure 4.9: A simple representation of the DNA. *A double-stranded DNA duplex with ssDNA 5' flap. Simple nucleotide representations with four nucleobases (orange, purple, blue and green), deoxyribose (yellow pentagon) and phosphate molecules (magenta circles). The cut site is one nucleotide into the duplex represented by the dotted blue line. The +1 and -1 nucleotides are highlighted, but the +1 phosphate is the main research topic for this section.*

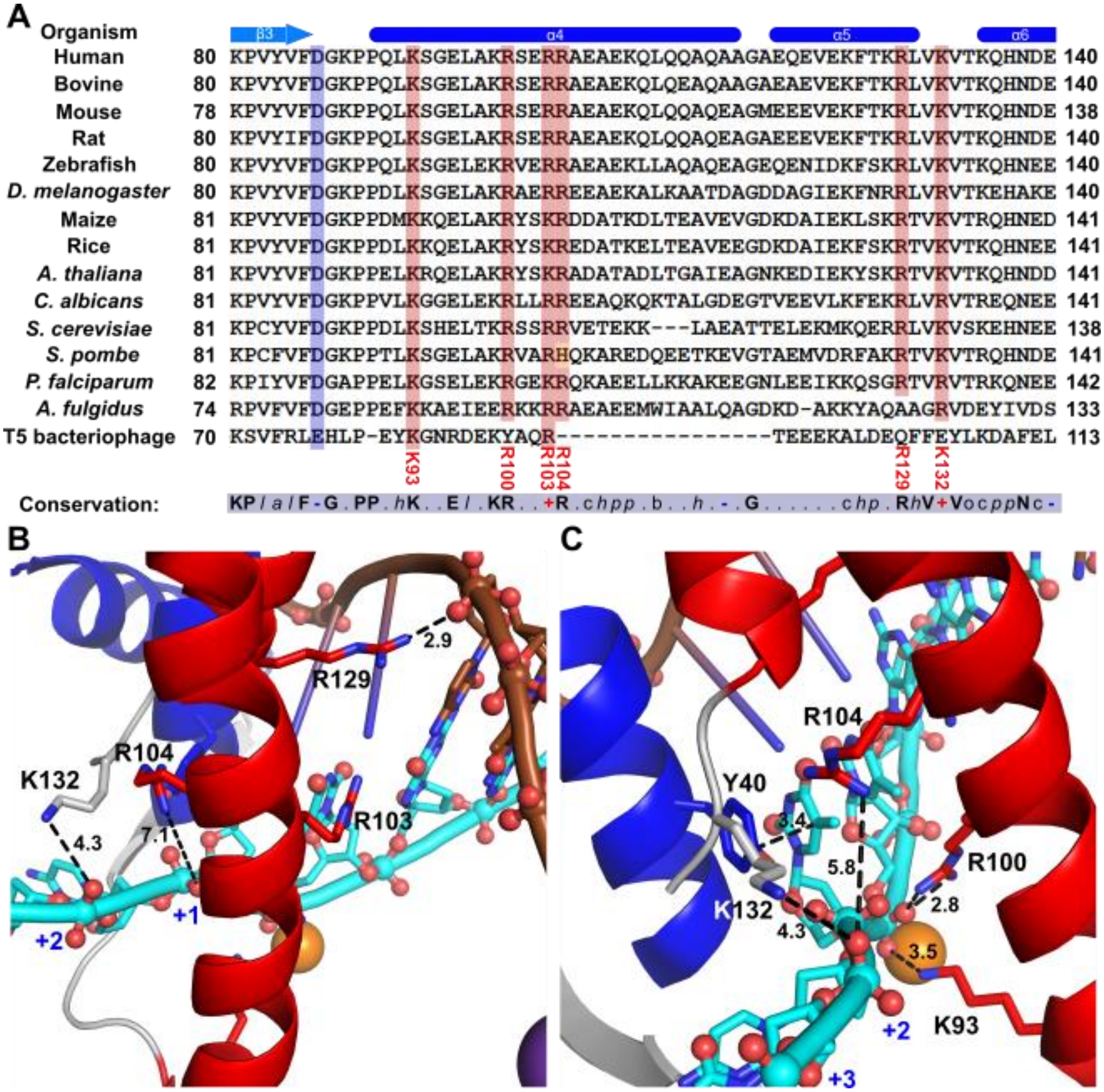


Figure 4.10: Selection of residues for +1 phosphate experiments. A. *A sequence alignment of FEN1 homologs from 15 species aligned using ClustalX2. The region examined corresponds to the helical arch and cap. The better conserved basic residues are shown below the alignment in red; R100 and K93 were not mutated in this study. Crystal structure representations were generated using the D86N mutant crystal (5UM9.pdb).* ***B.*** *A side view of the archway with the DNA flap threaded through; K132 and R104 are shown with their respective distances to phosphate oxygens. K132 is interacting with the +2 phosphate and the R104 residue with the +1 phosphate. R103 is also shown but no visible electrostatic interactions are observed. Whereas R129 is proposed to interact with the DNA template strand and is within electrostatic range.* ***C.*** *A back view of the threaded flap through the archway,*

with Y40 stacking on a nucleobase. K132 and R104 are illustrated again, with R100 and K93 also visible, the latter residues are well known to be essential active site residues. A magnesium ion is represented by an orange sphere.

Of the residues illustrated in *figure 4.10*, four were selected and tested to determine their importance in +1 phosphate interaction. The residues are R103, R104, K132 and R129. The first three were selected for their possible direct interactions with the +1 phosphate. The last residue is expected to be important in binding to the complementary DNA strand opposed to the reacting strand. Alanine single and double mutants were designed to remove the positive charge associated with the residues. Also two glutamate double mutants and a quadruple mutant were also generated to completely reverse the positive charge. The resulting hFEN1 mutants will be referred to as R103A, R104A, R129A, K132A, R103A-R129A, R104A-K132A, R103E-R129E, R104E-K132E and quad-E (R103E-R104E-R129E-K132E).

The hFEN1 mutants were tested with DF1, which contains a five nucleotide 5'-flap and a single nucleotide 3'-flap, by another member of my research group (data is published in [REF]), and test the effect of the mutations on the endonuclease activity of hFEN1. In this work, wild-type FEN1 and mutant variants were tested with a single-flapped (SF) substrates which are suitable for exonuclease cleavage by FEN1. To determine whether mutants directly interact with the +1 phosphate at the 5' end (highlighted in *figure 4.11*), SF substrates with a phosphate (SF1) and hydroxyl (SF2) at the 5' position of the reacting strand were tested. The data for hFEN1 with the 5' modified SF constructs are shown in *figure 4.12*. The key difference between DF1 and SF1 is that DF1 contains a phosphodiester bond as opposed to a phosphate group in SF1, which presents a difference in charge. However, the difference between all the mutants and with wild-type is approximately 10-fold in all cases, suggesting that the difference in charge is not important for hFEN1 activity. The DF1 data is available alongside the SF1 and SF2 data on *page 163.*

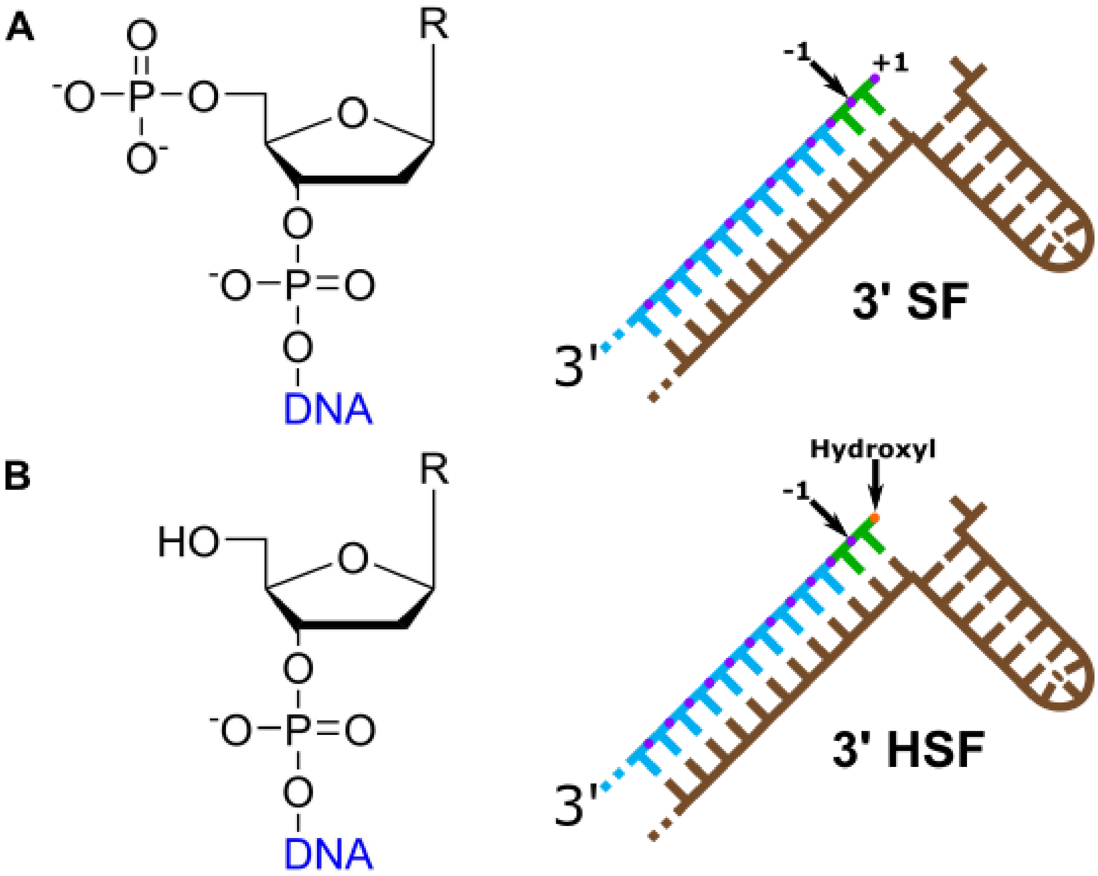


Figure 4.11: Single 3' flapped constructs. *Modifications at the 5' of the SF substrates with either a phosphate (**A**) or hydroxyl (**B**). Phosphates are coloured purple in the substrate schematics, and the hydroxyl is represented with an orange circle. The +1 and -1 nucleotides are highlighted in green. The reacting strands are cyan and the template strands are coloured in brown.*

Experiments were performed with 50 nM substrate concentrations of either the 3' SF (5' phosphate) or 3' HSF (5' hydroxyl) and variable concentrations of enzyme depending on the particular mutant used, to produce ~10-20% cleavage over variable time scales. Due to the nature of FEN1 as an exonuclease, single nucleotide products were removed and capillary electrophoresis was required to separate the 18-nucleotide substrate from the 17-nucleotide product. Electropherograms were manually integrated and converted into normalised rates. The data for wild type FEN1 and its alanine and glutamate variants are plotted in grouped bar charts in *figure 4.12*.

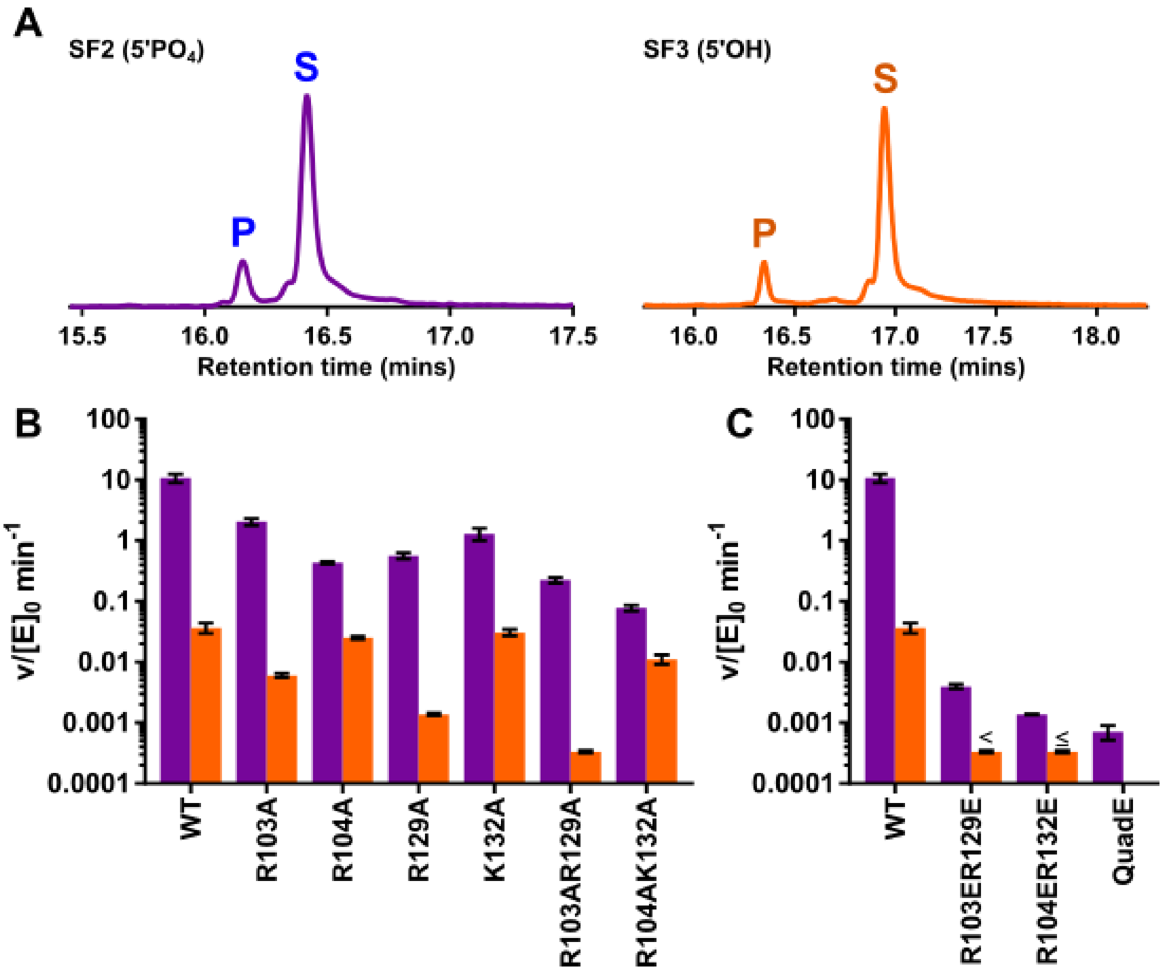


Figure 4.12: FEN1 mutants with +1 modifications. A. *Electropherograms of hFEN1 with a 5'-phosphate or -hydroxyl at the +1 position of single 3'-flap construct (see section 2.3) (SF1 and SF2 respectively) these peaks are integrated and processed as described in equation 2.1. Note, retention times shifted over time; this problem was fixed much later. Wild type (WT) FEN1 with the alanine (**B**) or glutamate mutants (**C**). Rates with SF1 (purple) or SF2) (orange) displaying differences depending on the mutant. Experiments for R103E-R129E and R104E-K132E mutants with the 5'-hydroxyl substrate were too slow to be accurately measured as multiple turnover experiments and are therefore represented with < and ≤ symbols.*

The data show that each mutant had differing levels of impairment with each modification. For the +1 phosphate experiments; R103A and K132A show small decreases in rate of 5- and 8-fold compared to wildtype FEN1. The impact of the R104A and R129A mutations are greater with 25- and 20-fold respective decreases in reaction rate. Double alanine mutants show an additive decrease of their two encompassed mutations with R103A-

R129A and R104A-K132A displaying a 50- and 140-fold decrease respectively. The double glutamate mutations exhibit larger decreases in activity compared to their alanine variants, which can be associated with a repulsion effect of the residue charges. The rate decreases of R103E-R129E and R104E-K132E were 2700- and 7800-fold respectively, presenting a dramatic decrease opposed to the alanine mutants. The QuadE mutant showed an even greater decrease in rate, with a 15000-fold decrease in reaction rate.

The 5'-hydroxyl data revealed some differences in the impact of the mutants. The R103A and R129A mutants show larger (8- and 35-fold) decreases in reaction rate compared to wild-type, R103A-R129A exhibits an additive effect of the single mutations with a decrease of 140-fold. However, R104A and K132A show minor reductions of around 2- and 1.5-fold compared to wild-type. Removing R104 and K132 in combination with the +1 phosphate shows a similar decrease in rate compared to wild-type with the 5'-hydroxyl. This suggests that these residues directly interact with the 5'-phosphate, as they show no effect compared to wild-type in its absence. The double alanine mutant R104A-K132A also displays a similar decrease in rate, with a small 4-fold decrease. On the other hand, R103E-R129E, R104E-K132E and the QuadE mutants were too slow to record on a reliable timescale with the poorer substrate.

The single mutation rate decreases are suggested to be representative of electrostatic guidance interactions; which is an apt explanation for the removal of positive charges [99]. Replacing basic residues with glutamates provides a larger decrease in reaction rate, probably due to a complete reversal of the residue’s charge, repulsing the 5'-phosphate. Interestingly, no significant change in the K_D of these mutants (data collected by collaborators and not shown) was observed compared to the reduction in reaction rate, suggesting that the loss of activity is not due to a reduced ability to bind the substrates [99]. The QuadE mutant has a rate comparable to the blocked state for hFEN1 with biotinylated substrates preincubated with streptavidin [101]. These data suggest that these mutations disrupt the threading process, and are potentially responsible for steering the phosphates into the correct position for cleavage. This is especially so for R104 and K132, which our functional data suggests directly interact with the +1 phosphate. Streptavidin trapping and blocking studies with these mutants may provide interesting insights into the characteristics of each mutant.

These data can be extrapolated slightly to examine the effects of these residues within the superfamily. A shortened sequence alignment of the region examined in *figure 4.10* of the various superfamily members shows conservation of the residues examined. In particular, R103 and R104 are consistently conserved through human members. R129 is conserved in hEXO1 but this region is not present in hGEN1 and hXPG; this may be due to a lack of the helical cap in these enzymes. K132 is not conserved in any of the other family members, although it is heavily conserved in FEN1 homologs. Interestingly K132 is important for direct +1 phosphate interaction and its absence in hEXO1 may be responsible for the lack of precision of flapped substrates. This alignment is shown in *figure 4.13* with other key residues highlighted in the region.

The most highly conserved residues in the region, K93 and R100 in FEN1 or K85 and R92 in EXO1, have been characterised previously for both proteins and mutations produce a large decrease in activity [16, 89]. R100 and R92 mutations in these enzymes are effectively catalytically-inert on a biological time-scale. The R92 residue in hEXO1 shows electrostatic interaction with the scissile phosphate (*figure 4.13B*), and R96 shows similar interactions with the +1 phosphate as R104 in FEN1 (*figure 4.13C*). R121, the EXO1 R129 homolog has some electrostatic interaction with a nucleobase in the template strand (*figure 4.13B*), before interacting directly with a phosphate in the template strand at the junction, which may be an indicator of threading before catalysis occurs.

Examination of DNA changes between the two structures shows that the template strand unwinds and the flapped region moves closer to the metal ions (*figure 4.13F-G*). The Y40 homolog in hEXO1, H36, appears to guide the nucleotides into place by stacking with the nucleobase. The second stacking residue, Y32, may have similar interactions before the flap is in position, and may account for the change in DNA conformation. The key difference between the cleaved and uncleaved states is that the important residues studied in hFEN1 are all positioned for catalysis in the cleaved state. The crystal structures examined are snapshots of substrate equilibration before cleavage and demonstrate that the residues first reposition the DNA before catalysis can occur.

A simple test with hEXO1 and a 5'-hydroxyl nicked duplex (ND3) gave an interesting result compared to that observed with hFEN1 (*figure 4.14*). Where hFEN1 and its mutants perform a single exonucleolytic cut one nucleotide into the duplex, hEXO1 either removes one large tract of DNA or the hydroxyl end group drives the protein's processive nature (described in *section 3.4*), as the first product of a single nucleotide excision has a greater affinity for hEXO1 in this case. However, this has yet to be confirmed, but future work in this area for hEXO1 could provide interesting insight.

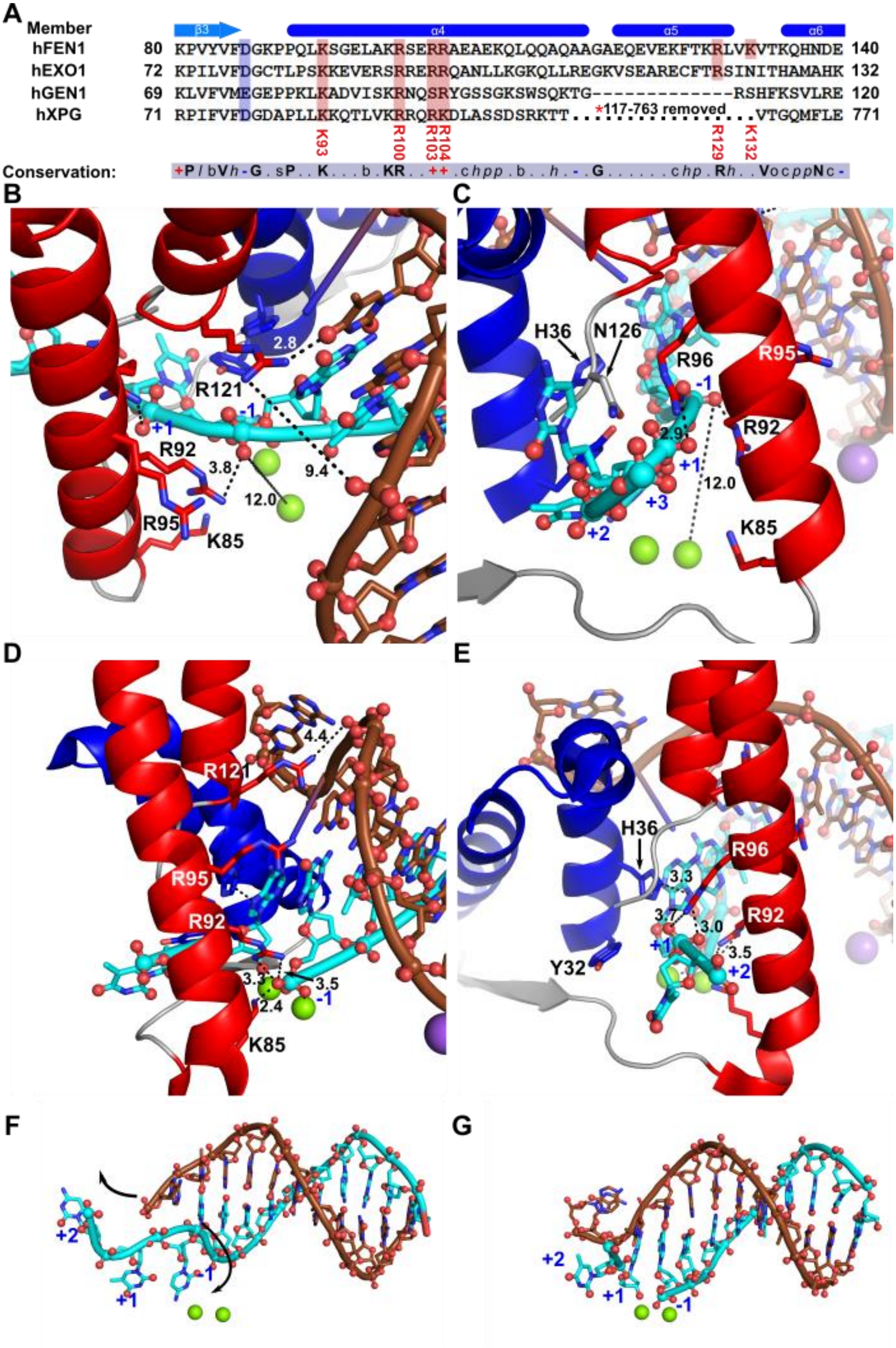

A
Member
β3
α4
α5
α6
hFEN1 80 KPVYVFDGKPPQLKSGELAKRSERRAEAEKQLQQAQAAGAEQEVEKFTKRLVKVTKQHNDE 140
hEXO1 72 KPILVFDGCTLPSKKEVERSRRERRQANLLKGKQLLREGKVSEARECFTRSINITHAMAHK 132
hGEN1 69 KLVFVMEGEPPKLKADVISKRNQSRYGSSGKSWSQKTG-------------RSHFKSVLRE 120
hXPG 71 RPIFVFDGDAPLLKKQTLVKRRQRKDLASSDSRKTT...*117-763 removed...VTGQMFLE 771
K93
R100
R103
R104
R129
K132
Conservation: +P/bVh-G.sP..K...b.KR..++.chpp.b..h.-.G......chp.Rh..VocppNc-
B
2.8
R121
-1
+1
R92
3.8
9.4
12.0
R95
K85
C
H36
N126
R96
R95
-1
R92
2.9
+1
12.0
+3
+2
K85
D
4.4
R121
R95
R92
3.3
3.5
2.4
-1
K85
E
H36
R96
3.3
3.7
3.0
R92
+1
3.5
+2
Y32
F
+2
-1
+1
G
+2
+1
-1

Figure 4.13: Conservation of +1 phosphate residues and examination of hEXO1 crystals. ***A.*** *Sequence alignment of FEN superfamily members, highlighting key residues within the helical arch and cap region. The hEXO1 homologs of the mutants used in hFEN1 are highlighted. Crystal structure representations of hEXO1 in a threaded state where the DNA is either threaded (5V0E;* ***B-C****) or cleaved (5V0D;* ***D-E****). A side (****B****) and back (****C****) view of the uncleaved threaded structure displaying the interactions and distances of the particular residues shown to be conserved. In this crystal the DNA is not in the correct conformation, with the scissile phosphate being 12 angstroms from the cations. Some interactions are present with the various basic residues. A side (****D****) and back (****E****) view of the cleaved crystal structure shows closer distances between the key residues and their respective targets, suggesting that once the flap is threaded and in the correct position, catalysis can occur. The DNA of the uncleaved (****F****) and cleaved (****G****) crystals with the differences highlighted for substrate steering, which back up the concept of a correct position for catalysis.*

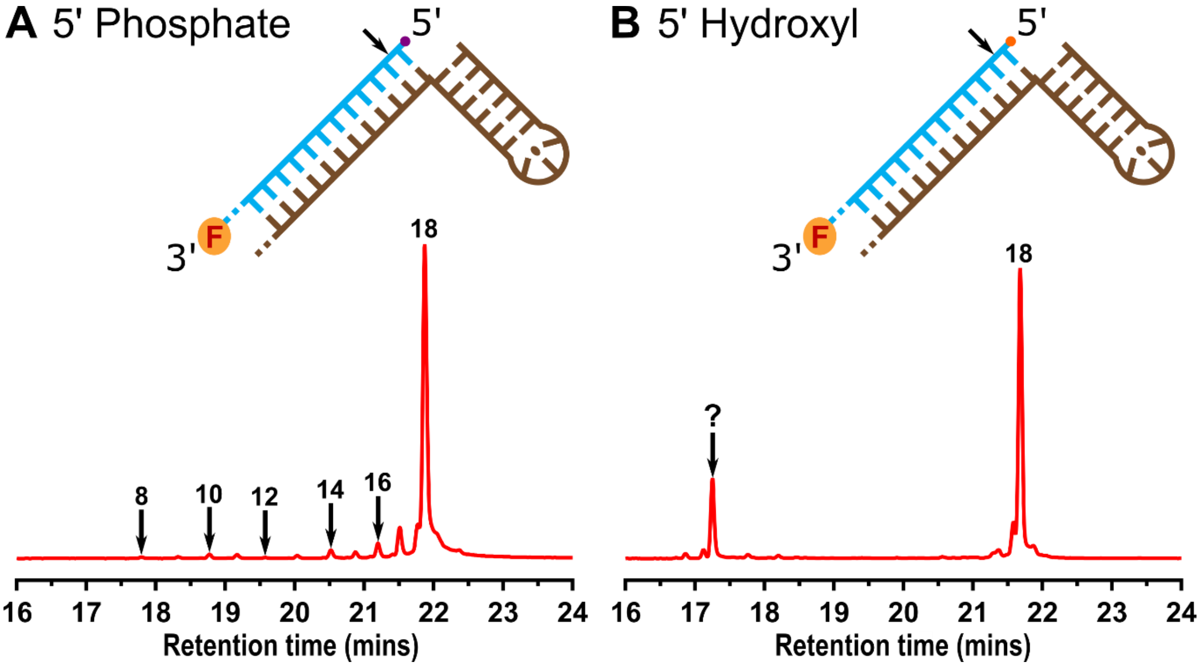


Figure 4.14: EXO1 cleavage patterns. *Electropherograms of hEXO1 with nicked duplexes with a 5' phosphate (****A****) and 5' hydroxyl (****B****), which are both labelled with a 3' fluorescein. The substrate peaks are both 18 nucleotides. Assays with the 5' phosphate show a ladder forming, where each duplex product can be processed as readily as the initial substrate. However, assays with the 5' hydroxyl nicked duplex showed formation of a single smaller product, where it is unclear whether the enzyme is cutting as an endonuclease or by acting processively.*

4.4 Summary of endonuclease studies

The endonuclease capability of hEXO1-352 is much slower compared to hFEN1 (90-fold), whereas the exonuclease ability of the two enzymes is relatively similar. However, this may be due to the processive nature of hEXO1, and the fact that the simple multiple turnover experiments performed do not truly measure this ability. Rather importantly, a 10-fold increase in rate is observed when only one 5' end is available for catalysis, which would suggest that more than one activity is occurring in bimolecular substrates, resulting from additional binding modes to the substrate.

Human EXO1 must first thread a flapped substrate through its helical archway before it can perform hydrolysis. However, compared to hFEN1 it is not capable of completely trapping the flapped substrate, with only ~20% of the substrate being processed on a biologically relevant timescale in a “trapping experiment”. The remainder of the substrate is treated around 2300-fold slower than the fast phase, which is comparable to the blocked state in which the substrate is unable to thread. Saturation experiments showed that the substrate was saturated onto the enzyme, and suggests threading is an on-enzyme effect that can be expressed by an equilibrium. These findings complement recent crystal structures of hEXO1 in a threaded state. These data and the crystallographic evidence demonstrate that hEXO1 threads flapped substrates to position the flap for catalysis, consistent with claims that hEXO1 acts as a back-up for hFEN1 *in vivo*.

Similar crystal structures were recently assigned for hFEN1 in a threaded state, and from these snapshots key residues were selected to test whether they are important in steering the phosphates into place. These residues show varying differences in rate when used to process a single flap substrate with either a phosphate or hydroxyl at the +1 position. From these studies it became apparent that R104 and K132 directly interact with the +1 phosphate, whereas R103 and R129 must have other roles in catalysis or interaction with other phosphates. These residues are conserved in hEXO1 (except for K132), and may have similar roles in stabilising the hEXO1 threaded state.

Chapter 5: The Role of Substrate Dynamics in Catalysis

5.1 Enzyme-induced conformational change

5.2 The limitations of substrate binding techniques with hEXO1

5.3 Summary of substrate dynamics

Chapter 5: The Role of Substrate Dynamics in Catalysis

5.1 Enzyme-induced conformational change

Previously, a reaction mechanism scheme (see *figure 3.5*) was used to explain the general scheme expected for FEN family members. Kinetic studies were used to examine various stages in the cycle, and I identified that product release is rate limiting at high substrate concentrations in hEXO1 (chapter 3). Analysis in the presence of calcium ions, which are known to bind in a similar place to catalytic magnesium ion but do not permit reaction, allowed for examination of the expected conformational change step highlighted in *figure 5.1* for hFEN1.

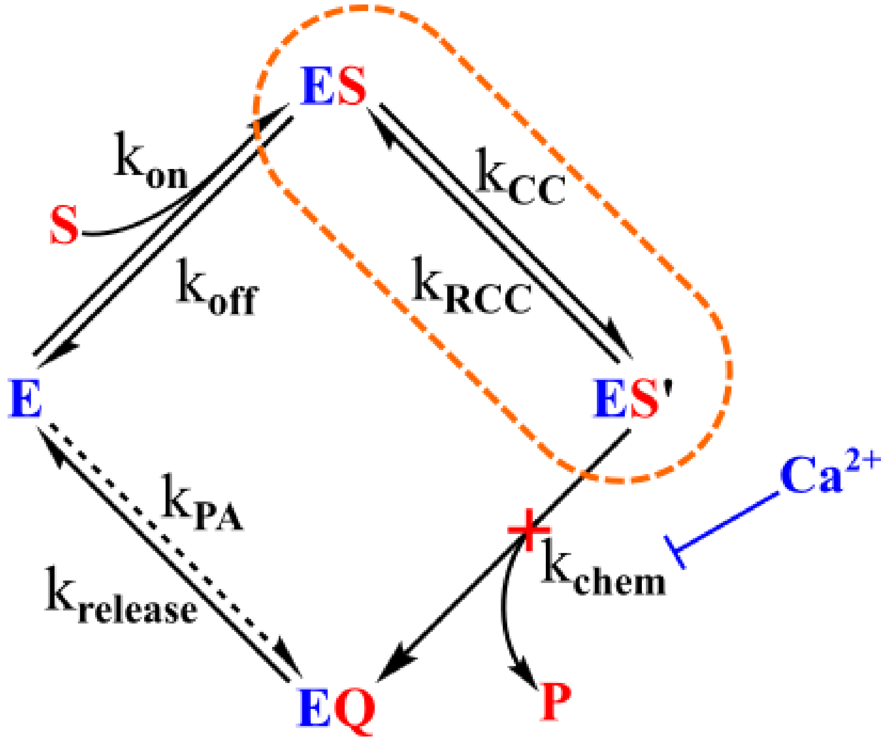


Figure 5.1: Reaction scheme revisited. *Reaction scheme highlighting the conformational change step in the presence of* Ca^{2+}*, which is known to inhibit the chemistry of the FEN superfamily.*

The conformational change of the enzyme substrate complex for hFEN1 was first attributed to double nucleotide unpairing (DNU) [118], where the +1 and -1 nucleotides are unpaired for placement of the scissile phosphate between the divalent metals (*figure 5.2A*). This phenomenon has been examined by exciton-coupled circular dichroism (ECCD) using tandem 2-aminopurine nucleotides in either the +1, -1 or -1, -2 positions. Once the enzyme binds the substrate a change in the CD signal is assigned to a break in the nucleotide stacking at the +1, -1 positions attributed to local unpairing of the DNA. An example of this change in signal is shown in *figure 5.2B*, where a comparison of the signal in the presence and absence of enzyme shows a drop in intensity.

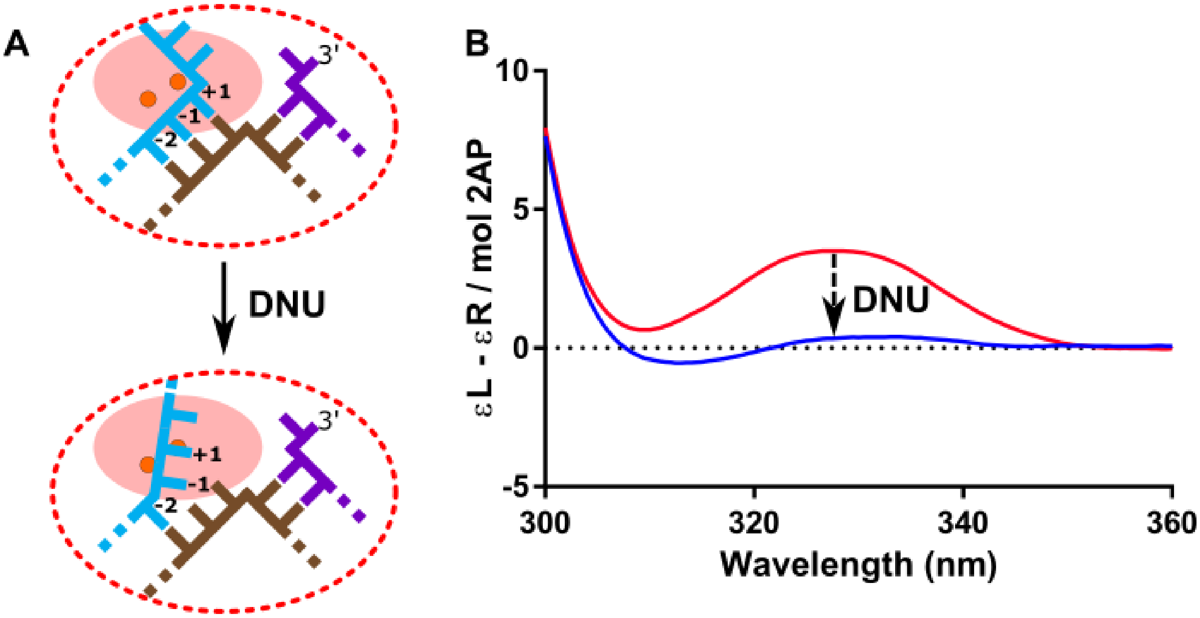


Figure 5.2: Double nucleotide unpairing in hFEN1. A. *A cartoon representation of DNU, where the +1 and -1 nucleotides are unstacked and positioned between the metal ions.* ***B.*** *An ECCD trace of hFEN1 with a double flapped substrate with tandem 2-aminopurines in the +1, -1 positions. The red trace represents the substrate alone in buffer containing* Ca^{2+}*, and the blue trace is once enzyme is added; a drop in signal is observed. The Y-axis units is the left circularly polarised absorbance subtracted by the right circularly polarised absorbance divided by the concentration of 2-AP molecules.*

The DNU model has been a point of dispute in the study of hFEN1, and recently crystallographic evidence with threaded flapped substrates does not show the DNA to be in this state. However, in its place the crystal exhibits a distortion in the structure of the DNA which could account for the change in CD signal in *figure 5.2B*. Modelling with perfect B-form DNA (generated by web3DNA [119]) with the same sequence used for the flapped substrate in the crystal structure, the DNA distortion can be observed. In *figure 5.3* the threaded crystal structure of hFEN1 [99] is examined in detail.

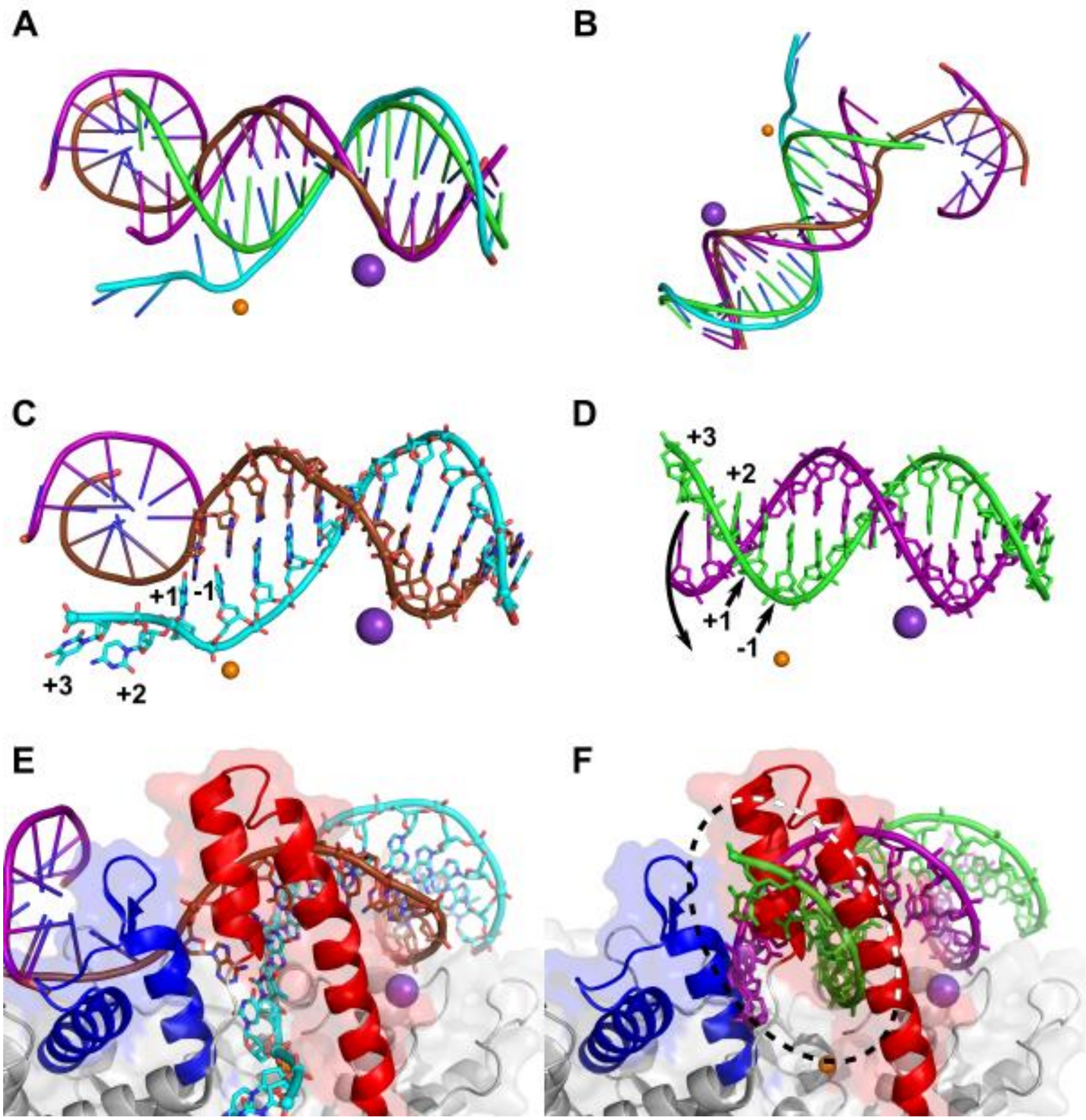


Figure 5.3: Distortion of the DNA substrate with hFEN1. *The double flapped (DF) substrate overlaid with a perfect B-form helix (BF) from the front (***A***) and above (***B***). The DF is coloured with a cyan reacting strand, brown template strand and purple 3' flap strand. Whereas, the BF DNA is coloured with a green reacting strand and purple template strand. The metals are shown to give spatial awareness of the structure. It is clear that distortion is present in all regions of the substrate and not just the 5' flap.* ***C.*** *The DF substrate highlighting the -1, +1, +2 and +3 nucleotides, displays that the nucleotides are not unstacked from the duplex, discounting DNU.* ***D.*** *The BF DNA showing the clear level of distortion needed to reach the position of the -1, +1, +2 and +3 nucleotides in the DF substrate.* ***E-F.*** *FEN1 in the presence*

*of the DF (**E**) and BF (**F**); the distortion is required for the flap to be correctly positioned through the enzyme.*

To determine whether a similar distortion is required for cleavage by hEXO1, ECCD experiments were performed. Spectra were taken of 10 μM substrate (e.g. ND$_{+1-1}$, OVH$_{+1-1}$, etc) in buffer (CDRB), which contains Ca^{2+}. If samples were in the presence of enzyme, a 12.5 μM final concentration was used to ensure that the substrate was saturated with enzyme. Various constructs were tested with hEXO1, due to its lack of specificity, and a selection of the CD spectra are shown in *figure 5.4*.

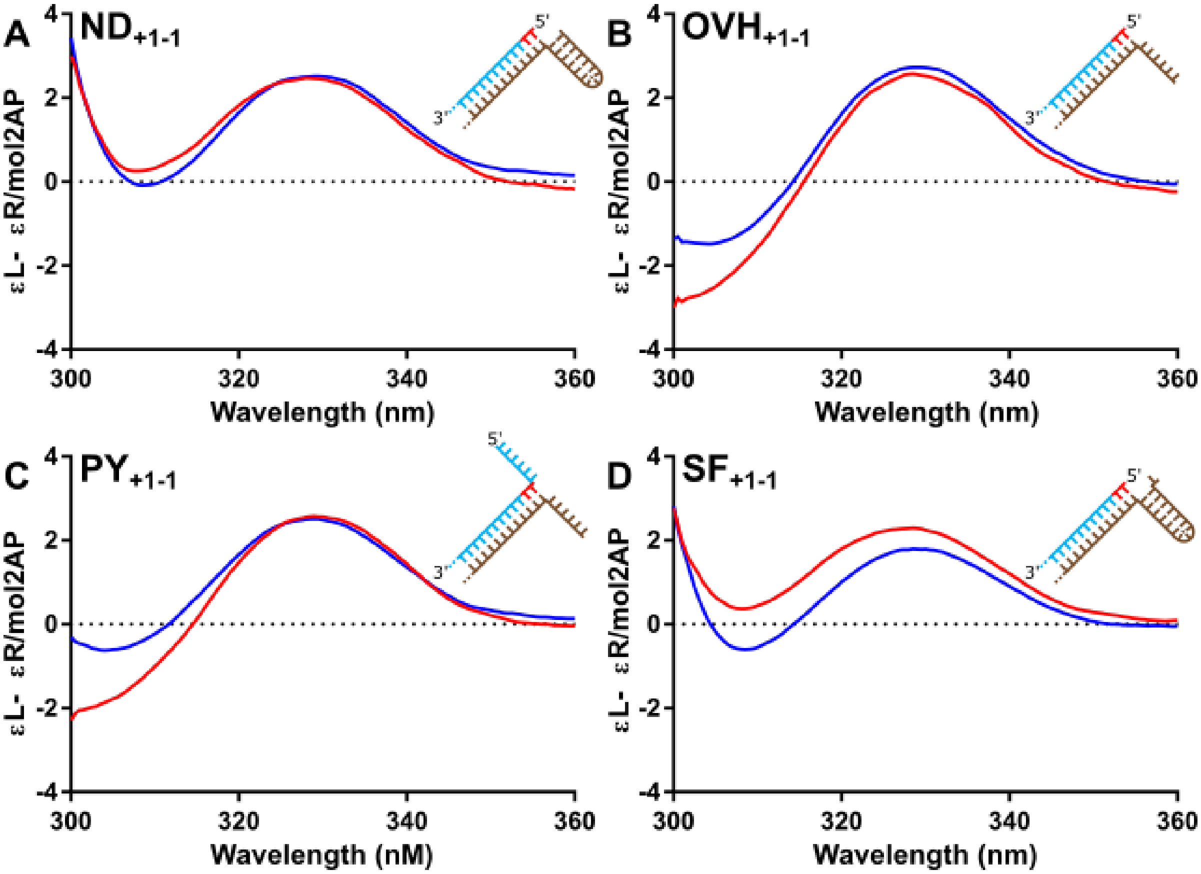


Figure 5.4: ECCD spectra of hEXO1 with tandem 2-aminopurines in the +1, -1 position. *Spectra of hEXO1 with (**A**) ND$_{+1-1}$, (**B**) OVH$_{+1-1}$, (**C**) PY$_{+1-1}$ and (**D**) SF$_{+1-1}$ (see* section 2.3*). The red trace represents the substrate alone and the blue trace is in the presence of enzyme. Spectra were taken stepwise over 300-480 nm, but only 300-360 nm is displayed with the remainder being used to normalise the traces.*

The spectra for hEXO1 in *figure 5.4* do not follow the pattern observed for hFEN1 in *figure 5.2B*, with no significant change in CD signal between the ES and S-alone samples. A larger range of substrates were tested; a nicked duplex, 3' overhang, pseudo-Y, single flap and unimolecular 3' overhang, with tandem 2-aminopurines (2-AP) in both the +1, -1 or -1, -2 positions (see *section 2.3*). Spectra were taken in the presence of 10 mM Ca^{2+} before another trace was taken after the supplementation of excess EDTA. The spectra were baseline-corrected, converted to molar ellipticity, gently smoothed and normalised against the spectra taken in the 380-480 nm region. The full data is shown in *figure 5.5*, where the CD value at 326 nm (the maxima of tandem 2-AP) was averaged with three repeats.

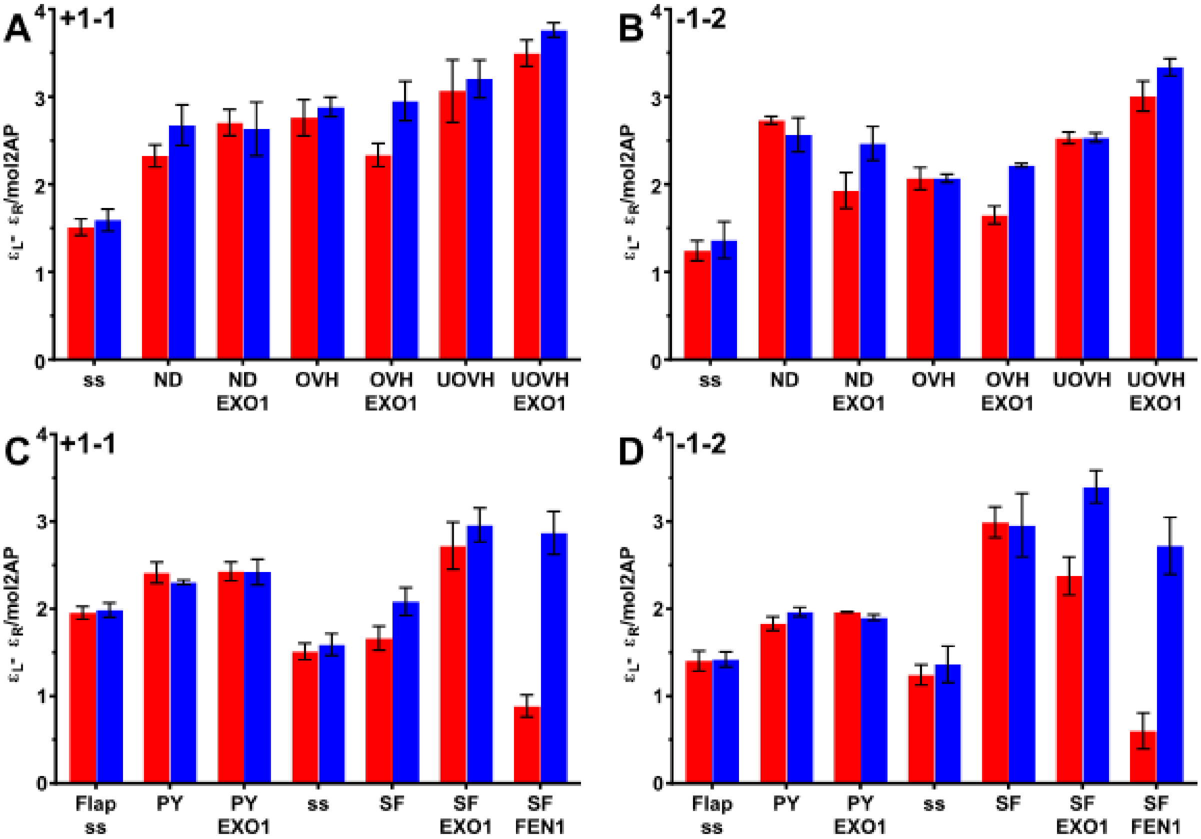


Figure 5.5: ECCD data for hEXO1 taken at 326 nm. *EXO1 substrates tested in the +1-1 (**A**) and -1-2 (**B**) positions, with single-stranded DNA (ss), nicked duplex (ND), 3' overhang (OVH) and unimolecular 3' overhang (UOVH) substrates. Samples were taken in the presence of* Ca^{2+} *(red) and excess EDTA (blue). FEN1 substrates tested in the +1-1 (**C**) and -1-2 (**D**) positions, with single stranded DNA (ss) or single-stranded flapped DNA (ss Flap), a pseudo-Y (PY), and a single 3' flapped duplex (SF). Samples are in the presence of 12.5 µM of enzyme when*

indicated. Bars are colour coded as above, and a hFEN1 control with the single flap is also displayed.

The data in *figure 5.5* are evidence that the conformational change observed with hFEN1 is not conserved in hEXO1, with little difference in the spectra of hEXO1 with either Ca^{2+} or EDTA (EDTA abolishes the conformational change in hFEN1, due to chelation of the metal ions). Therefore, the enzyme-induced substrate conformational change observed in hFEN1 is not universal to the family members, and may be important in the precise activity of hFEN1. Examination of the +1, -1 and -2 bases from crystal structures for hFEN1 D86N, R100A, hEXO1 WT and B-form DNA to determine the level of local nucleotide distortion is displayed in *figure 5.6*.

The nucleotides examined by ECCD in hFEN1 and hEXO1 are highlighted in *figure 5.6*, with their stacking on one another compared. A distortion in the +1, -1 and -1, -2 nucleotide stacking has been observed in hFEN1 previously [99], with the D86N and R100A mutant crystal structures showing differing levels of distortion compared to the B-DNA. This distortion was expected to be representative of the changes in ECCD signal produced when enzyme is added to substrates containing tandem 2-aminopurines. Surprisingly, examination of the same nucleotides in hEXO1 (5V08) displays similar local distortion performed by this enzyme.

Currently, no updated hypothesis for the difference in ECCD signal has been suggested. However, the DNA backbone in the hEXO1 crystals is more uniform compared to the hFEN1 examples. Examination of the threaded crystal structures of hEXO1 with perfect B-form DNA demonstrates an interesting stabilisation of the DNA as the ES complex moves through catalysis (*figure 5.7*).

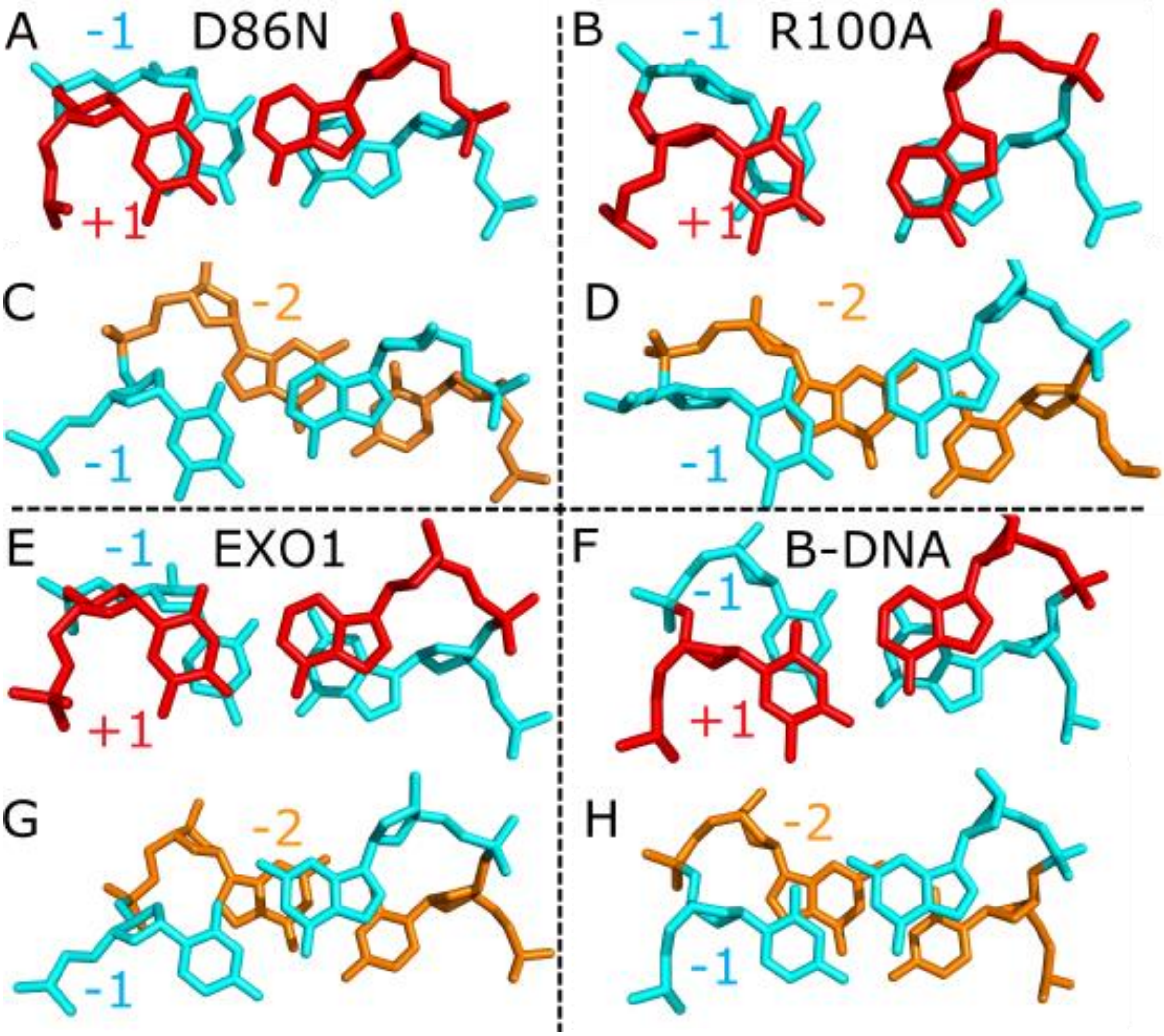
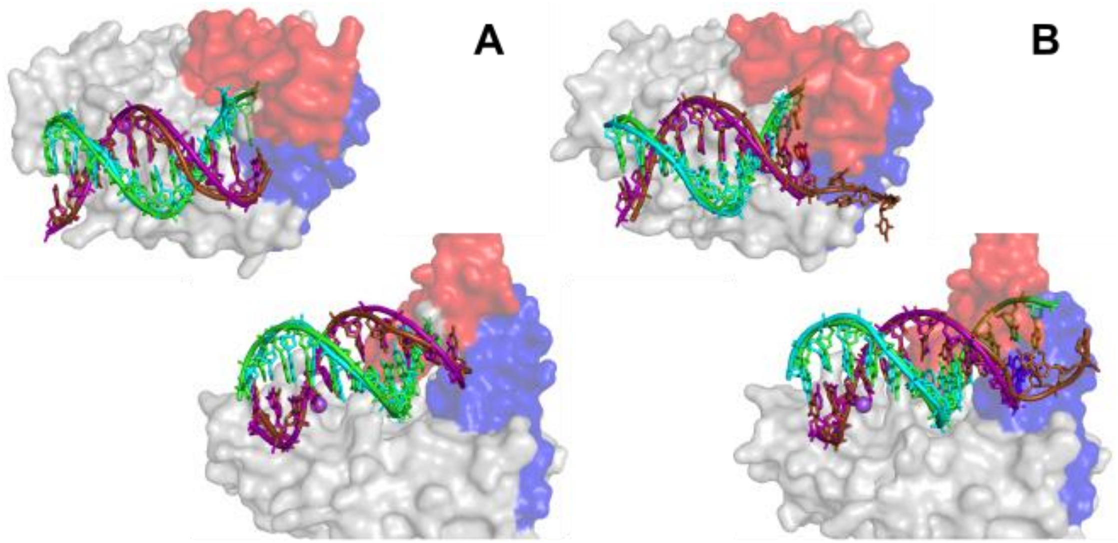
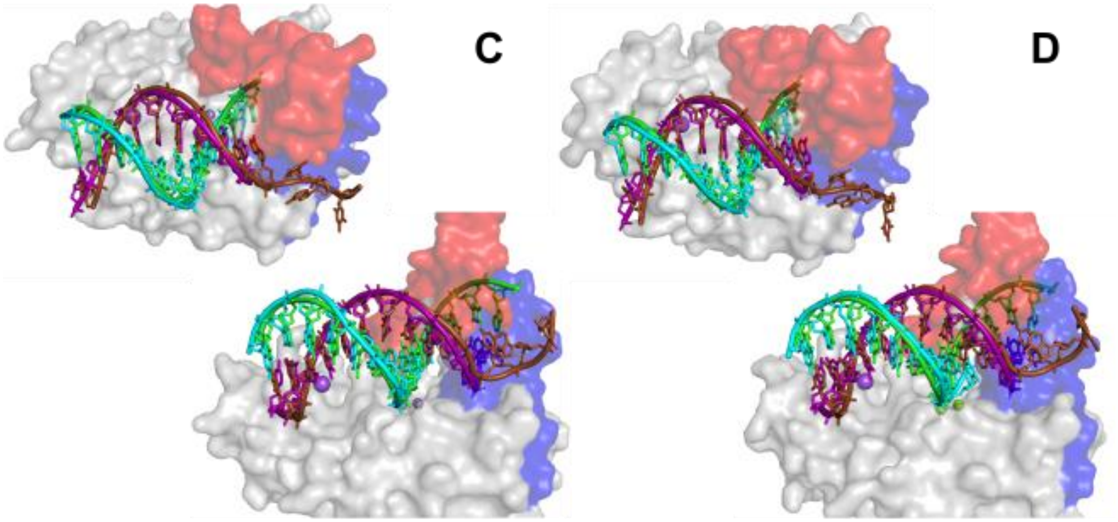


Figure 5.6: Nucleotide distortion of FEN1 and EXO1. *DNA nucleotides +1, -1 (**A-B** & **E-F**) and -1, -2 (**C-D** & **G-H**), where the nucleotides +1, -1 and -2 are red, cyan and orange respectively. The nucleotides from hFEN1 variants D86N (**A** & **C**), R100A (**B** & **D**), and EXO1 WT (**E** & **G**), with B-DNA for comparison (**F** & **H**). In each case the reacting strand is on the left and the template strand is on the right. The level of local distortion can be examined by comparison with the B-DNA. The crystal structures examined were 5UM9 (D86N), 5KSE (R100A) and 5V08 (EXO1).*



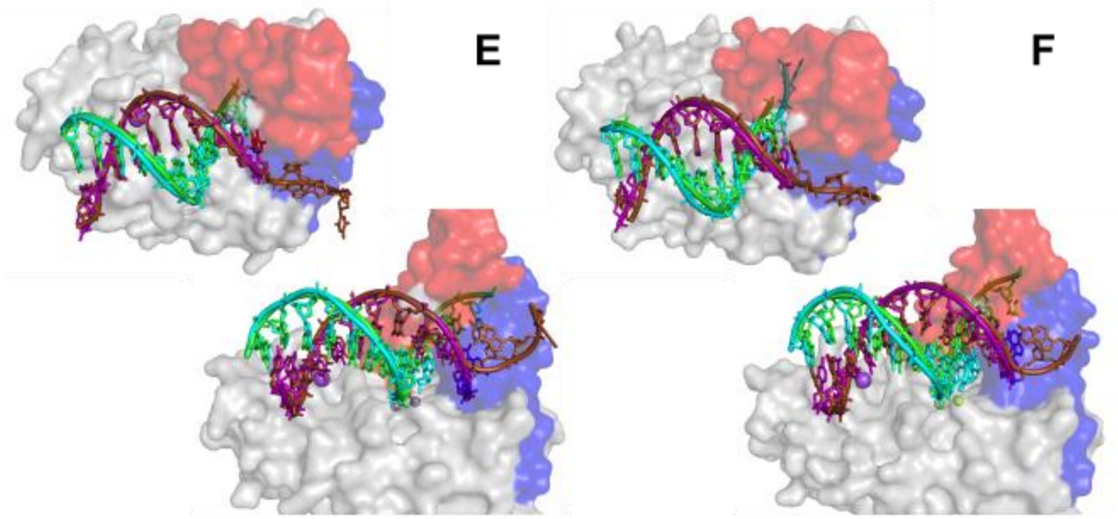
Substrate binding

Rotation 1



Rotation 2 Intermediate

Active Site Positioning



EXO Substrate Cleavage

Flapped Substrate Cleavage

Figure 5.7: DNA distortion of the crystal structures of hEXO1. *Varying ES states were crystallised at differing phases of the catalytic cycle [100]. The helical arch and hydrophobic wedge are red and blue respectively. The crystallised DNA is represented with a cyan reacting strand and brown template strand. The B-DNA's match the sequence of the crystallised DNA and have a green reacting strand and purple template strand. The crystals for each phase: 5UZV (**A**), 5V04 (**B**), 5V05 (**C**), 5V08 (**D**), 5V0A (**E**) and 5V0D (**F**). **A.** Initial DNA binding, displays backbone distortion. **B.** The first rotation of the DNA towards the active site also contains minor distortion. **C.** An intermediate of the second rotation, with similar distortion to **B**. **D.** Active site positioning shows reduced backbone distortion compared to previous states. **E.** Substrate cleavage of an exonucleolytic substrate is almost in perfect B-form conformation. **F.** Cleavage of a flapped substrate requires DNA distortion, which is consistent with FEN1.*

As demonstrated in *figure 5.7*, the different stages of hEXO1 interaction with its substrate display increasing enzyme-induced stabilisation of the DNA strands in each phase towards hydrolysis, as the DNA becomes more B-form throughout the cycle. The various stages shown by *figure 5.7* from *A-E* are a continuous cycle starting with substrate binding, followed by multiple DNA rotations before active site positioning, and finally hydrolysis. These crystals were produced by time resolved crystallography, where samples were frozen at different time-points and imaged to examine the results [100]. All the DNA substrates are the same except in the case of the flapped substrate cleavage.

The figure represents a consistent stabilisation of the substrate into a B-form conformation, with exonucleolytic cleavage of the DNA appearing to be in a B-form helix. However, panel figure 5.7*F* shows higher distortion when compared to the other states, which is an interesting observation as a 5' flapped substrate is present. Therefore, flapped substrates may require backbone distortion to be processed. A comparison of the substrate DNA from the threaded hFEN1 crystal and hEXO1 threaded crystal is shown in *figure 5.8*.

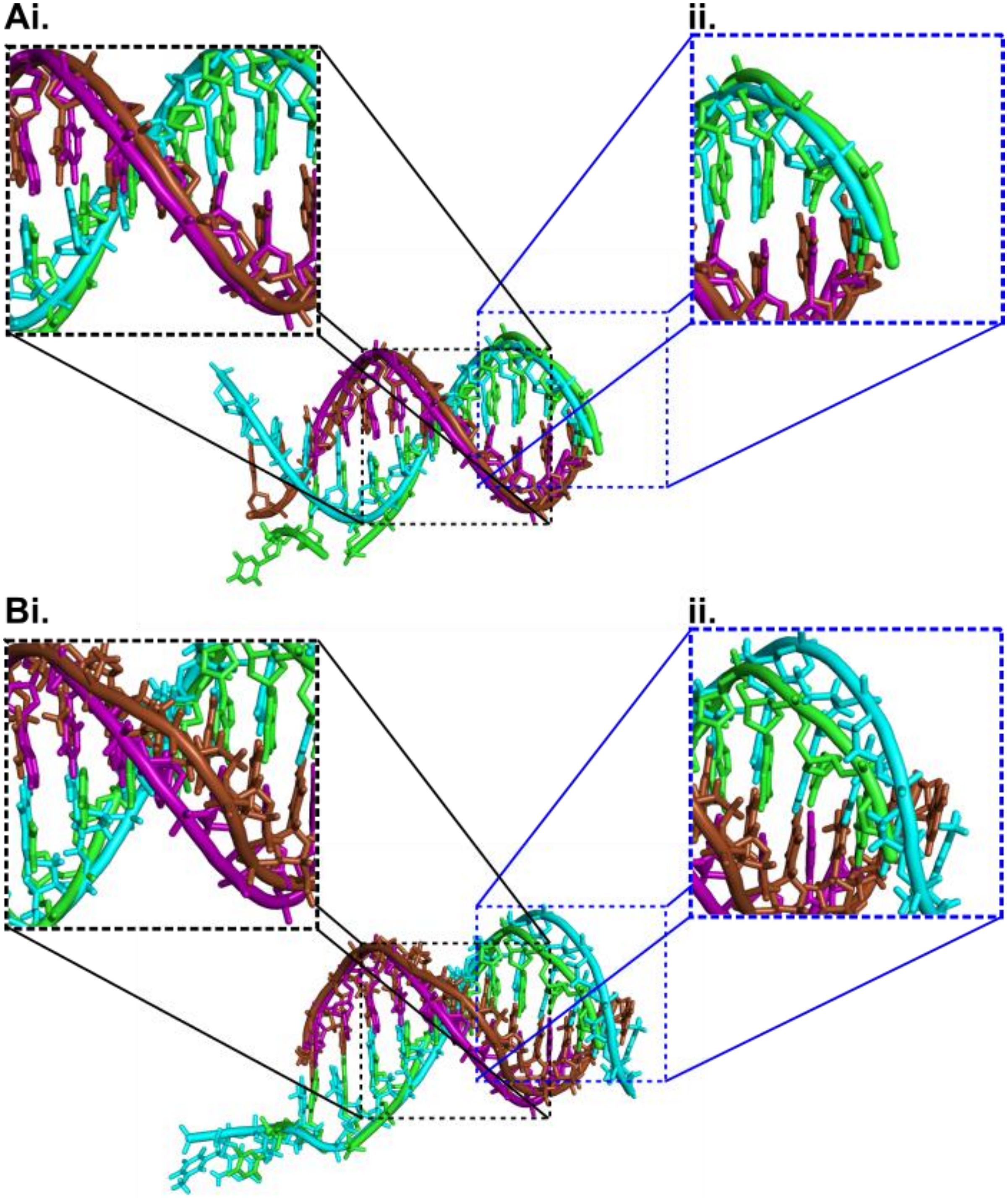


Figure 5.8: Comparison of hFEN1 substrate distortion with the hEXO1 substrate. A. *A close up of the DNA used to crystallise hEXO1, which is a cleaved flapped DNA substrate (5V0D), aligned to a B-DNA construct. The B-DNA is coloured with a brown template strand and cyan reacting strand, whereas the cleaved DNA is coloured with a purple template strand and a green reactive strand. The snapshots (****i****) and (****ii****) show an examination of the template strands*

and reactive strands respectively. ***B.*** *The flapped substrate from hEXO1 aligned with the flapped substrate from hFEN1 (5UM9). The EXO1 flapped substrate is coloured as in panel* ***A****. The FEN1 substrate is coloured brown and cyan for the template and reactive strand respectively. Again, the snapshots (****i****) and (****ii****) show a comparison of the template and reactive strands respectively. The DNA backbone distortion is noticeably larger for hFEN1.*

The comparisons made in *figure 5.8* show that DNA backbone distortion is present in hEXO1 when processing a flapped substrate; however, the distortion presented by hFEN1 is more severe. Therefore, it can be hypothesised that processing flapped substrates requires DNA distortion for accommodation of the 5'-flap through the helical arch. Crystals with 5' flapped constructs appear to compress the duplex DNA, possibly to stabilise the flapped strand through the active site; this is currently only my conjecture and no physical evidence has been provided.

The idea that hEXO1 stabilises the DNA strands towards B-form upon binding may play an important role in realigning the DNA substrate for repeated cleavage by processivity; this process would allow for the duplex to slide further into the active site after every successive cleavage. The reduced ability for distortion of flapped substrates may simply be due to the enzyme’s preference for exonucleolytic cleavage. The more severe backbone distortion in hFEN1 may be important for its hydrolytic precision, which successfully cleaves at one position the majority of the time, whereas, data presented for hEXO1 show that it cleaves multiple positions of an endonucleolytic substrate. All these hypothesises require further testing to confirm the nature of the DNA distortion, and no conclusion can be made on the exact cause of the ECCD shift that is observable in hFEN1.

5.2 The limitations of substrate binding techniques with hEXO1

Attempts at defining the substrate binding phase described by k_{off}/k_{on} in *figure 5.1* were performed to determine the effect substrate binding has on the cycle as a whole. Two techniques were utilised to try and calculate this factor; Förster resonance energy transfer (FRET) and fluorescence anisotropy (FA). FRET relies on measuring the energy transfer from a donor fluorescent dye to an acceptor fluorescent dye. In an attempt to develop an understanding of how the substrate bends when it interacts with the enzyme.

In our FRET assays three differently labelled DNA constructs are used, which are each titrated with increasing concentrations of enzyme. In *figure 5.9* the three forms of a nicked duplex construct for hEXO1 and a double flapped construct for hFEN1 are displayed. The three constructs serve different roles in the determination of FRET efficiency. Firstly, an unlabelled DNA trimolecular substrate is used which is referred to as NL (non-labelled). Secondly, a singly labelled substrate which contains the donor label (in this case fluorescein (FAM)) which is referred to as DOL (donor-only labelled). Finally, a doubly labelled substrate with both the donor and acceptor dyes (FAM and 5-carboxytetramethylrhodamine (TAMRA) respectively), which is known as the DAL (donor-acceptor labelled) substrate. All experiments were performed with a buffer containing Ca^{2+}, which renders FEN1 and EXO1 catalytically inert, therefore, the substrates are not destroyed by the enzymes.

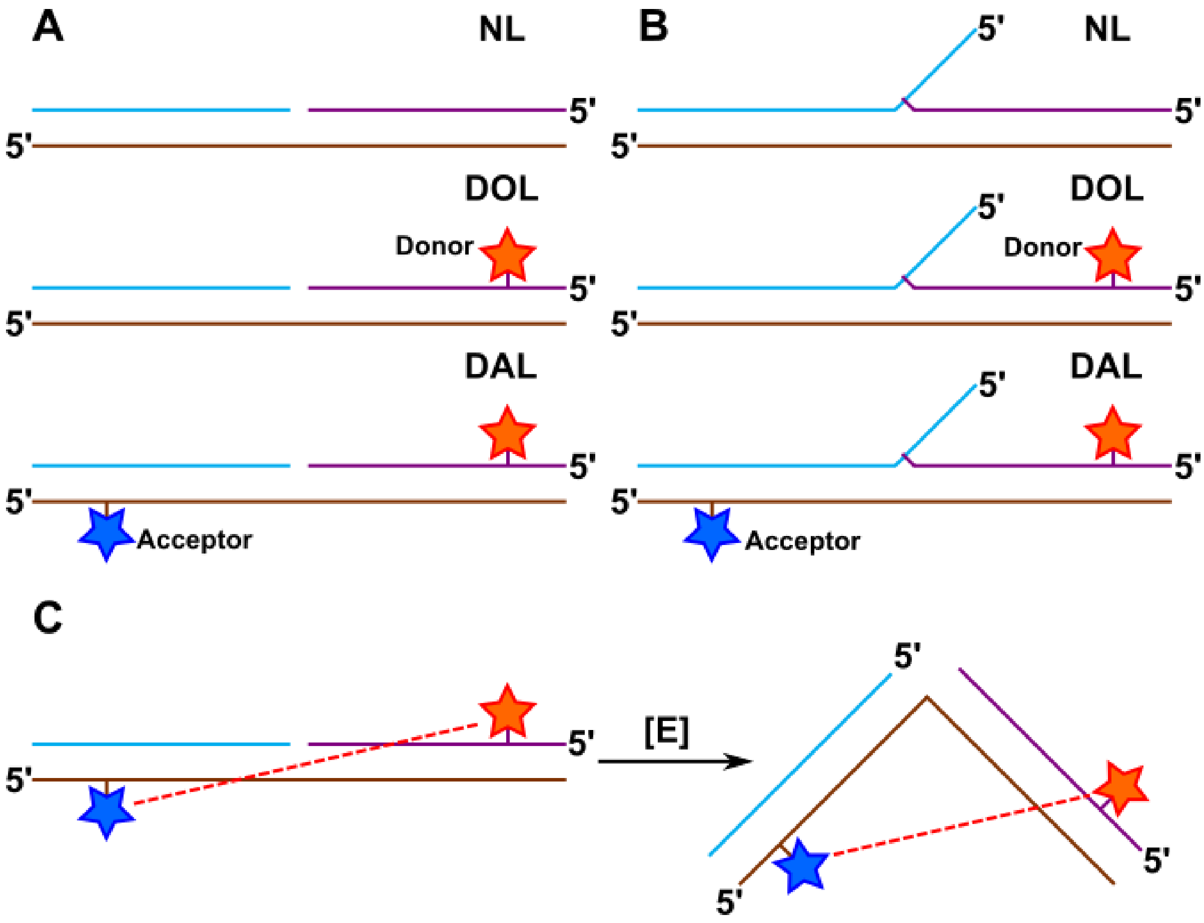
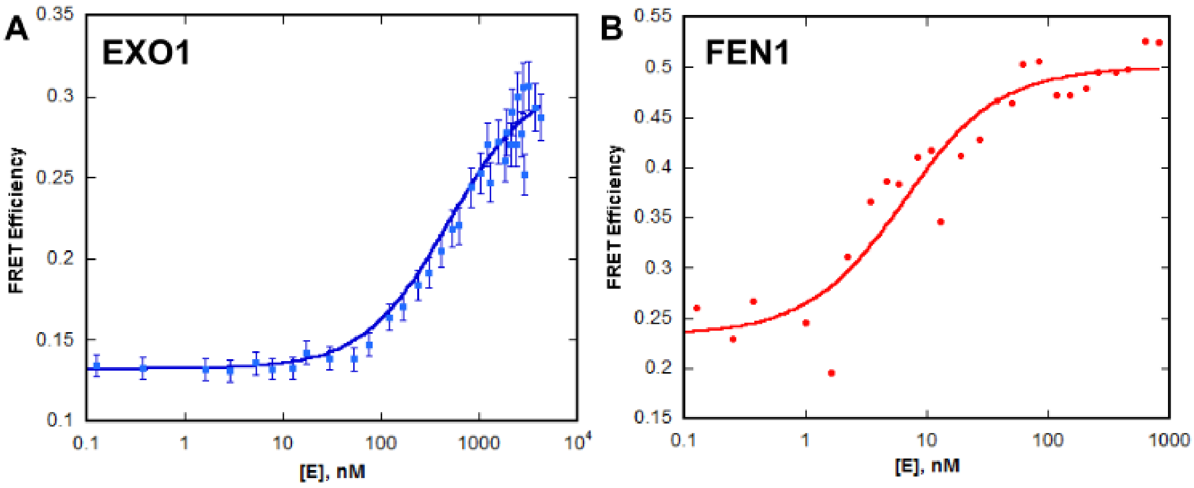


Figure 5.9: DNA constructs used for FRET. *The* ND_{NL}*,* ND_{DOL} *and* ND_{DAL} *substrates (**A**) and the* DF_{NL}*,* DF_{DOL} *and* DF_{DAL} *substrates (**B**) for hEXO1 and hFEN1 respectively (for strands and sequences, see* section 2.3*). With the reactive strand coloured cyan, the*

template/acceptor strand coloured brown and the upstream template/donor strand is coloured purple. The donor fluorescein dye is represented with the orange star, and the acceptor 5-carboxytetramethylrhodamine dye by a blue star. ***C.*** *A schematic illustrating the bending associated with enzymatic interaction with the substrate, which is expected to bring the two dyes closer together, increasing energy transfer efficiency.*

Spectra were taken over a range of enzyme concentrations, with the substrate concentrations being adjusted to account for the substrate being diluted when enzyme is added. The NL acts as a control by removing any change in signal due to the unlabelled substrate alone. The DOL and DAL substrates are required to determine the FRET change by the ratio-A methodology [109]. Experiments with 10 nM substrate are titrated with 0.1-5000 nM enzyme. Detection of fluorescence in both the FAM and TAMRA regions was used to determine the change in FRET efficiency with increasing enzyme concentrations, which was plotted using Kaleidagraph (synergy software). The plotted data was fit to the quadratic equation (*equation 2.7*) with the plots and some of the parameters displayed in *figure 5.10*.



C

	k_{bend}, nM	FRET efficiencies	
		E_{min}	E_{max}
EXO1	479 ± 70	0.13	0.31

	k_{bend}, nM	FRET efficiencies	
		E_{min}	E_{max}
FEN1	5.2 ± 1.5	0.23	0.50

Figure 5.10: Plots of FRET data for hEXO1 and hFEN1. *Plots of FRET efficiencies for hEXO1 (**A**) and hFEN1 (**B**) with their respective substrates and increasing enzyme concentrations; these efficiencies were determined by use of* equations 2.4-2.7 *(see* section 2.6). ***A.*** *EXO1 data with 10 nM $[S]_0$ and between 0.1-5000 nM [E].* ***B.*** *FEN1 control with 2.5 nM $[S]_0$ and between 0.1-1000 nM [E].* ***C.*** *The K_{bend}, E_{min} and E_{max} of the two fits in **A** and **B**.*

The K_{bend} is the rate of substrate bending when titrated with increasing concentrations of enzyme, and represents the enzyme concentration at half the max FRET efficiency. The K_{bend} value determined for hEXO1 is roughly 100-fold higher than the FEN1 control, which is a surprising result. Although no published crystallographic evidence demonstrates that hEXO1 bends duplex DNA, it is expected to do so as the nicked duplex is its preferred substrate. The key difficulty with determining the K_{bend} of hEXO1 in this case is the lack of a plateau in the data, with no end point being observed. In fact, increasing the enzyme concentrations lead to a decrease in the FRET efficiency, which could be due to collisional quenching, or possibly down to multiple enzyme binding to the same substrate. Therefore, further experiments with fluorescence anisotropy were attempted to provide a reliable binding constant.

Fluorescence anisotropy can be used to measure the tumbling of a fluorophore in solution. Vertical excitation of the sample before passing the light through an emission polariser in either the parallel (vertical) or perpendicular (horizontal) provides intensities that can be used to measure the anisotropy (r). Fluorophore labelled (specifically FAM) DNA constructs were titrated with increasing concentrations of enzyme. It is expected that as the enzyme concentration increases, the tumbling of the fluorophore will decrease as the substrate is saturated with enzyme. This can be measured by anisotropy in a fashion that mirrors the changes in the FRET experiments. The DNA constructs tested are shown in *figure 5.11*.

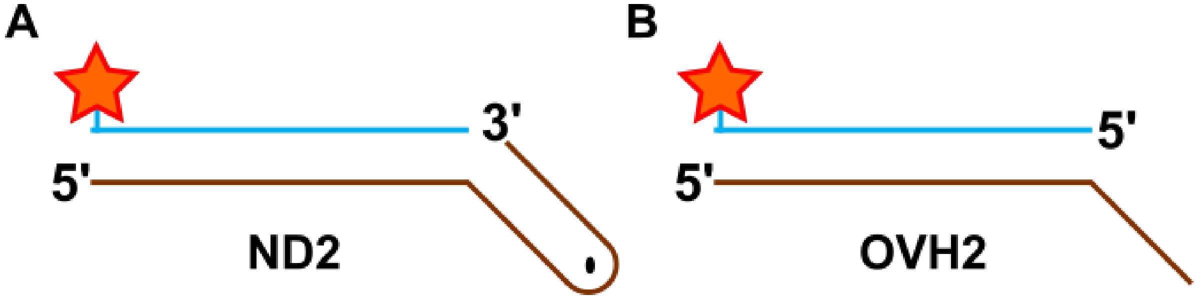


Figure 5.11: Constructs used in fluorescence anisotropy. *The ND2 (**A**) and OVH2 construct (**B**), which are both 3'-fluorescein labelled (see* section 2.3 *for strands and sequences). The dot represents the unimolecular turn in the DNA sequence.*

FA experiments were performed with an initial substrate concentration of 10 nM and titrated with increasing concentrations of enzyme (between 0.1 and 1000 nM). The total intensity (I_{TOT}), the emission of molecules in the same orientation to the polariser (I_{VV}), or the emission perpendicular to the polariser (I_{VH}) and a correction factor against the horizontal polarised emissions (G-factor) were determined using the Horiba Jobin Yvon FluoroMax-3 fluorimeter. These values were used to determine the anisotropy (r) using *equation 2.8*. The anisotropy was plotted against the enzyme concentration and fitted to a variation of the quadratic equation (*equation 2.8*) used with the FRET data (*figure 5.12*).

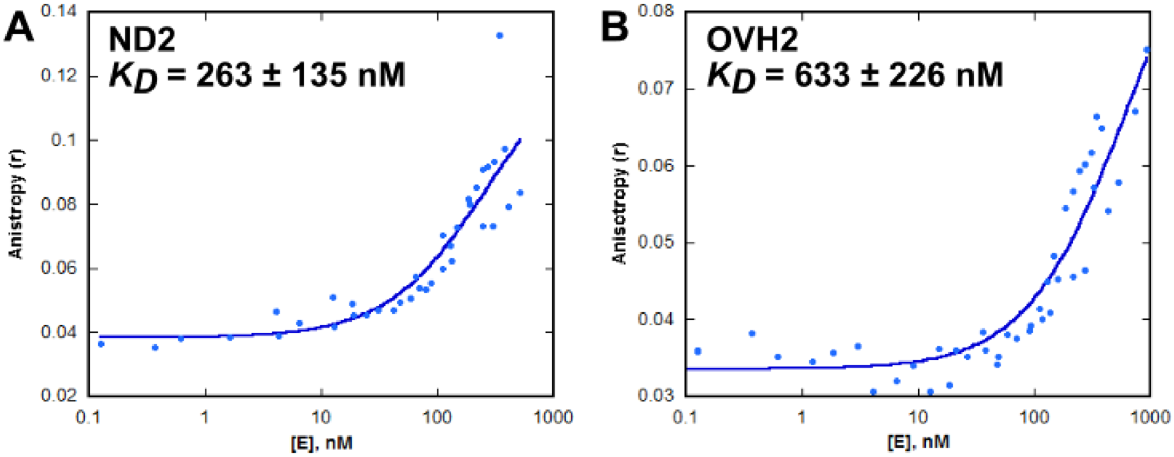


Figure 5.12: Anisotropy data for hEXO1. *The data and fits for ND2 (**A**) and OVH2 (**B**) experiments with the dissociation constant (K_D) of the respective fits.*

Experiments were also performed with the DOL nicked duplex in *figure 5.9A* but the data is not significant (the error is double the K_D value it predicted, possibly due to a lack of end point). The K_D determined by the fluorescence anisotropy experiments are all within the same order as well as within the same range as the K_{bend} determined by FRET for the ND_{DAL} substrate. It is predicted that these constants are exaggerated in some way. Other research groups have reported the K_D of hEXO1 to be similar to hFEN1, with a value around 20 nM, and a simple explanation may explain the differences [120].

Previously, it was suggested that the FRET efficiency may simply be disrupted by multiple enzyme molecules binding to the same substrate at different 5' ends. As highlighted in *figure 5.13A*, all the substrates tested have at least two 5' ends that are suitable for EXO1 binding. Binding of protein to the blunt end would not be predicted to induce a FRET change and could potentially be an obstacle to binding to the other 5’-end of the DNA junction that would result in enhanced FRET. The study mentioned [120] above uses single-molecule TIRF FRET to determine the K_{bend} of hEXO1 with a substrate with two 5' ends. However, the nature of the experiment meant that one of the ends was fixed to a surface before enzyme was added, which could potentially block the enzyme interacting with this end. This process is shown in *figure 5.13B* and suggests a suitable method to overcome this issue.

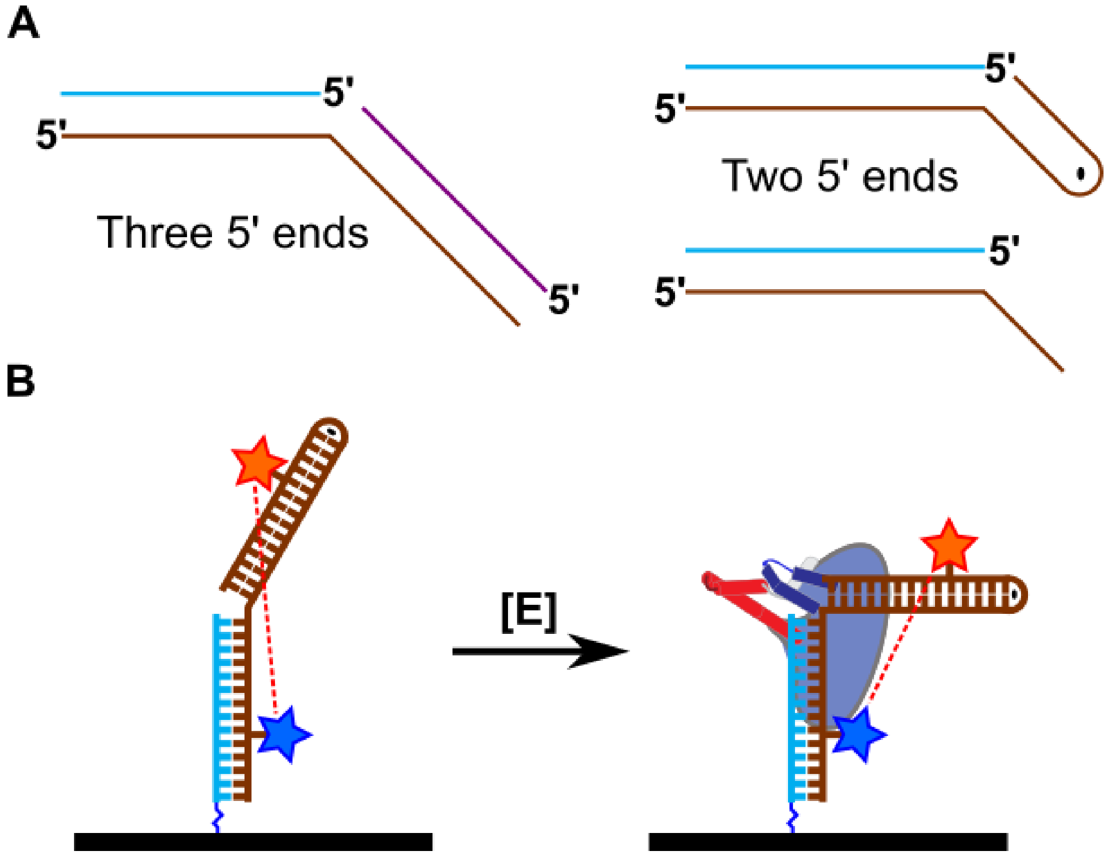


Figure 5.13: Overcoming multiple 5' ends. A. *The constructs used in both FRET and FA experiments have multiple 5' ends, which are suitable substrates of hEXO1.* ***B.*** *The reaction mechanism used by [120] to determine the* K_{bend} *of hEXO1, where only one 5' end is accessible, possibly providing more reliable values.*

Therefore, future work should be aimed at blocking the 5' ends to ensure precise binding at one position and bending of the DNA substrates. This approach may provide lower K_{bend} values that are comparable to hFEN1. Testing these substrates by fluorescence anisotropy should also be more reliable.

5.3 Summary of substrate dynamics

Exciton-coupled circular dichroism studies with hEXO1 do not provide evidence that an enzyme-induced local substrate conformational change that is observed in hFEN1 is shared by hEXO1. Further examination of the crystallographic structures of hEXO1 and hFEN1 modelled with B-DNA display that substrate distortion is required to process flapped substrates. However, when hEXO1 binds to an exonucleolytic substrate, it appears to stabilise the DNA strands before catalysis occurs. Förster resonance energy transfer and fluorescence anisotropy were utilised to determine the bending and dissociation constants of hEXO1, providing relatively high values compared to hFEN1. However, data from another study suggests that the binding constant for hEXO1 is more comparable to hFEN1 [120], which may be due to blocking interaction with the alternative 5' ends.

Chapter 6: Determination of the specificity of a range of N-hydroxyurea inhibitors for hFEN1

- 6.1 Identification of a group of N-hydroxyurea inhibitors for hFEN1
- 6.2 Determination of the compounds specificity for hFEN1
- 6.3 The effects of PCNA on FEN1 inhibition
- 6.4 Summary of the inhibition experiments

Chapter 6: Determination of the specificity of a range of N-hydroxyurea inhibitors for hFEN1

6.1 Identification of a group of N-hydroxyurea inhibitors for hFEN1

Human FEN1 is an important enzyme with roles in DNA replication and repair mechanisms, generally acting as the key nuclease. Due to its importance for DNA maintenance, it comes as no surprise that FEN1 has been observed as both overexpressed and under-expressed in cancer cells, allowing for either preservation of cell life or providing greater variation [16, 121-122]. Therefore, it would be an effective target for anti-cancer drugs by preventing cancer cells from growing. Early work in 2005 [111] was aimed at constructing compounds to inhibit the activity of hFEN1, whilst minimising the off-target effect of hXPG inhibition.

A high-throughput screen was performed to identify compounds that were capable of inhibiting hFEN1 [111]. The most potent compound was named ATH-0013974 and is shown in *figure 6.1A*. This compound has an affinity for hFEN1, with a reasonable IC_{50} (0.31 μM) and the IC_{50} for XPG is roughly 6-fold higher (1.98 μM). The structure of the compound was likened to Flutimide (*figure 6.1B*), an antiviral drug that targets the influenza endonuclease [123]. The proposed binding scheme of the two compounds relies on co-ordination of oxygens to the divalent metal ions in the active sites of many nucleases (*figure 6.1C*). Further experiments displayed that removal of the N-hydroxyl eliminated the inhibitory effect of ATH-0013974. The study continued to produce a range of N-hydroxyurea compounds based on ATH-0013974 to optimise hFEN1 inhibition, whilst reducing inhibition of XPG [123].

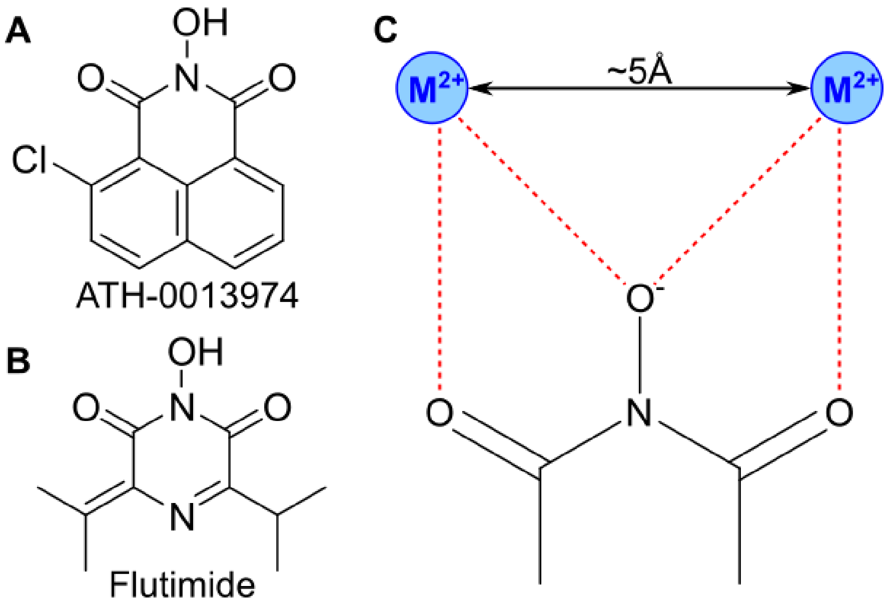


Figure 6.1: Compound identified by previous study and its mode of action. *The Structures of ATH-0013974 (**A**) and Flutimide (**B**). **C.** The predicted mode of action; the compounds co-ordinate to the divalent metal ions in the active sites of nucleases which rely on a two metal configuration.*

A compound identified in this study (henceforth referred to as compound 1, see *figure 6.2A*) demonstrated a 30-fold difference in IC_{50} for FEN1 against XPG, requiring small concentrations to dramatically inhibit hFEN1 (11 and 292 nM for FEN1 and XPG respectively). Cell studies with compound 1 showed that it complemented methyl methanesulfonate (a DNA alkylating agent) and temozolomide (a DNA methylating agent), producing cells that are sensitive to DNA damage [111]. Compound 1 was selected for further experiments with hFEN1. To provide an appropriate anticancer treatment by targeting hFEN1, it is essential that the drug be specific for the enzyme alone. The compound's affinity for XPG was previously demonstrated to be weak compared to FEN1, but new experiments with hEXO1, which is a closer relative to hFEN1, would determine whether the compound is suitable. Therefore, studies into whether compound 1, and another newly derived compound (compound 2) are hFEN1 specific were undertaken.

6.2 Determination of the compounds specificity for hFEN1

The two compounds tested with hFEN1 and hEXO1 are shown below:

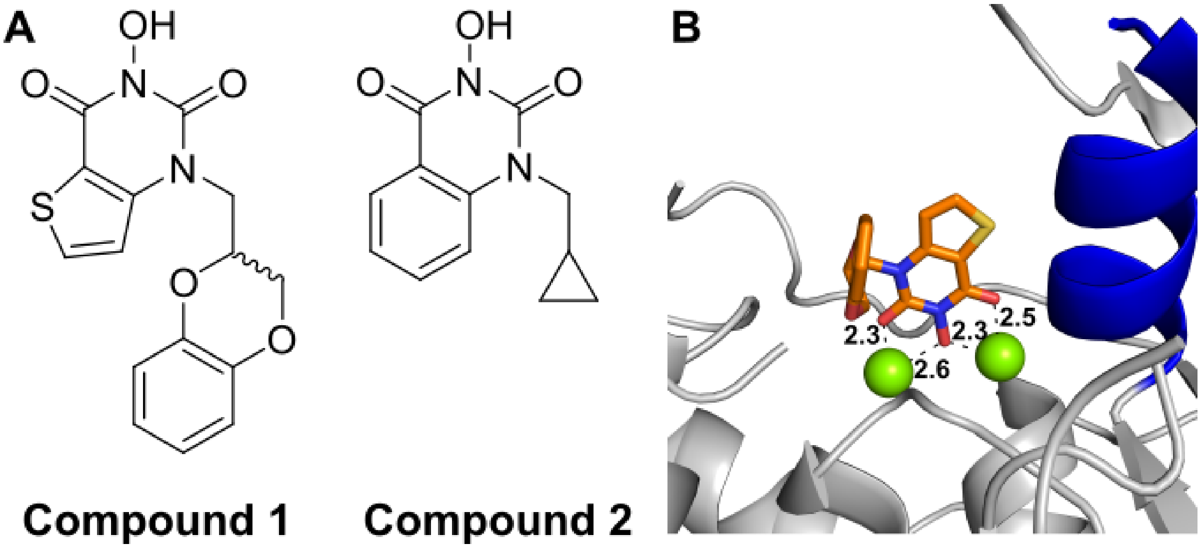


Figure 6.2: The compounds tested. A. *N-hydroxy based compounds 1 (athersys) and 2 (cyclopropylmethyl). **B.** Crystal structure of compound 1 in the active site of hFEN1 (5FV7), consistent with the hypothesised interactions in figure 6.1C.*

Work with hFEN1 was previously performed by another student [110, 112] with the double flap (DF) construct and the two compounds, demonstrating that they both successfully inhibit FEN1 activity. It was shown that the compounds have relatively similar IC_{50}'s (46.4 and 30.0 nM for compounds 1 and 2 respectively) [112]. To determine whether the compounds inhibit hFEN1 as an exonuclease, a single-nucleotide 3'-flapped 5'-FAM labelled substrate (SF1) was tested over a range of compound concentrations for comparison. The compounds reduced the activity of hFEN1, with the IC_{50} of compound 1 being essentially equal between the single and double flapped substrates. The IC_{50} of SF1 with compound 2 is roughly 15-fold higher, suggesting a weaker interaction. The data for hFEN1 and SF1 with the two compounds was analysed via dHPLC as described in *section 2.4.3*, and is displayed in *figure 6.3*.

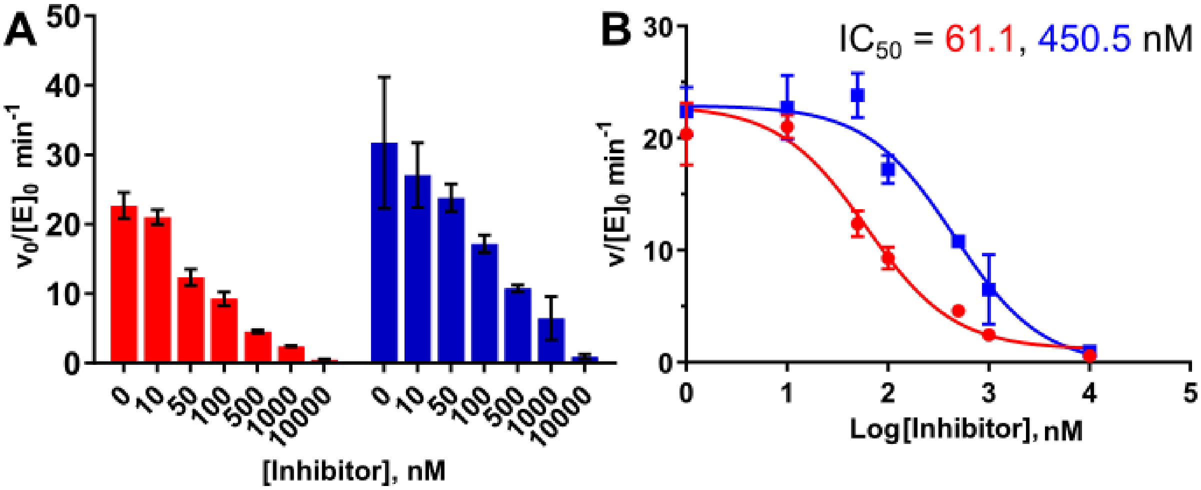


Figure 6.3: Data for hFEN1 with the 5' SF substrate and compounds 1 & 2. *The histogram bars are red and blue for compounds 1 and 2 respectively.* ***A.*** *Plot of the initial rates for 150 nM substrate with increasing concentrations of the inhibitors, the enzyme concentrations varied to allow for ~10% cleavage in 10 minutes (see* section 2.8*).* ***B.*** *The IC_{50} plot of the 5' SF with the two compounds, and the exact IC_{50} prediction is highlighted in the top right corner.*

To determine the specificity of the compounds for hFEN1, similar experiments to those performed in *figure 6.3* were performed with hEXO1 and OVH1 in the presence of the compounds. Four concentrations of the compounds were tested (0, 100, 1000 & 10000 nM) at a substrate concentration of 150 nM. The data is plotted in *figure 6.4* and shows a reduction in activity as the concentrations of the inhibitors increased. Approximate IC_{50} values were predicted to be similar to the values determined for hFEN1. Although, more concentrations of inhibitor would be required to precisely determine the IC_{50} values, the inhibitory effect of the compounds on hEXO1 is clear. Moreover, the inhibition of hEXO1 occurs at a similar concentration range of compound to the inhibition of hFEN1.

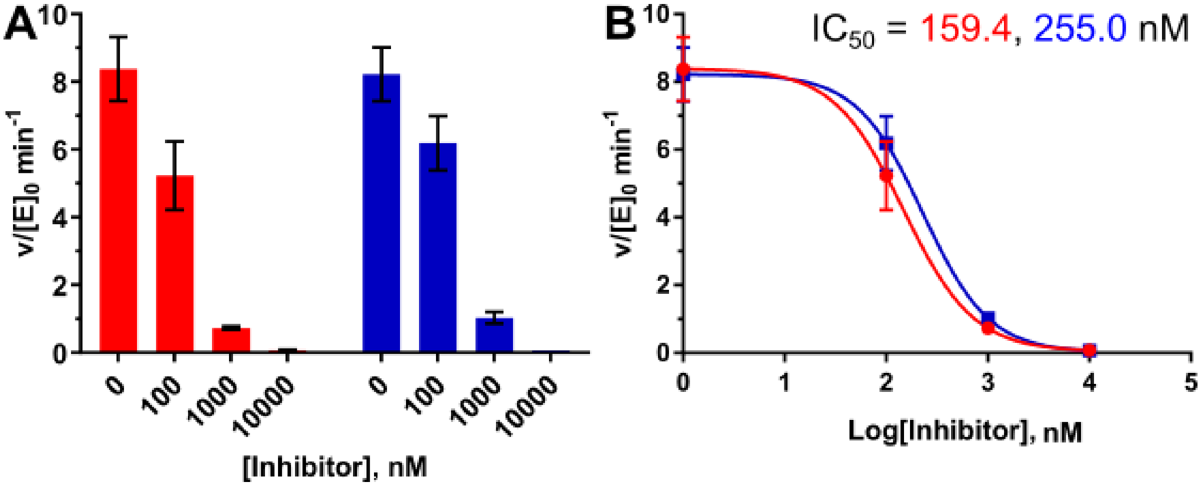


Figure 6.4: Data for hEXO1 with OVH1 and compounds 1 & 2. *The histogram bars are red and blue for compounds 1 and 2 respectively.* ***A.*** *Plot of the initial rates with 150 nM OVH1 with increasing concentrations of the inhibitors, the enzyme concentrations varied to allow for ~10% cleavage in 10 minutes (see* section 2.8*).* ***B.*** *The* IC_{50} *plot for the OVH with the two compounds, and the rough* IC_{50} *predictions are highlighted in the top corner.*

Therefore, it has been biochemically proven that the compounds are not specific to hFEN1 and also inhibit hEXO1 activity with a similar potency. To verify that the compounds directly interact with the enzymes and do not act via interaction with the DNA, differential scanning fluorimetry (DSF) was performed. DSF can be used to determine the melting temperature of the enzyme alone or in the presence of a compound by the use of Sypro Orange [124]. This dye can bind to the hydrophobic surface of proteins, but its emission is quenched by water. Therefore, as the protein melts the dye can interact with the surface and fluoresce, resulting in determination of changes in the melting temperatures by the change in fluorescence between the samples. This technique is similar to the *in vivo* cellular thermal shift assay technique (CETSA), which was also employed by our collaborators to demonstrate that the compounds interact with hFEN1 *in vivo* [112]. It is postulated that binding of a compound to the enzyme would increase the thermal stability of the enzyme, which would be represented by an increase in the melting temperature. Plots of the change in melting temperature (ΔT_m) against the compound concentrations are shown in *figure 6.5*.

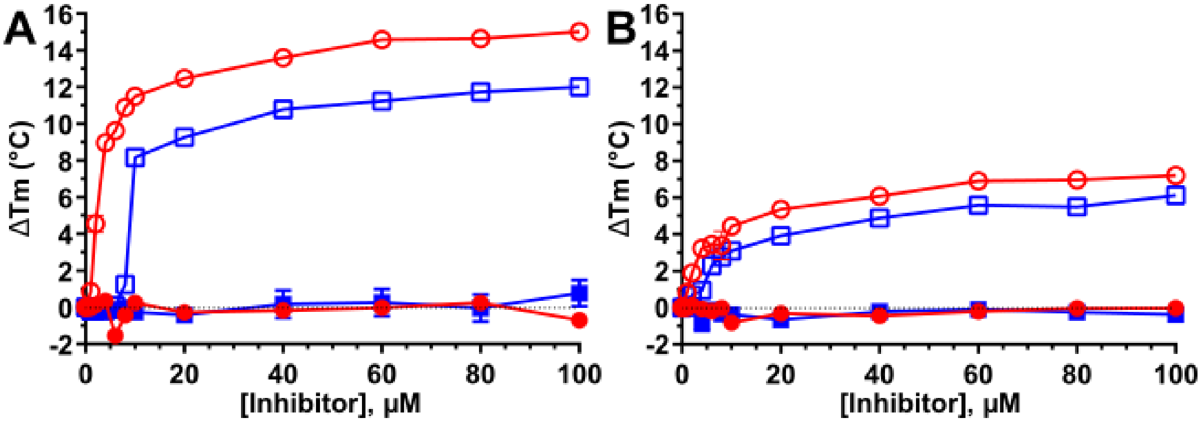


Figure 6.5: Differential scanning fluorimetry of hFEN1 and hEXO1. *Changes in melting temperature were plotted against the inhibitor concentration. Compounds 1 and 2 are represented by red circles and blue squares respectively, with data for hFEN1 (**A**) and hEXO1 (**B**). Experiments in the presence of Mg^{2+} ions are displayed with unfilled shapes, whereas experiments in the presence of EDTA are shown with filled shapes. Compound concentrations were used in a range where saturation of the enzyme is expected (little catalytic activity).*

The DSF experiments demonstrate that an interaction is occurring between the enzymes and the compounds, measured by an increase in the apparent melting temperature. It should be noted that this occurs in a magnesium-dependent manner, as the presence of EDTA appears to abolish interaction with the active site directed inhibitor. This observation is consistent with the compounds directly interacting and inhibiting catalysis by co-ordination with the divalent metal ions in the active site. It can be concluded that these inhibitors are not specific to hFEN1, and would probably be too potent for use as a cancer treatment due to their off-target effects.

6.3 The effects of PCNA on FEN1 inhibition

Our collaborators from Astrazeneca in this study determined an EC_{50} (effective concentration for half maximal response in cells, i.e. death) for compound 1 and compound 3 (IC_{50} = 16.9 nM) in SW620 colon cancer cells. Interestingly, compounds 1 and 3 had EC_{50}'s of 5.1 and 6.8 µM respectively, which are 100- and 400-fold higher for each compound compared to *in vitro* studies. Proliferating cell nuclear antigen (PCNA) forms a trimeric complex known to interact with a vast range of proteins involved in DNA replication, repair and metabolism. One such protein that PCNA is suggested to interact with is hFEN1, where it supposedly helps to stabilise FEN1's interaction with the DNA (*figure 6.6*). Therefore, it has been reported that

PCNA can stimulate the activity of hFEN1 [125-126]. If PCNA increases the residence time of the DNA substrate on hFEN1, so that it is greater than the residence time of the inhibitor this could explain the marked differences between cellular EC_{50}'s and *in vitro*-determined IC_{50}'s where PCNA was not present. Thus, to determine whether PCNA could be a factor in the in cellular efficacy of the inhibitors, FEN1 inhibition experiments were performed in the presence of PCNA. These experiments used an extended version of the double flapped duplex substrate (DF2 – see *section 2.3*), with a 3-fold larger upstream region prior to the 3'-flap (see *figure 6.6B*) which allows for PCNA loading onto the upstream region. Therefore, determination of the effect of compound 1 with DF2 was first performed, the data is shown in *figure 6.6*.

The data in *figure 6.6* demonstrates that the IC_{50} for hFEN1 with the DF2 and compound 1 construct is roughly 5-fold higher compared to hFEN1 with DF1. This is within a similar biological range and is still suitable for experiments with PCNA. Experiments with hFEN1, DF2 and compounds 1 and 3 were performed with increasing concentrations of PCNA. *In vivo*, PCNA requires loading onto the DNA by replication factor C (RFC). However, it was suggested for the same process to occur *in vitro* with excess PCNA without RFC [126]. Therefore, samples with a fixed concentration of 10 μM of either compounds 1 or 3, with increasing concentrations of PCNA (0, 100, 1000 & 10000 nM) to ensure effective loading onto the DNA. The data for this experiment, and the structure of compound 3, are displayed in *figure 6.7*.

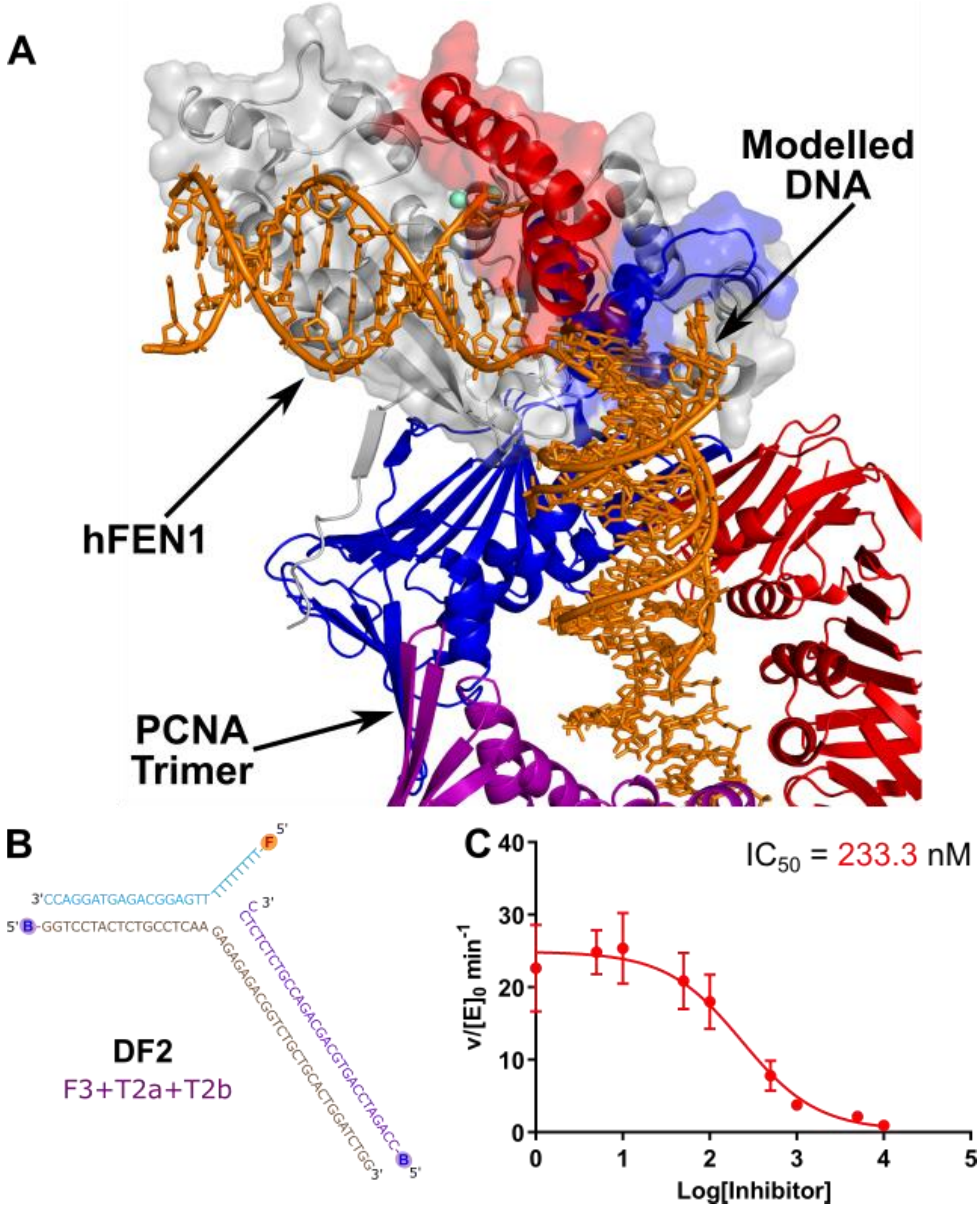


Figure 6.6: PCNA with FEN1 and data of hFEN1 with the DF2 and compound 1. A. *FEN1 and PCNA crystal sturctures (3Q8K and 1UL1) modelled together to give a representation of the expected interaction of PCNA and FEN1.* ***B.*** *The DF2 construct used in PCNA experiments.* ***C.*** *IC$_{50}$ plot for hFEN1 with DF2 and compound 1 with the exact IC$_{50}$ of the substrate in the top right corner.*

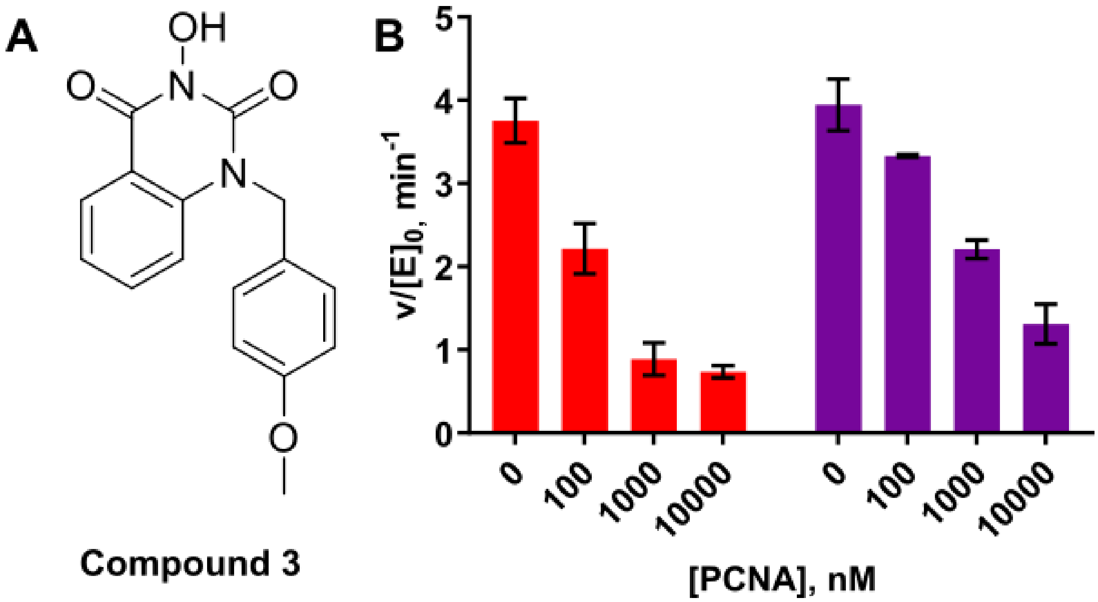


Figure 6.7: The effects of PCNA on FEN1 inhibition. A. *The structure of compound 3, N-hydroxy based 4-methoxybenzyl.* ***B.*** *The rate of 150 pM hFEN1 with 150 nM DF2 and increasing concentrations of PCNA, with 10 µM of either compound 1 (red) or 3 (purple). Increasing PCNA concentrations appeared to reduce activity of hFEN1.*

The data shows that PCNA does not protect hFEN1 from inhibition with the compounds *in vitro*; in fact PCNA appears to reduce the activity of hFEN1. No inhibitor-free PCNA control was performed, but as the PCNA-free sample has a similar rate to hFEN1 with DF2 and 10 µM of compound 1 (*figure 6.6C* – where log[inhibitor] = 4) it is assumed that the PCNA is not responsible for the drop in hFEN1 rate. These results contrast with earlier literature reports that suggest stimulation of the FEN1 reaction by PCNA. Inhibition by PCNA is an interesting observation, but may just be representative of an oversaturation of PCNA to the DNA substrate. For example, the binding of PCNA to the reacting duplex arm of the substrate rather than the 3’-flap side might be expected to inhibit the FEN1 reaction by inhibiting the interaction of protein and DNA. However, these results suggest that hFEN1 cannot be protected from inhibition by the addition of PCNA. It would be interesting to determine whether this is due to inefficient loading of the PCNA to the DNA substrate, but that would require future work.

6.4 Summary of the inhibition experiments

A range of N-hydroxyurea based inhibitors were previously developed and optimised for their ability to inhibit hFEN1 whilst having a minimal off-target effect on hXPG. One of these compounds was selected from this study, which has been crystallised with hFEN1 displaying that the compound co-ordinates directly with the divalent metal ions [112]. Studies with this compound (compound 1) and another later derived compound (compound 2) showed they had similar inhibitory effects on both the exonucleolytic and endonucleolytic activities of hFEN1. Further tests with hEXO1 demonstrated that the compounds are not specific to hFEN1, reducing the activity of hEXO1 with similar potency.

Differential scanning fluorimetry was used to confirm that direct interaction is necessary between the enzymes and the compounds. These experiments demonstrated that the compounds interact with hFEN1 and hEXO1 in a magnesium-dependent manner, consistent with the compounds co-ordinating in the active site. Finally, studies to determine the effect of PCNA stimulation on hFEN1 inhibition displayed that increasing concentrations of PCNA with a fixed inhibitor concentration of either compound 1 or 3 actually reduced the activity of hFEN1. Therefore, PCNA interaction does not protect against FEN1 inhibition *in vitro*.

Chapter 7: Conclusions

7.1 The conclusions of this research

7.2 Future work

7.3 Implications of this research on the flap endonuclease superfamily

Chapter 7: Conclusions

7.1 The conclusions of this research

The work detailed in this thesis is focussed at further understanding the mechanism of the FEN superfamily of 5' nucleases, which operate in DNA replication, repair and metabolism. To do this, extensive research primarily aimed at studying human exonuclease-1 (hEXO1) and making comparisons with what is already known about human flap endonuclease-1 (hFEN1). Early experiments determined the substrate specificity of hEXO1 followed by general kinetic studies of its *exo*- and *endo*-nucleolytic capabilities. Further work into whether there is a requirement for threading 5' flapped substrates and if DNA distortion is present, which are both precursors of hFEN1 catalysis, were tested in hEXO1. Finally, a group of inhibitors known to knock-down the activity of hFEN1 [111] were assayed with hEXO1 to determine if they are specific to hFEN1, for possible use in the treatment of cancer.

In vivo, hEXO1 is involved in multiple DNA repair mechanisms, including eukaryotic mismatch repair and double strand break repair [105]. In each pathway a different substrate is processed by hEXO1. Therefore, in chapter 3 the substrate preference of hEXO1 was determined, prior to further kinetic studies being undertaken. Three exonucleolytic substrates were tested, a blunt duplex, 3' overhang, and nicked duplex, each with either a 5' fluorescein (FAM) label or a 3' FAM label (with a 5' phosphate). The results suggested that the nicked duplex, which is the major target of mismatch repair, was the preferred substrate for hEXO1, consistent with both labelled constructs.

Determination of the Michaelis-Menten parameters for hEXO1 with the nicked duplex (with differing positions for fluorescent labels) provided k_{cat}'s of approximately 25 min^{-1} and K_m's of 8 and 18 nM (for 5' and 3' FAM labels respectively). Further experiments using quenched flow kinetics attained a single turnover k_{ST} rate of approximately 260 min^{-1}, which is roughly 10-fold higher than the k_{cat}. As highlighted in *figure 3.5*, k_{cat} encompasses the rates of the conformational change of the enzyme-substrate complex, followed by the chemical rate of the catalytic step, and finally the rate of product release. In contrast, the k_{STmax} is comprised of only the conformational change and chemistry step of the reaction mechanism. Therefore, as k_{ST} was 10-fold faster than the enzymes k_{cat} this implies that product release is a rate-limiting step in hEXO1 catalysis at high concentrations of substrate. This may be due to hEXO1 being a processive enzyme *in vivo* [116] (although it only has mild observable processivity *in vitro*) and as a result the enzyme may hold onto the major duplex product and continue to process it as a substrate. Previous studies with hFEN1 observed a comparable K_m value, but with a k_{cat} roughly 6.5-fold higher than determined for hEXO1. Overall the two enzymes have relatively similar k_{cat}/K_m values.

It has been suggested that hEXO1 can act as a loose back up for hFEN1 *in vivo*, acting in Okazaki fragment maturation [127-129]. Processing the ends of Okazaki fragments requires endonucleolytic capability, removing a long flapped structure (as opposed to single nucleotides). In chapter 4, the endonuclease capability of hEXO1 was tested by use of pseudo-Y constructs, equipped with either a 5' or 3' FAM label (as before). The results showed that the endonucleolytic activity of hEXO1 is roughly 35-fold slower than exonucleolytic processing of the nicked duplex. However, use of a unimolecular substrate, which only has a flapped 5' end with elimination of a duplex 5' end, provided a rate much more comparable to the nicked duplex (only 4-fold down). Comparing the exonuclease and endonuclease capabilities of hEXO1 and hFEN1 suggests that the two enzymes have similar exonuclease activity, but hFEN1 has a roughly 70-fold higher endonuclease capability.

The mode of 5' flap positioning was an area of controversy in FEN1 research for many years, which has been proven biochemically and crystallographically to use a threading mechanism, where the 5' flap is threaded through a helical gateway and under the helical cap [99, 101]. Similar biochemical experiments to those used to study hFEN1 were performed with hEXO1 in chapter 4. These relied on either trapping a biotinylated pseudo-Y construct threaded through the helical arch of the enzyme with streptavidin (which forms a high affinity interaction with biotin) or blocking the substrate from adopting the threaded state by preincubation with streptavidin prior to the addition of the enzyme. The results provide interesting insights into the requirement for hEXO1 to thread flapped substrates, which was essential for catalysis on a biologically relevant time scale. Attempts to compete away this threaded state with an excess of unlabelled exonucleolytic construct were unsuccessful, implying that threading does indeed occur. However, when it was attempted to trap the substrate on hEXO1, the data were biphasic in nature; only 20% of the substrate reacted quickly and was therefore trapped, and the remainder behaved as though it was blocked from reaction on a fast timescale. This is postulated to be due to an on-enzyme equilibrium for flap threading. This contrasts with hFEN1 where 100% of the substrate adopts a trapped state.

Another laboratory also published crystal structures of hEXO1 in a threaded state, which backs up this biochemical data [100]. Therefore, hEXO1, like hFEN1, must thread flapped substrates before catalysis can occur.

Building on the requirement for hFEN1 to thread 5' flaps, residues were identified by their positioning in the threaded crystal structures [99] based on possible interactions they make with the +1 phosphate, which may stabilise the threaded state. Various basic FEN1 residues were mutated to alanine residues, removing their associated charges before testing the effect with a single 3' flapped substrate with either a phosphate or hydroxyl at the 5' end of the reacting strand. Both single and double mutant variants were examined and it was determined that R104A and K132A have similar lower rates when processing a 5' hydroxyl compared with wild type hFEN1, suggesting that they both interact directly with the +1 phosphate. This effect was additive in the R104A-K132A double mutant, and when the residues were mutated to glutamate it was further detrimental. It should be noted that other residues tested (R103 and R129) also reduced the activity of hFEN1 with the substrate with a 5' phosphate, and activity reduction was more severe with the 5' hydroxyl. R104 is conserved through the superfamily and may be required for phosphate stabilisation; however, K132 is only present in hFEN1. The lack of conservation of K132 in the superfamily suggests its importance in maintaining the precision of hFEN1 catalysis one nucleotide into the duplex.

Previous studies suggested that human FEN1 induces double nucleotide unpairing on its DNA substrates to facilitate positioning of the 5' flap in the active site [16, 118], which is expected to cause a change in ECCD signal using substrates containing tandem 2-amino purine nucleotides. However, later crystallographic data provided evidence that the DNA is not unpaired as expected; instead hFEN1 distorts the DNA to position it in the active site [99]. This distortion was expected to be the cause of the changes in ECCD signal. In chapter 5, attempts to duplicate this signal change in hEXO1 showed little change, if any, with a range of exonuclease and endonuclease substrates with tandem 2-aminopurines. However, examination of the local nucleotide stacking in hEXO1 also displayed the nucleobase distortion observed in hFEN1 crystals. Further examination of the DNA from various published hEXO1 crystal structures displayed small amounts of distortion in the DNA strands of exonuclease substrates, appearing similar to B-form DNA. In fact, the later the stage of the catalytic cycle that the crystal was in, the more stable the DNA backbone appeared. On the other hand, EXO1 crystals with an endonuclease substrate demonstrate more distortion before catalysis, in line with hFEN1 crystals. It is postulated that hEXO1 stabilises the DNA strands of exonuclease substrates allowing for consecutive hydrolysis in a processive manner. Alternatively, a flapped substrate may require DNA strand distortion to force the flap through the helical arch. It remains unclear what the cause of the ECCD signal shifts in hFEN1 is at this time.

To better understand the conformational change step of catalysis, which is expected in FEN superfamily members, substrate dynamic studies were performed in chapter 5. Determination of K_{bend} and K_D values for hEXO1 were achieved using Förster resonance energy transfer (FRET) and fluorescence anisotropy (FA) respectively. Both techniques demonstrated values of approximately 500 nM for K_{bend} or K_D, which suggesting that the enzyme binds to and bends its substrates weakly. This was unexpected as rate constants determined for hFEN1 are much lower [118]. Another study acquired K_D values for hEXO1 that were much more comparable to hFEN1 by blocking additional 5'-ends from EXO1 interaction [120]. In chapter 4, it was demonstrated that providing a singular 5' end raised the rate of hEXO1 with a pseudo-Y construct by 10-fold. Therefore, it is plausible that hEXO1 is interacting with the multiple 5' ends, preventing binding (and bending) of the substrate in the expected site, resulting in a much larger value for K_{bend} and K_D. However, further work is required to investigate this hypothesis.

A range of N-hydroxyurea-based compounds were previously shown to specifically inhibit hFEN1 compared to hXPG [111]. Further development of these compounds provided two compounds with decent potency for hFEN1, and little effect on hXPG [112]. However, EXO1 is more closely related to FEN1 than XPG is; therefore, determination of specificity to FEN1 over EXO1 was important. In chapter 6, kinetic experiments with hEXO1 over a range of inhibitor concentrations demonstrated a comparable decrease in rate to that of hFEN1 with both compounds. Further studies with differential scanning fluorimetry, a technique which determines the change in thermal stability of an enzyme by comparison of samples in the presence and absence of a compound. Both hFEN1 and hEXO1 showed increases in thermal stability with increasing concentrations of both compounds, in a magnesium-dependent manner, consistent with the compounds binding the metal ions in the active site. Therefore, these inhibitors are not specific to hFEN1.

Experiments using compound 1 and 3 were tested on a human cancer cell line by a collaborator. However, these studies displayed roughly 100- and 400-fold higher EC_{50} than the IC_{50}'s determined *in vitro* [112]. In the literature it has been suggested that proliferating cell nuclear antigen (PCNA) is capable of enhancing the activity of hFEN1 *in vivo* [125-126]. In chapter 6, experiments with hFEN1, a double flapped substrate and either compound 1 or 2 at fixed concentrations were assayed with increasing concentrations of PCNA to determine whether it was capable of protecting FEN1 from inhibition. Intriguingly, this experiment demonstrated that hFEN1 is not protected from inhibition *in vitro*, and also displays a reduction in activity with increasing PCNA concentrations, which may be due to oversaturation of the DNA substrate. Therefore, interaction with PCNA appears unlikely to be the responsible for this higher EC_{50} value.

7.2 Future work

Future work needed as a result of this project includes developing a better understanding of hEXO1-352 as a processive enzyme. Also tests to determine whether unimolecular substrates have a faster exonuclease rate compared to their bimolecular counterparts. These two results could determine whether the observed Michaelis-Menten parameters are representative of the full activity of hEXO1. Secondly, replication of the key mutants in hFEN1 used to examine +1 phosphate interactions would be interesting to test with hEXO1. Particularly, the R104 equivalent, which is conserved throughout the superfamily, may provide useful insight as it directly interacts with the +1 phosphate [16, 99]. However, K132 is not conserved in hEXO1. Thirdly, determination of the cause of the CD shift in hFEN1 which is not observed in hEXO1 would be important (*section 5.1*). Finally, blocking the alternative 5' ends of the FRET and FA substrates may settle whether they are the cause of the high K_D and K_{bend} values determined for hEXO1 (*section 5.2*), or if they are in fact accurate results. This is only the beginning of research into hEXO1 and future work will be even more insightful than what has been demonstrated in this thesis.

7.3 Implications of this research on the flap endonuclease superfamily

The *exo*- and *endo*-nucleolytic rate of hEXO1 and hFEN1 are representative of their biological roles, and may not be representative of the FEN superfamily. The superfamily can be divided into two subclasses of enzyme; those that process continuous or discontinuous substrates. EXO1 and FEN1 hydrolyse discontinuous substrates with bifurcated constructs, therefore, it is unsurprising that both enzymes are capable of threading flapped substrates. However, as GEN1 and XPG process continuous DNA substrates, they are not expected to have threading capabilities. It would be interesting to examine the effect of the conserved +1 phosphate residues in hEXO1 and the other members of the superfamily.

Equally, studies examining the ECCD signal changes with GEN1 and XPG may be more fruitful compared to EXO1 due to their ability to pass the substrate between their helical wedged regions. This would be especially interesting with hGEN1, which acts as a dimer *in vivo*. Finally, compound 1 which was assayed with hFEN1 and hEXO1 is known to inhibit XPG with a lower affinity [111], which is also expected for compound's 2 and 3. As these compounds co-ordinate to the active site metals of FEN1 and EXO1, it is hypothesised that GEN1 will also be inhibited due to its similar catalytic architecture, but testing would be required. However, the effect of these compounds may differ due to GEN1 acting as a dimer on its substrate. Therefore, further experiments with the compounds and the remaining superfamily members may be interesting.

References

1. Avery, O. T.; MacLeod, C. M.; McCarty, M., Studies on the chemical nature of the substance inducing transformation of Pneumococcal types. *Journal of Experimental Medicine* **1944,** *79* (2), 137-158.
2. Hershey, A. D.; Chase, M., Independent functions of viral protein and nucleic acid in growth of bacteriophage. *The Journal of General Physiology* **1952,** *36* (1), 39-56.
3. Lodish, H.; Berk, A.; Zipursky, S. L.; Matsudaira, P.; Baltimore, D.; Darnell, J., The three roles of RNA in protein synthesis. *Book* **2000**.
4. Watson, J. D.; Crick, F. H. In The structure of DNA, Cold Spring Harbor Symposia on Quantitative Biology, Cold Spring Harbor Laboratory Press: 1953; pp 123-131.
5. Williams, N. H.; Wyman, P., Base catalysed phosphate diester hydrolysis. *Chemical Communications* **2001,** (14), 1268-1269.
6. Pingoud, A.; Jeltsch, A., Structure and function of type II restriction endonucleases. *Nucleic Acids Research* **2001,** *29* (18), 3705-3727.
7. Thomas, M.; Davis, R. W., Studies on the cleavage of bacteriophage lambda DNA with EcoRI Restriction endonuclease. *Journal of Molecular Biology* **1975,** *91* (3), 315-328.
8. Lieber, M. R., The FEN-1 family of structure-specific nucleases in eukaryotic DNA replication, recombination and repair. *Bioessays* **1997,** *19* (3), 233-240.
9. Kunkel, T. A.; Burgers, P. M., Dividing the workload at a eukaryotic replication fork. *Trends in Cell Biology* **2008,** *18* (11), 521-527.
10. McElhinny, S. A. N.; Gordenin, D. A.; Stith, C. M.; Burgers, P. M.; Kunkel, T. A., Division of labor at the eukaryotic replication fork. *Molecular Cell* **2008,** *30* (2), 137-144.
11. Burgers, P. M., Polymerase dynamics at the eukaryotic DNA replication fork. *Journal of Biological Chemistry* **2009,** *284* (7), 4041-4045.
12. Kao, H.-I.; Bambara, R. A., The protein components and mechanism of eukaryotic Okazaki fragment maturation. *Critical Reviews in Biochemistry and Molecular Biology* **2003,** *38* (5), 433-452.
13. Zheng, L.; Shen, B., Okazaki fragment maturation: nucleases take centre stage. *Journal of Molecular Cell Biology* **2011,** *3* (1), 23-30.
14. Waga, S.; Bauer, G.; Stillman, B., Reconstitution of complete SV40 DNA replication with purified replication factors. *Journal of Biological Chemistry* **1994,** *269* (14), 10923-10934.
15. Sparks, J. L.; Chon, H.; Cerritelli, S. M.; Kunkel, T. A.; Johansson, E.; Crouch, R. J.; Burgers, P. M., RNase H2-initiated ribonucleotide excision repair. *Molecular Cell* **2012,** *47* (6), 980-986.
16. Tsutakawa, Susan E.; Classen, S.; Chapados, Brian R.; Arvai, A. S.; Finger, L. D.; Guenther, G.; Tomlinson, Christopher G.; Thompson, P.; Sarker, Altaf H.; Shen, B.; Cooper, Priscilla K.; Grasby, Jane A.; Tainer, John A., Human Flap Endonuclease Structures, DNA Double-Base Flipping, and a Unified Understanding of the FEN1 Superfamily. *Cell* **2011,** *145* (2), 198-211.
17. Sullivan, M.; Morgan, D. O., Finishing mitosis, one step at a time. *Nature reviews. Molecular cell biology* **2007,** *8* (11), 894.
18. Marston, A. L.; Amon, A., Meiosis: cell-cycle controls shuffle and deal. *Nature Reviews of Molecular Cell Biology* **2004,** *5* (12), 983-997.
19. Clancy, S., Genetic recombination. *Nature Education* **2008,** *1* (1), 40.

20. Wyatt, H. D.; Sarbajna, S.; Matos, J.; West, S. C., Coordinated actions of SLX1-SLX4 and MUS81-EME1 for Holliday junction resolution in human cells. *Molecular Cell* **2013,** *52* (2), 234-247.

21. Matos, J.; West, S. C., Holliday junction resolution: regulation in space and time. *DNA Repair* **2014,** *19*, 176-181.

22. Ip, S. C.; Rass, U.; Blanco, M. G.; Flynn, H. R.; Skehel, J. M.; West, S. C., Identification of Holliday junction resolvases from humans and yeast. *Nature* **2008,** *456* (7220), 357.

23. Bocquet, N.; Bizard, A. H.; Abdulrahman, W.; Larsen, N. B.; Faty, M.; Cavadini, S.; Bunker, R. D.; Kowalczykowski, S. C.; Cejka, P.; Hickson, I. D.; Thomä, N. H., Structural and mechanistic insight into Holliday junction dissolution by Topoisomerase IIIα and RMI1. *Nature Structural & Molecular Biology* **2014,** *21* (3), 261-268.

24. Nimonkar, A. V.; Genschel, J.; Kinoshita, E.; Polaczek, P.; Campbell, J. L.; Wyman, C.; Modrich, P.; Kowalczykowski, S. C., BLM–DNA2–RPA–MRN and EXO1–BLM–RPA–MRN constitute two DNA end resection machineries for human DNA break repair. *Genes & development* **2011,** *25* (4), 350-362.

25. Reha-Krantz, L. J., DNA polymerase proofreading: Multiple roles maintain genome stability. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* **2010,** *1804* (5), 1049-1063.

26. Kolodner, R. D., Mismatch repair: mechanisms and relationship to cancer susceptibility. *Trends in Biochemical Sciences* **1995,** *20* (10), 397-401.

27. Peltomaki, P., Role of DNA mismatch repair defects in the pathogenesis of human cancer. *Journal of Clinical Oncology* **2003,** *21* (6), 1174-1179.

28. Hughes, M.; D'Arrigo, A.; Truong, O.; Hsuan, J. J.; Jiricny, J., GTBP, a 160-Kilodalton Protein Essential for Mismatch-BindingActivity in Human Cells. *Science* **1995,** *268*, 30.

29. Drummond, J. T.; Li, G.-M.; Longley, M. J.; Modrich, P., Isolation of an hMSH2-p160 heterodimer that restores DNA mismatch repair to tumor cells. *Science* **1995**, 1909-1909.

30. Acharya, S.; Wilson, T.; Gradia, S.; Kane, M. F.; Guerrette, S.; Marsischky, G. T.; Kolodner, R.; Fishel, R., hMSH2 forms specific mispair-binding complexes with hMSH3 and hMSH6. *Proceedings of the National Academy of Sciences* **1996,** *93* (24), 13629-13634.

31. Dufner, P.; Marra, G.; Räschle, M.; Jiricny, J., Mismatch recognition and DNA-dependent stimulation of the ATPase activity of hMutSα is abolished by a single mutation in the hMSH6 subunit. *Journal of Biological Chemistry* **2000,** *275* (47), 36550-36555.

32. Gradia, S.; Acharya, S.; Fishel, R., The human mismatch recognition complex hMSH2-hMSH6 functions as a novel molecular switch. *Cell* **1997,** *91* (7), 995-1005.

33. Genschel, J.; Littman, S. J.; Drummond, J. T.; Modrich, P., Isolation of MutSβ from human cells and comparison of the mismatch repair specificities of MutSβ and MutSα. *Journal of Biological Chemistry* **1998,** *273* (31), 19895-19901.

34. Cannavo, E.; Marra, G.; Sabates-Bellver, J.; Menigatti, M.; Lipkin, S. M.; Fischer, F.; Cejka, P.; Jiricny, J., Expression of the MutL homologue hMLH3 in human cells and its role in DNA mismatch repair. *Cancer Research* **2005,** *65* (23), 10759-10766.

35. Räschle, M.; Marra, G.; Nyström-Lahti, M.; Schär, P.; Jiricny, J., Identification of hMutLβ, a heterodimer of hMLH1 and hPMS1. *Journal of Biological Chemistry* **1999,** *274* (45), 32368-32375.

36. Blackwell, L. J.; Wang, S.; Modrich, P., DNA chain length dependence of formation and dynamics of hMutSα· hMutLα· heteroduplex complexes. *Journal of Biological Chemistry* **2001,** *276* (35), 33233-33240.

37. Kleczkowska, H. E.; Marra, G.; Lettieri, T.; Jiricny, J., hMSH3 and hMSH6 interact with PCNA and colocalize with it to replication foci. *Genes & Development* **2001,** *15* (6), 724-736.

38. Lau, P. J.; Kolodner, R. D., Transfer of the MSH2· MSH6 complex from proliferating cell nuclear antigen to mispaired bases in DNA. *Journal of Biological Chemistry* **2003,** *278* (1), 14-17.

39. Dzantiev, L.; Constantin, N.; Genschel, J.; Iyer, R. R.; Burgers, P. M.; Modrich, P., A defined human system that supports bidirectional mismatch-provoked excision. *Molecular Cell* **2004,** *15* (1), 31-41.

40. Johnson, A.; O'Donnell, M., Cellular DNA replicases: components and dynamics at the replication fork. *Annual Review of Biochemistry* **2005,** *74*, 283-315.

41. Genschel, J.; Modrich, P., Mechanism of 5′-directed excision in human mismatch repair. *Molecular Cell* **2003,** *12* (5), 1077-1086.

42. Kadyrov, F. A.; Dzantiev, L.; Constantin, N.; Modrich, P., Endonucleolytic function of MutLα in human mismatch repair. *Cell* **2006,** *126* (2), 297-308.

43. Waga, S.; Stillman, B., Cyclin-dependent kinase inhibitor p21 modulates the DNA primer-template recognition complex. *Molecular and Cellular Biology* **1998,** *18* (7), 4177-4187.

44. Umar, A.; Buermeyer, A. B.; Simon, J. A.; Thomas, D. C.; Clark, A. B.; Liskay, R. M.; Kunkel, T. A., Requirement for PCNA in DNA mismatch repair at a step preceding DNA resynthesis. *Cell* **1996,** *87* (1), 65-73.

45. Guo, S.; Presnell, S. R.; Yuan, F.; Zhang, Y.; Gu, L.; Li, G.-M., Differential requirement for proliferating cell nuclear antigen in 5′ and 3′ nick-directed excision in human mismatch repair. *Journal of Biological Chemistry* **2004,** *279* (17), 16912-16917.

46. Desai, A.; Gerson, S., Exo1 independent DNA mismatch repair involves multiple compensatory nucleases. *DNA Repair* **2014,** *21*, 55-64.

47. Symington, L. S.; Gautier, J., Double-strand break end resection and repair pathway choice. *Annual Review of Genetics* **2011,** *45*, 247-271.

48. Lieber, M. R., The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annual Review of Biochemistry* **2010,** *79*, 181-211.

49. Dynan, W. S.; Yoo, S., Interaction of Ku protein and DNA-dependent protein kinase catalytic subunit with nucleic acids. *Nucleic Acids Research* **1998,** *26* (7), 1551-1559.

50. Shim, E. Y.; Chung, W. H.; Nicolette, M. L.; Zhang, Y.; Davis, M.; Zhu, Z.; Paull, T. T.; Ira, G.; Lee, S. E., Saccharomyces cerevisiae Mre11/Rad50/Xrs2 and Ku proteins regulate association of Exo1 and Dna2 with DNA breaks. *The EMBO Journal* **2010,** *29* (19), 3370-3380.

51. Buis, J.; Wu, Y.; Deng, Y.; Leddon, J.; Westfield, G.; Eckersdorff, M.; Sekiguchi, J. M.; Chang, S.; Ferguson, D. O., Mre11 nuclease activity has essential roles in DNA repair and genomic stability distinct from ATM activation. *Cell* **2008,** *135* (1), 85-96.

52. Sartori, A. A.; Lukas, C.; Coates, J.; Mistrik, M.; Fu, S.; Bartek, J.; Baer, R.; Lukas, J.; Jackson, S. P., Human CtIP promotes DNA end resection. *Nature* **2007,** *450* (7169), 509.

53. Brosh, R. M.; Li, J.-L.; Kenny, M. K.; Karow, J. K.; Cooper, M. P.; Kureekattil, R. P.; Hickson, I. D.; Bohr, V. A., Replication protein A physically interacts with the Bloom's syndrome protein and stimulates its helicase activity. *Journal of Biological Chemistry* **2000,** *275* (31), 23500-23508.

54. Nimonkar, A. V.; Özsoy, A. Z.; Genschel, J.; Modrich, P.; Kowalczykowski, S. C., Human exonuclease 1 and BLM helicase interact to resect DNA and initiate DNA repair. *Proceedings of the National Academy of Sciences* **2008,** *105* (44), 16906-16911.
55. Wu, L.; Hickson, I. D., The Bloom's syndrome helicase suppresses crossing over during homologous recombination. *Nature* **2003,** *426* (6968), 870.
56. Migliore, L.; Coppedè, F., Environmental-induced oxidative stress in neurodegenerative disorders and aging. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis* **2009,** *674* (1), 73-84.
57. Jastroch, M.; Divakaruni, A. S.; Mookerjee, S.; Treberg, J. R.; Brand, M. D., Mitochondrial proton and electron leaks. *Essays in Biochemistry* **2010,** *47,* 53-67.
58. Grollman, A. P.; Moriya, M., Mutagenesis by 8-oxoguanine: an enemy within. *Trends in Genetics* **1993,** *9* (7), 246-249.
59. Kunisada, M.; Sakumi, K.; Tominaga, Y.; Budiyanto, A.; Ueda, M.; Ichihashi, M.; Nakabeppu, Y.; Nishigori, C., 8-Oxoguanine formation induced by chronic UVB exposure makes Ogg1 knockout mice susceptible to skin carcinogenesis. *Cancer Research* **2005,** *65* (14), 6006-6010.
60. Bruskov, V. I.; Malakhova, L. V.; Masalimov, Z. K.; Chernikov, A. V., Heat-induced formation of reactive oxygen species and 8-oxoguanine, a biomarker of damage to DNA. *Nucleic Acids Research* **2002,** *30* (6), 1354-1363.
61. Chen, D. S.; Herman, T.; Demple, B., Two distinct human DNA diesterases that hydrolyze 3′-blocking deoxyribose fragments from oxidized DNA. *Nucleic Acids Research* **1991,** *19* (21), 5907-5914.
62. Abbotts, R.; Madhusudan, S., Human AP endonuclease 1 (APE1): from mechanistic insights to druggable target in cancer. *Cancer Treatment Reviews* **2010,** *36* (5), 425-435.
63. Prasad, R.; Shock, D. D.; Beard, W. A.; Wilson, S. H., Substrate channeling in mammalian base excision repair pathways: passing the baton. *Journal of Biological Chemistry* **2010,** *285* (52), 40479-40488.
64. Liu, Y.; Wilson, S. H., DNA base excision repair: a mechanism of trinucleotide repeat expansion. *Trends in Biochemical Sciences* **2012,** *37* (4), 162-172.
65. Liu, Y.; Beard, W. A.; Shock, D. D.; Prasad, R.; Hou, E. W.; Wilson, S. H., DNA polymerase β and flap endonuclease 1 enzymatic specificities sustain DNA synthesis for long patch base excision repair. *Journal of Biological Chemistry* **2005,** *280* (5), 3665-3674.
66. Wilson, S. H.; Beard, W. A.; Shock, D. D.; Batra, V. K.; Cavanaugh, N. A.; Prasad, R.; Hou, E. W.; Liu, Y.; Asagoshi, K.; Horton, J. K., Base excision repair and design of small molecule inhibitors of human DNA polymerase β. *Cellular and Molecular Life Sciences* **2010,** *67* (21), 3633-3647.
67. Farrington, S. M.; Tenesa, A.; Barnetson, R.; Wiltshire, A.; Prendergast, J.; Porteous, M.; Campbell, H.; Dunlop, M. G., Germline susceptibility to colorectal cancer due to base-excision repair gene defects. *The American Journal of Human Genetics* **2005,** *77* (1), 112-119.
68. Andrew, S. E.; Goldberg, Y. P.; Kremer, B.; Telenius, H.; Theilmann, J.; Adam, S.; Starr, E.; Squitieri, F.; Lin, B.; Kalchman, M. A., The relationship between trinucleotide (CAG) repeat length and clinical features of Huntington's disease. *Nature Genetics* **1993,** *4* (4), 398-403.
69. Duyao, M.; Ambrose, C.; Myers, R.; Novelletto, A.; Persichetti, F.; Frontali, M.; Folstein, S.; Ross, C.; Franz, M.; Abbott, M., Trinucleotide repeat length instability and age of onset in Huntington's disease. *Nature Genetics* **1993,** *4* (4), 387-392.

70. Liu, Y.; Prasad, R.; Beard, W. A.; Hou, E. W.; Horton, J. K.; McMurray, C. T.; Wilson, S. H., Coordination between polymerase β and FEN1 can modulate CAG repeat expansion. *Journal of Biological Chemistry* **2009,** *284* (41), 28352-28366.

71. Kovtun, I. V.; McMurray, C. T., Trinucleotide expansion in haploid germ cells by gap repair. *Nature Genetics* **2001,** *27* (4), 407.

72. Pearson, C. E.; Ewel, A.; Acharya, S.; Fishel, R. A.; Sinden, R. R., Human MSH2 binds to trinucleotide repeat DNA structures associated with neurodegenerative diseases. *Human Molecular Genetics* **1997,** *6* (7), 1117-1123.

73. Manley, K.; Shirley, T. L.; Flaherty, L.; Messer, A., Msh2 deficiency prevents in vivo somatic instability of the CAG repeat in Huntington disease transgenic mice. *Nature Genetics* **1999,** *23* (4), 471-473.

74. Martin, A.; Scharff, M. D., AID and mismatch repair in antibody diversification. *Nature Reviews of Immunology* **2002,** *2* (8), 605.

75. Neuberger, M. S.; Harris, R. S.; Di Noia, J.; Petersen-Mahrt, S. K., Immunity through DNA deamination. *Trends in Biochemical Sciences* **2003,** *28* (6), 305-312.

76. Jiricny, J., The multifaceted mismatch-repair system. *Nature Reviews of Molecular cell biology* **2006,** *7* (5), 335.

77. Setlow, R. B.; Setlow, J. K., Evidence that ultraviolet-induced thymine dimers in DNA cause biological damage. *Proceedings of the National Academy of Sciences* **1962,** *48* (7), 1250-1257.

78. Trosko, J.; Chu, E.; Carrier, W., The induction of thymine dimers in ultraviolet-irradiated mammalian cells. *Radiation Research* **1965,** *24* (4), 667-672.

79. Sinha, R. P.; Häder, D.-P., UV-induced DNA damage and repair: a review. *Photochemical & Photobiological Sciences* **2002,** *1* (4), 225-236.

80. Batty, D. P.; Wood, R. D., Damage recognition in nucleotide excision repair of DNA. *Gene* **2000,** *241* (2), 193-204.

81. Sugitani, N.; Sivley, R. M.; Perry, K. E.; Capra, J. A.; Chazin, W. J., XPA: A key scaffold for human nucleotide excision repair. *DNA Repair* **2016,** *44*, 123-135.

82. Tirode, F.; Busso, D.; Coin, F.; Egly, J.-M., Reconstitution of the transcription factor TFIIH: assignment of functions for the three enzymatic subunits, XPB, XPD, and cdk7. *Molecular Cell* **1999,** *3* (1), 87-95.

83. Mocquet, V.; Laine, J. P.; Riedl, T.; Yajin, Z.; Lee, M. Y.; Egly, J. M., Sequential recruitment of the repair factors during NER: the role of XPG in initiating the resynthesis step. *The EMBO Journal* **2008,** *27* (1), 155-167.

84. Graf, N.; Ang, W. H.; Zhu, G.; Myint, M.; Lippard, S. J., Role of endonucleases XPF and XPG in nucleotide excision repair of platinated DNA and cisplatin/oxaliplatin cytotoxicity. *Chembiochem* **2011,** *12* (7), 1115-1123.

85. Giannattasio, M.; Follonier, C.; Tourrière, H.; Puddu, F.; Lazzaro, F.; Pasero, P.; Lopes, M.; Plevani, P.; Muzi-Falconi, M., Exo1 competes with repair synthesis, converts NER intermediates to long ssDNA gaps, and promotes checkpoint activation. *Molecular Cell* **2010,** *40* (1), 50-62.

86. Lindsey-Boltz, L. A., Bringing It All Together: Coupling Excision Repair to the DNA Damage Checkpoint. *Photochemistry and Photobiology* **2017,** *93* (1), 238-244.

87. Lazzaro, F.; Giannattasio, M.; Puddu, F.; Granata, M.; Pellicioli, A.; Plevani, P.; Muzi-Falconi, M., Checkpoint mechanisms at the intersection between DNA damage and repair. *DNA Repair* **2009,** *8* (9), 1055-1067.

88. Tomlinson, C. G.; Atack, J. M.; Chapados, B.; Tainer, J. A.; Grasby, J. A., Substrate recognition and catalysis by flap endonucleases and related enzymes. *Biochemistry Society Transactions* **2010,** *38* (1), 433-437.

89. 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* **2011,** *145* (2), 212-223.

90. Tran, P. T.; Erdeniz, N.; Symington, L. S.; Liskay, R. M., EXO1-A multi-tasking eukaryotic nuclease. *DNA Repair* **2004,** *3* (12), 1549-1559.

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

92. O'Donovan, A.; Davies, A. A.; Moggs, J. G.; West, S. C.; Wood, R. D., XPG endonuclease makes the 3′ incision in human DNA nucleotide excision repair. *Nature* **1994,** *371* (6496), 432-435.

93. Sun, M.; Schwalb, B.; Pirkl, N.; Maier, K. C.; Schenk, A.; Failmezger, H.; Tresch, A.; Cramer, P., Global analysis of eukaryotic mRNA degradation reveals Xrn1-dependent buffering of transcript levels. *Molecular Cell* **2013,** *52* (1), 52-62.

94. West, S.; Gromak, N.; Proudfoot, N. J., Human 5'to 3'exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature* **2004,** *432* (7016), 522.

95. Pei, J.; Kim, B.-H.; Grishin, N. V., PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Research* **2008,** *36* (7), 2295-2300.

96. Larkin, M. A.; Blackshields, G.; Brown, N.; Chenna, R.; McGettigan, P. A.; McWilliam, H.; Valentin, F.; Wallace, I. M.; Wilm, A.; Lopez, R., Clustal W and Clustal X version 2.0. *Bioinformatics* **2007,** *23* (21), 2947-2948.

97. Liu, Y.; Kao, H.-I.; Bambara, R. A., Flap endonuclease 1: a central component of DNA metabolism. *Annual Review of Biochemistry* **2004,** *73* (1), 589-615.

98. Shaw, S. J.; Finger, L. D.; Grasby, J. A., Human Exonuclease 1 Threads 5′-Flap Substrates through Its Helical Arch. *Biochemistry* **2017,** *56* (29), 3704-3707.

99. 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., Phosphate steering by Flap Endonuclease 1 promotes 5′-flap specificity and incision to prevent genome instability. *Nature Communications* **2017,** *8*.

100. Shi, Y.; Hellinga, H. W.; Beese, L. S., Interplay of catalysis, fidelity, threading, and processivity in the exo-and endonucleolytic reactions of human exonuclease I. *Proceedings of the National Academy of Sciences* **2017**, 201704845.

101. 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 Research* **2012,** *40* (10), 4507-4519.

102. Pelletier, H.; Sawaya, M. R.; Wolfle, W.; Wilson, S. H.; Kraut, J., Crystal structures of human DNA polymerase β complexed with DNA: implications for catalytic mechanism, processivity, and fidelity. *Biochemistry* **1996,** *35* (39), 12742-12761.

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

104. Moldovan, G.-L.; Pfander, B.; Jentsch, S., PCNA, the maestro of the replication fork. *Cell* **2007,** *129* (4), 665-679.

105. Keijzers, G.; Liu, D.; Rasmussen, L. J., Exonuclease 1 and its versatile roles in DNA repair. *Critical Reviews in Biochemistry and Molecular Biology* **2016,** *51* (6), 440-451.

106. Araújo, S. J.; Nigg, E. A.; Wood, R. D., Strong functional interactions of TFIIH with XPC and XPG in human DNA nucleotide excision repair, without a preassembled repairosome. *Molecular and Cellular Biology* **2001,** *21* (7), 2281-2291.

107. Sambrook, J.; Russell, D. W., *The Inoue Method for Preparation and Transformation of Competent E. Coli: "Ultra-Competent" Cells*. Cold Spring Harbor: 2006.

108. Studier, W. F., Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification* **2005,** *41* (1), 207-234.

109. Clegg, R. M., [18] Fluorescence resonance energy transfer and nucleic acids. *Methods in Enzymology* **1992,** *211*, 353-388.

110. Exell, J. C., Spectroscopic evidence for catalytically-required FEN1-mediated DNA conformational change; a novel strategy for FEN1 inhibition. *University of Sheffield Press* **2015**, 277.

111. Tumey, L. N.; Bom, D.; Huck, B.; Gleason, E.; Wang, J.; Silver, D.; Brunden, K.; Boozer, S.; Rundlett, S.; Sherf, B., The identification and optimization of a N-hydroxy urea series of flap endonuclease 1 inhibitors. *Bioorganic & Medicinal Chemistry Letters* **2005,** *15* (2), 277-281.

112. Exell, J. C.; Thompson, M. J.; Finger, L. D.; Shaw, S. J.; Debreczeni, J.; Ward, T. A.; McWhirter, C.; Siöberg, C. L.; Molina, D. M.; Abbott, W. M., Cellularly active N-hydroxyurea FEN1 inhibitors block substrate entry to the active site. *Nature Chemical Biology* **2016,** *12* (10), 815-821.

113. Sengerová, B.; Tomlinson, C.; Atack, J. M.; Williams, R.; Sayers, J. R.; Williams, N. H.; Grasby, J. A., Brønsted analysis and rate-limiting steps for the T5 flap endonuclease catalyzed hydrolysis of exonucleolytic substrates. *Biochemistry* **2010,** *49* (37), 8085-8093.

114. Bennet, I. A., Evidence of dynamics and disorider using NMR-spectroscopic techniques applied to human Flap Endonuclease-1. *University of Sheffield Press* **2017**, 118.

115. Bennet, I. A.; Finger, L. D.; Baxter, N. J.; Hounslow, A. M.; Exell, J. C.; Waltho, J. P.; Grasby, J. A., Regional intrinsic disorder couples substrate specificity and scissile phosphate diester selectivity in Human Flap Endonuclease-1. *Nucleic Acids Research In review*.

116. Myler, L. R.; Gallardo, I. F.; Zhou, Y.; Gong, F.; Yang, S.-H.; Wold, M. S.; Miller, K. M.; Paull, T. T.; Finkelstein, I. J., Single-molecule imaging reveals the mechanism of Exo1 regulation by single-stranded DNA binding proteins. *Proceedings of the National Academy of Sciences* **2016,** *113* (9), E1170-E1179.

117. Fregel, R.; González, A.; Cabrera, V. M., Improved ethanol precipitation of DNA. *Electrophoresis* **2010,** *31* (8), 1350-1352.

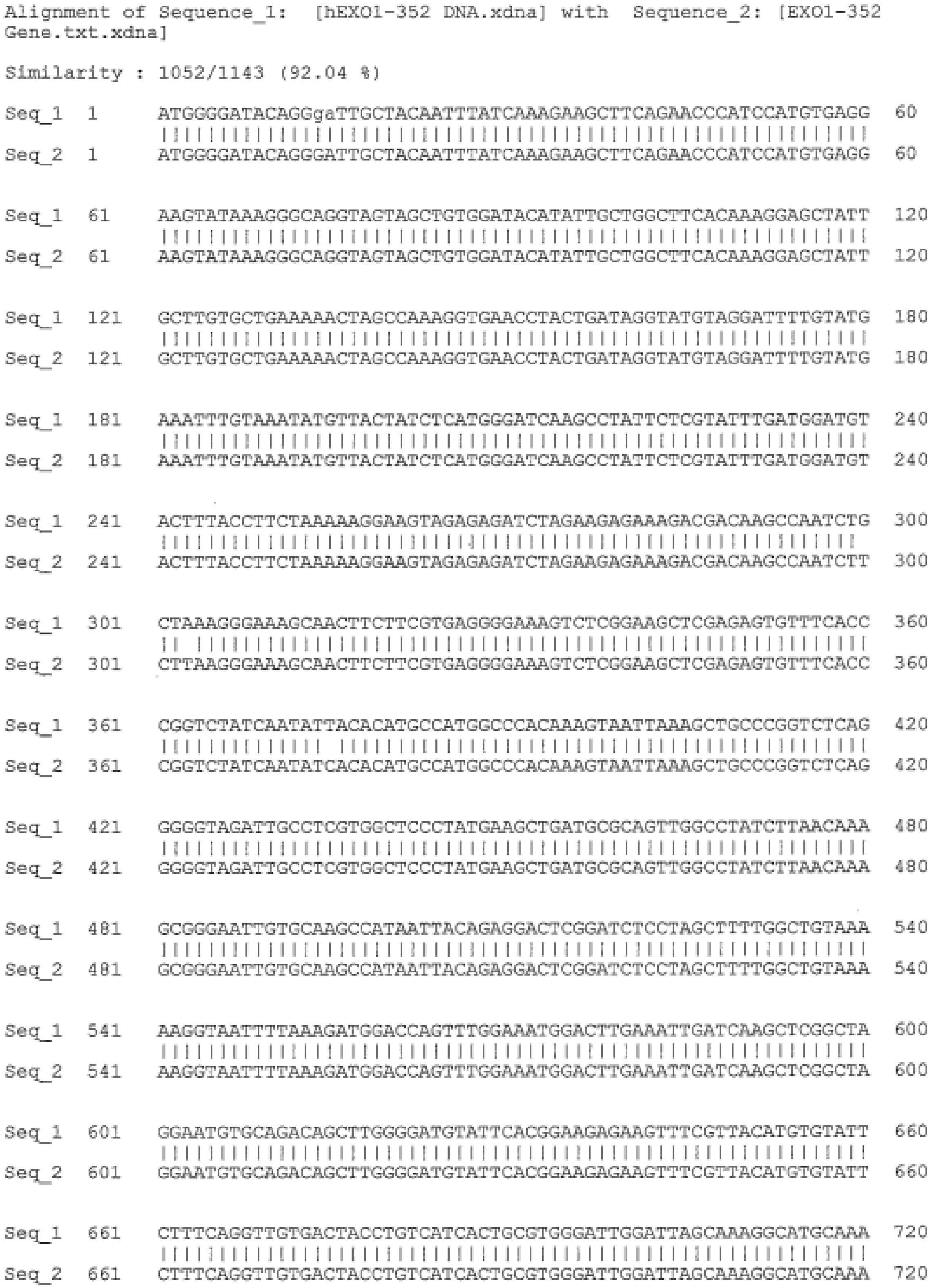
118. Algasaier, S. I.; Exell, J. C.; Bennet, I. A.; Thompson, M. J.; Gotham, V. J.; Shaw, S. J.; Craggs, T. D.; Finger, L. D.; Grasby, J. A., DNA and protein requirements for substrate conformational changes necessary for human flap Endonuclease-1-catalyzed reaction. *Journal of Biological Chemistry* **2016,** *291* (15), 8258-8268.

119. Zheng, G.; Lu, X.-J.; Olson, W. K., Web 3DNA—a web server for the analysis, reconstruction, and visualization of three-dimensional nucleic-acid structures. *Nucleic Acids Research* **2009,** *37* (suppl_2), W240-W246.

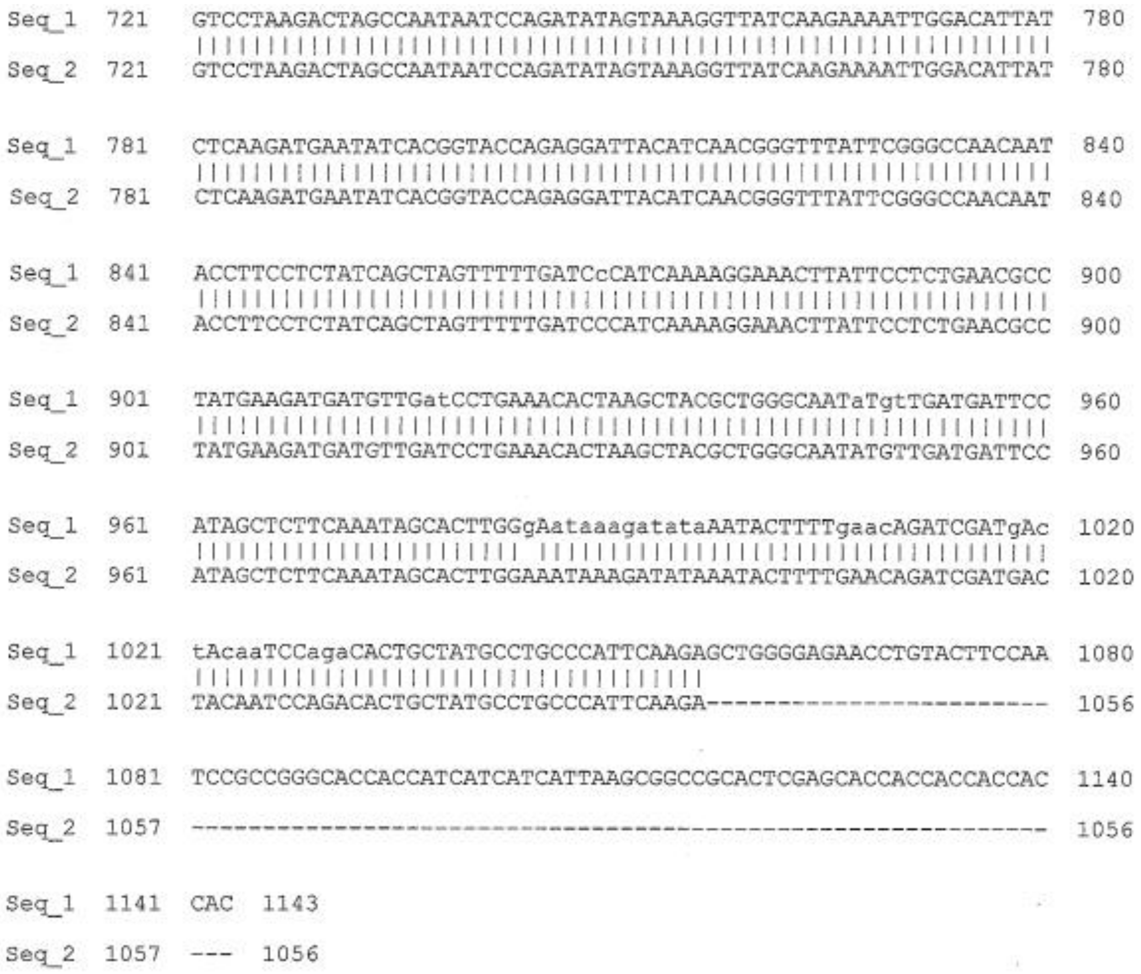
120. Rashid, F.; Harris, P. D.; Zaher, M. S.; Sobhy, M. A.; Joudeh, L. I.; Yan, C.; Piwonski, H.; Tsutakawa, S. E.; Ivanov, I.; Tainer, J. A., Single-molecule FRET unveils induced-fit mechanism for substrate selectivity in flap endonuclease 1. *eLife* **2017,** *6*, e21884.

121. 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* (46), 7243-7246.
122. Zheng, L.; Jia, J.; Finger, L. D.; Guo, Z.; Zer, C.; Shen, B., Functional regulation of FEN1 nuclease and its link to cancer. *Nucleic Acids Research* **2010,** *39* (3), 781-794.
123. Tomassini, J.; Davies, M.; Hastings, J.; Lingham, R.; Mojena, M.; Raghoobar, S.; Singh, S.; Tkacz, J.; Goetz, M., A novel antiviral agent which inhibits the endonuclease of influenza viruses. *Antimicrobial Agents and Chemotherapy* **1996,** *40* (5), 1189-1193.
124. Niesen, F. H.; Berglund, H.; Vedadi, M., The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nature Protocols* **2007,** *2* (9), 2212-2221.
125. Chapados, B. R.; Hosfield, D. J.; Han, S.; Qiu, J.; Yelent, B.; Shen, B.; Tainer, J. A., Structural basis for FEN-1 substrate specificity and PCNA-mediated activation in DNA replication and repair. *Cell* **2004,** *116* (1), 39-50.
126. Tom, S.; Henricksen, L. A.; Bambara, R. A., Mechanism whereby proliferating cell nuclear antigen stimulates flap endonuclease 1. *Journal of Biological Chemistry* **2000,** *275* (14), 10498-10505.
127. Liberti, S. E.; Andersen, S. D.; Wang, J.; May, A.; Miron, S.; Perderiset, M.; Keijzers, G.; Nielsen, F. C.; Charbonnier, J.-B.; Bohr, V. A., Bi-directional routing of DNA mismatch repair protein human exonuclease 1 to replication foci and DNA double strand breaks. *DNA Repair* **2011,** *10* (1), 73-86.
128. Qiu, J.; Qian, Y.; Chen, V.; Guan, M.-X.; Shen, B., Human exonuclease 1 functionally complements its yeast homologues in DNA recombination, RNA primer removal, and mutation avoidance. *Journal of Biological Chemistry* **1999,** *274* (25), 17893-17900.
129. Keijzers, G.; Bohr, V. A.; Rasmussen, L. J., Human exonuclease 1 (EXO1) activity characterization and its function on FLAP structures. *Bioscience Reports* **2015,** *35* (3), e00206.

Appendices



```
Alignment of Sequence_1:  [hEXO1-352 DNA.xdna] with  Sequence_2: [EXO1-352
Gene.txt.xdna]

Similarity : 1052/1143 (92.04 %)

Seq_1  1       ATGGGGATACAGGgaTTGCTACAATTTATCAAAGAAGCTTCAGAACCCATCCATGTGAGG  60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  1       ATGGGGATACAGGGATTGCTACAATTTATCAAAGAAGCTTCAGAACCCATCCATGTGAGG  60


Seq_1  61      AAGTATAAAGGGCAGGTAGTAGCTGTGGATACATATTGCTGGCTTCACAAAGGAGCTATT  120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  61      AAGTATAAAGGGCAGGTAGTAGCTGTGGATACATATTGCTGGCTTCACAAAGGAGCTATT  120


Seq_1  121     GCTTGTGCTGAAAAACTAGCCAAAGGTGAACCTACTGATAGGTATGTAGGATTTTGTATG  180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  121     GCTTGTGCTGAAAAACTAGCCAAAGGTGAACCTACTGATAGGTATGTAGGATTTTGTATG  180


Seq_1  181     AAATTTGTAAATATGTTACTATCTCATGGGATCAAGCCTATTCTCGTATTTGATGGATGT  240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  181     AAATTTGTAAATATGTTACTATCTCATGGGATCAAGCCTATTCTCGTATTTGATGGATGT  240


Seq_1  241     ACTTTACCTTCTAAAAAGGAAGTAGAGAGATCTAGAAGAGAAAGACGACAAGCCAATCTG  300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Seq_2  241     ACTTTACCTTCTAAAAAGGAAGTAGAGAGATCTAGAAGAGAAAGACGACAAGCCAATCTT  300


Seq_1  301     CTAAAGGGAAAGCAACTTCTTCGTGAGGGGAAAGTCTCGGAAGCTCGAGAGTGTTTCACC  360
               || |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  301     CTTAAGGGAAAGCAACTTCTTCGTGAGGGGAAAGTCTCGGAAGCTCGAGAGTGTTTCACC  360


Seq_1  361     CGGTCTATCAATATTACACATGCCATGGCCCACAAAGTAATTAAAGCTGCCCGGTCTCAG  420
               |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Seq_2  361     CGGTCTATCAATATCACACATGCCATGGCCCACAAAGTAATTAAAGCTGCCCGGTCTCAG  420


Seq_1  421     GGGGTAGATTGCCTCGTGGCTCCCTATGAAGCTGATGCGCAGTTGGCCTATCTTAACAAA  480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  421     GGGGTAGATTGCCTCGTGGCTCCCTATGAAGCTGATGCGCAGTTGGCCTATCTTAACAAA  480


Seq_1  481     GCGGGAATTGTGCAAGCCATAATTACAGAGGACTCGGATCTCCTAGCTTTTGGCTGTAAA  540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  481     GCGGGAATTGTGCAAGCCATAATTACAGAGGACTCGGATCTCCTAGCTTTTGGCTGTAAA  540


Seq_1  541     AAGGTAATTTTAAAGATGGACCAGTTTGGAAATGGACTTGAAATTGATCAAGCTCGGCTA  600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  541     AAGGTAATTTTAAAGATGGACCAGTTTGGAAATGGACTTGAAATTGATCAAGCTCGGCTA  600


Seq_1  601     GGAATGTGCAGACAGCTTGGGGATGTATTCACGGAAGAGAAGTTTCGTTACATGTGTATT  660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  601     GGAATGTGCAGACAGCTTGGGGATGTATTCACGGAAGAGAAGTTTCGTTACATGTGTATT  660


Seq_1  661     CTTTCAGGTTGTGACTACCTGTCATCACTGCGTGGGATTGGATTAGCAAAGGCATGCAAA  720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  661     CTTTCAGGTTGTGACTACCTGTCATCACTGCGTGGGATTGGATTAGCAAAGGCATGCAAA  720
```



```
Seq_1  721   GTCCTAAGACTAGCCAATAATCCAGATATAGTAAAGGTTATCAAGAAAATTGGACATTAT  780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  721   GTCCTAAGACTAGCCAATAATCCAGATATAGTAAAGGTTATCAAGAAAATTGGACATTAT  780

Seq_1  781   CTCAAGATGAATATCACGGTACCAGAGGATTACATCAACGGGTTTATTCGGGCCAACAAT  840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  781   CTCAAGATGAATATCACGGTACCAGAGGATTACATCAACGGGTTTATTCGGGCCAACAAT  840

Seq_1  841   ACCTTCCTCTATCAGCTAGTTTTTGATCcCATCAAAAGGAAACTTATTCCTCTGAACGCC  900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  841   ACCTTCCTCTATCAGCTAGTTTTTGATCCCATCAAAAGGAAACTTATTCCTCTGAACGCC  900

Seq_1  901   TATGAAGATGATGTTGatCCTGAAACACTAAGCTACGCTGGGCAATaTgtTGATGATTCC  960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  901   TATGAAGATGATGTTGATCCTGAAACACTAAGCTACGCTGGGCAATATGTTGATGATTCC  960

Seq_1  961   ATAGCTCTTCAAATAGCACTTGGgAataaagatataAATACTTTTgaacAGATCGATgAc  1020
             ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Seq_2  961   ATAGCTCTTCAAATAGCACTTGGAAATAAAGATATAAATACTTTTGAACAGATCGATGAC  1020

Seq_1  1021  tAcaaTCCagaCACTGCTATGCCTGCCCATTCAAGAGCTGGGGAGAACCTGTACTTCCAA  1080
             ||||||||||||||||||||||||||||||||||||
Seq_2  1021  TACAATCCAGACACTGCTATGCCTGCCCATTCAAGA------------------------  1056

Seq_1  1081  TCCGCCGGGCACCACCATCATCATCATTAAGCGGCCGCACTCGAGCACCACCACCACCAC  1140

Seq_2  1057  ------------------------------------------------------------  1056

Seq_1  1141  CAC  1143

Seq_2  1057  ---  1056
```

Figure A1: The optimised hEXO1-352 sequence used. *The codon optimised sequence of hEXO1-352 used herein (Seq_1 – obtained from gene art) aligned against the hEXO1 gene sequence (Seq_2). The Seq_2 sequence is truncated at 352 codons.*

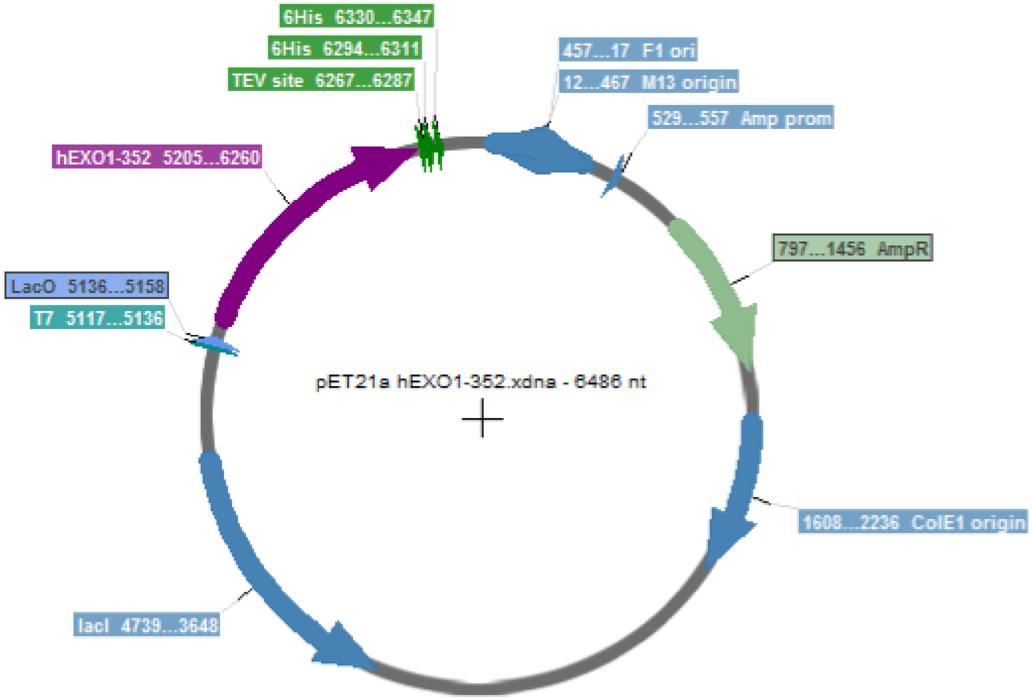
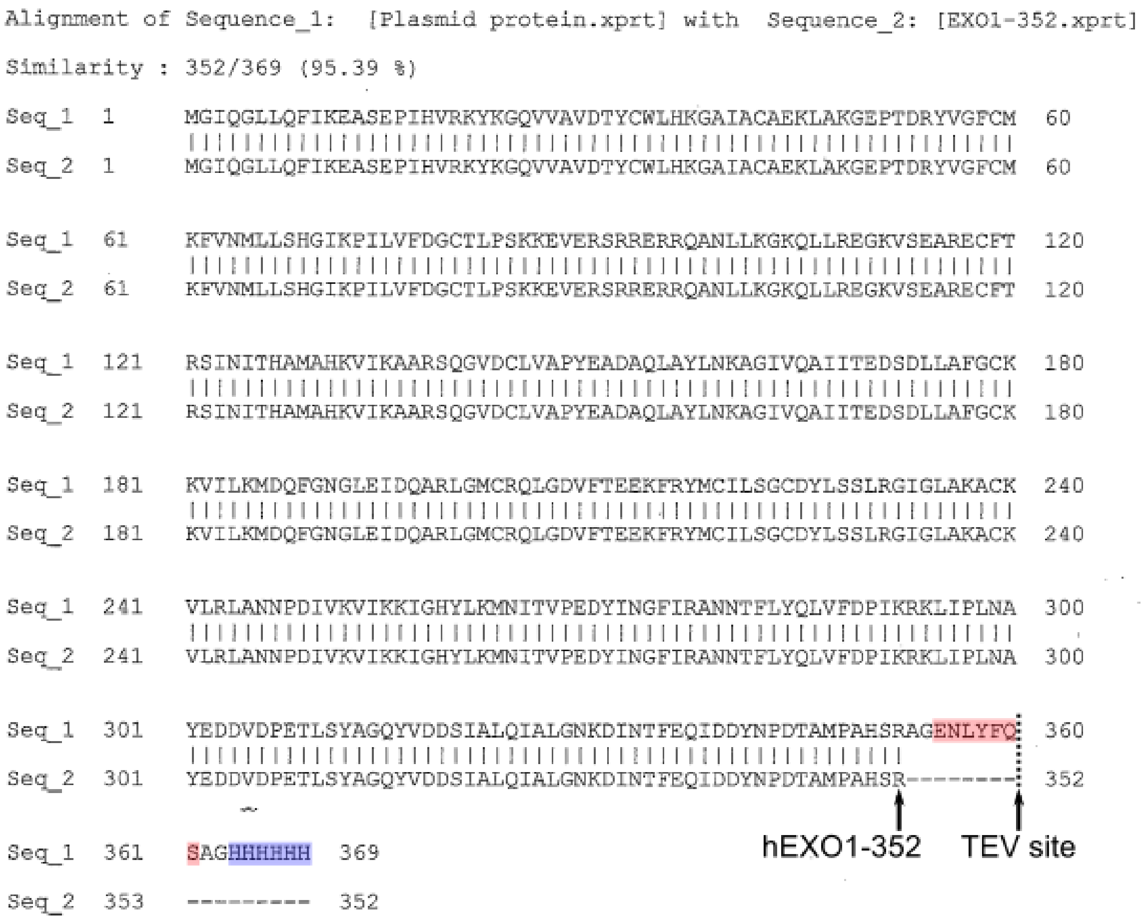


Figure A2: Human EXO1-352 plasmid created. *The pET21a plasmid with the hEXO1 gene, truncated to produce only 352 amino acids (purple). The gene insertion contains a turboTEV cleavable 6 histidine tag (green).*



```
Alignment of Sequence_1:  [Plasmid protein.xprt] with  Sequence_2: [EXO1-352.xprt]

Similarity : 352/369 (95.39 %)

Seq_1  1     MGIQGLLQFIKEASEPIHVRKYKGQVVAVDTYCWLHKGAIACAEKLAKGEPTDRYVGFCM  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  1     MGIQGLLQFIKEASEPIHVRKYKGQVVAVDTYCWLHKGAIACAEKLAKGEPTDRYVGFCM  60

Seq_1  61    KFVNMLLSHGIKPILVFDGCTLPSKKEVERSRRERRQANLLKGKQLLREGKVSEARECFT  120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  61    KFVNMLLSHGIKPILVFDGCTLPSKKEVERSRRERRQANLLKGKQLLREGKVSEARECFT  120

Seq_1  121   RSINITHAMAHKVIKAARSQGVDCLVAPYEADAQLAYLNKAGIVQAIITEDSDLLAFGCK  180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  121   RSINITHAMAHKVIKAARSQGVDCLVAPYEADAQLAYLNKAGIVQAIITEDSDLLAFGCK  180

Seq_1  181   KVILKMDQFGNGLEIDQARLGMCRQLGDVFTEEKFRYMCILSGCDYLSSLRGIGLAKACK  240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  181   KVILKMDQFGNGLEIDQARLGMCRQLGDVFTEEKFRYMCILSGCDYLSSLRGIGLAKACK  240

Seq_1  241   VLRLANNPDIVKVIKKIGHYLKMNITVPEDYINGFIRANNTFLYQLVFDPIKRKLIPLNA  300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  241   VLRLANNPDIVKVIKKIGHYLKMNITVPEDYINGFIRANNTFLYQLVFDPIKRKLIPLNA  300

Seq_1  301   YEDDVDPETLSYAGQYVDDSIALQIALGNKDINTFEQIDDYNPDTAMPAHSRAGENLYFQ  360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  301   YEDDVDPETLSYAGQYVDDSIALQIALGNKDINTFEQIDDYNPDTAMPAHSR--------  352

Seq_1  361   SAGHHHHHH  369

Seq_2  353   ---------  352
```

Figure A3: Human EXO1-352 protein sequence. *The protein sequence expected to express from the plasmid, with 6-His tag in blue, which is cleavable by TurboTEV (Biochemistry, Ltd) using the site highlighted in red. The TEV cleavage site is highlighted, the protein purified and used in the experiments is between G2 and Q360.*

Publications

⌘ *Author's Choice*

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 291, NO. 15, pp. 8258–8268, April 8, 2016
© 2016 by The American Society for Biochemistry and Molecular Biology, Inc. Published in the U.S.A.

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

Received for publication, October 20, 2015, and in revised form, February 4, 2016 Published, JBC Papers in Press, February 16, 2016, DOI 10.1074/jbc.M115.698993

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

---

[*] This work was supported by Biotechnology and Biological Sciences Research Council Grants BB/K009079/1 and BB/M00404X/1. This work was also supported by a studentship from the Ministry of Higher Education Libya (to S. I. A.), studentships from the EPSRC and Astra Zeneca (to J. C. E.), and Biotechnology and Biological Sciences Research Council White Rose Studentship DTP BB/J014443/1 (to I. A. B.). The authors declare that they have no conflicts of interest with the contents of this article.

[1] The first two authors should be regarded as joint first authors.

[2] Present address: Dept. of Microbiology and Molecular Genetics, University of California, Davis, Briggs Hall, One Shields Ave., Davis, CA 95616-8665.

[3] To whom correspondence should be addressed: Centre for Chemical Biology, Dept. of Chemistry, Krebs Institute, University of Sheffield, Sheffield S3 7HF, UK. Tel.: 44-114-222-9478; E-mail: j.a.grasby@sheffield.ac.uk.

[4] The abbreviations used are: FEN1, flap endonuclease-1; nt, nucleotide; 2AP, 2-aminopurine; DF, double flap; SF, single (3′) flap; NL, nonlabeled; DOL, donor-only labeled; DAL, doubly labeled; ECCD, exciton coupled CD; h, human; TAMRA, tetramethylrhodamine; MMDF, mismatch containing (+1) double flap; SADF, 5′-streptavidin double flap.

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-aminoethoxyethanol 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 $(\text{ratio})_A$ 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.

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

where

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

and

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

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

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.

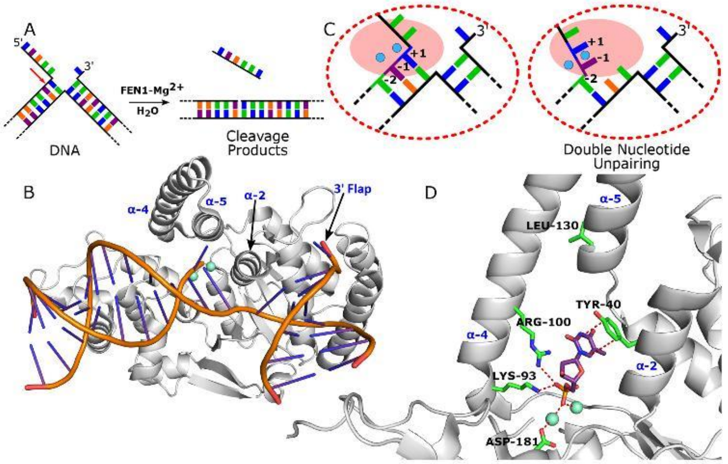


FIGURE 1. **FEN1 DNA bending and double nucleotide unpairing.** *A*, schematic of the FEN1 catalyzed hydrolysis of a double flap DNA yielding single-stranded DNA and double-stranded nicked DNA products. An *arrow* indicates the site of reaction. Each nucleobase is represented by a different color. *B*, 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 α2-α4) and cap (top of α4 and α5) 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 $(\text{ratio})_A$ 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 $\pm Ca^{2+}$ (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

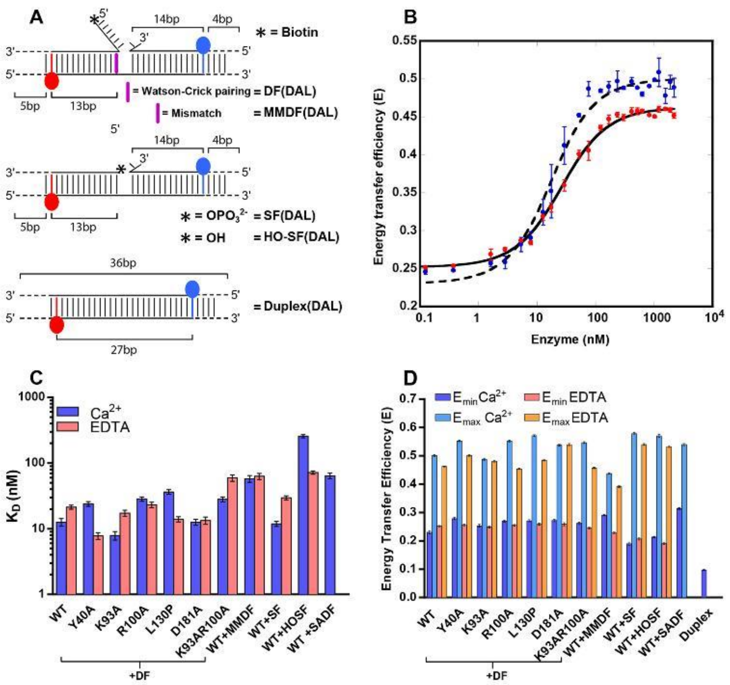


FIGURE 2. **FRET data showing DNA bending on complexation with hFEN1 and mutants.** *A*, schematic of double flap (DF, endonucleolytic) and single flap (SF, exonucleolytic) DNA constructs (Table 2) used in FRET studies, donor = fluorescein (*blue*) and acceptor = TAMRA (*red*). Nonlabeled (NL), donor 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 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′-**pAA**G AGG CAG AGT G-3′
ssHO-SF$_{+1-1}$	5′-**AA**G AGG CAG AGT G-3′
ssSF$_{-1-2}$	5′-pG**AA** AGG CAG AGT G-3′
ssHO-SF$_{-1-2}$	5′-G**AA** AGG CAG AGT G-3′
ssDF$_{+1-1}$	5′-TTT TT**A** **A**GA GGC AGA GTG-3′
ssDF$_{-1-2}$	5′-TTT TTG **AA**A GGC AGA GTC-3′
ssMM$_{+1}$DF$_{-1-2}$	5′-TTT TTC **AA**A 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′-pTTG AGG CAG AGT AGG ACC-3′
EO-Fret	5′-TTG AGG CAG AGT AGG ACC-3′
MM$_{+1}$-Fret	5′-pTTT 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*).

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 $K_{\text{bend}} \pm Ca^{2+}$ 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 $\alpha 4$ 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 ($\alpha 5$) 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 $\alpha 2$ 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 ($k_{\text{STmax}} = 740\ \text{min}^{-1}$) (data not shown). For Y40A, $k_{\text{STmax}} = 7.91 \pm 0.01\ \text{min}^{-1}$, and for D181A, $k_{\text{STmax}} = 0.075 \pm 0.003\ \text{min}^{-1}$, 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 $K_{\text{bend}} = 13 \pm 1.7$ nM, MMDF $K_{\text{bend}} = 58 \pm 6.8$ 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 ($K_{\text{bend}} = 12 \pm 1.1$ 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 $\alpha 4$ and $\alpha 5$) and gateway (base of $\alpha 4$ and $\alpha 2$) (Fig. 1, *B* and *D*) (1, 8, 12–14). Evidence for this so-called threading hypothesis came 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 ($\lambda_{\max} = 326$ 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 signal 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-$Ca^{2+}\cdot SF_{-1-2}$ 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

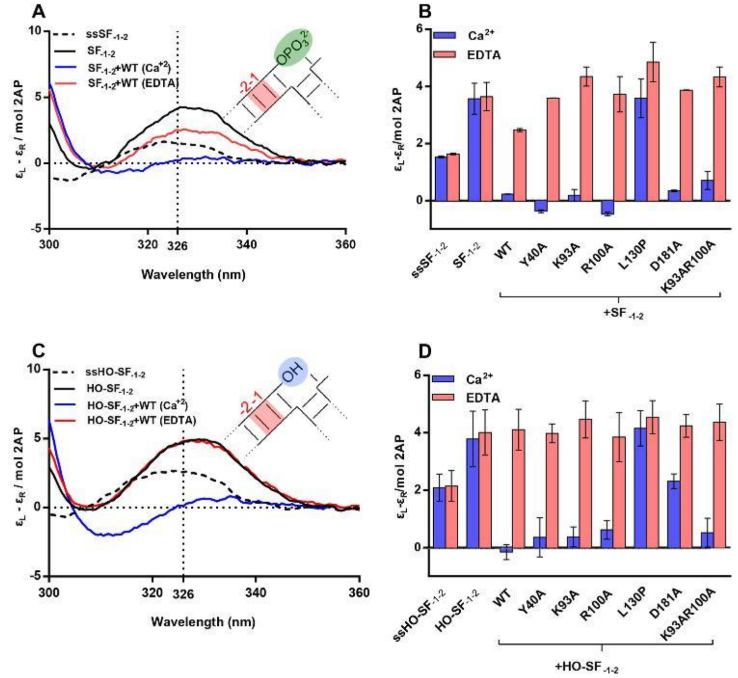


FIGURE 3. **hFEN1 and mutant mediated conformational change of 2AP-containing single flap SF$_{-1-2}$ monitored by ECCD.** All measurements were carried out at 20 °C and pH 7.5. *A*, divalent metal ion-dependent reduction in 2AP exciton coupling signal occurred when substrate SF$_{-1-2}$ was bound to 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$_{-1-2}$ as SF$_{-1-2}$ 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

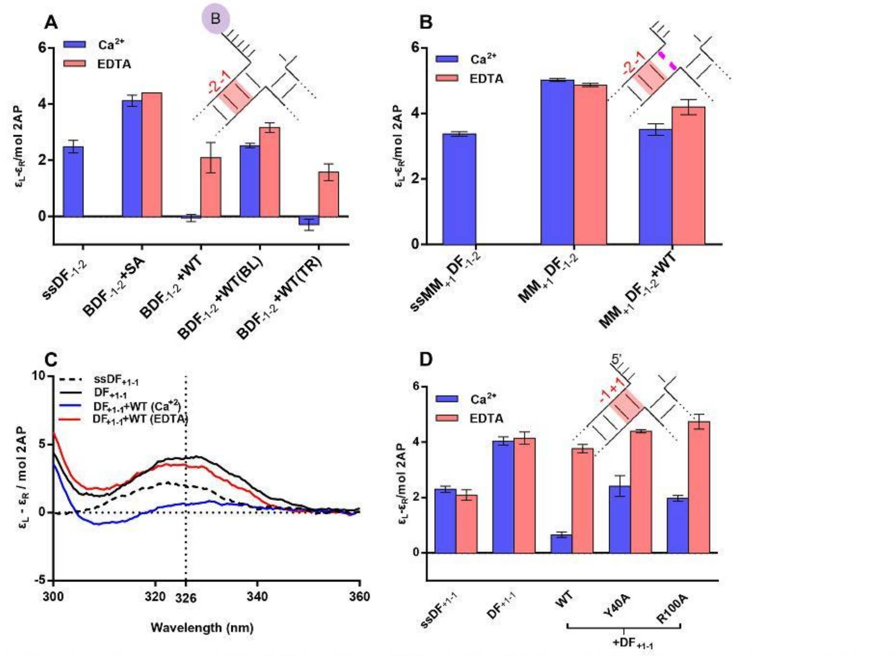


FIGURE 4. **ECCD monitored conformational change of +1−1 2AP and 5′-modified −1−2 double flap substrates.** All measurements were carried out at 20 °C and pH 7.5. *ss*, single strand. Standard errors from repeat experiments are shown. *A*, comparison of molar ellipticity per 2AP residue at 326 nm of 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. 1*C*). 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. 5*B*). 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. 5*B*). However, Y40A, L130P, and D181A did not significantly alter the signal with SF_{+1-1} at 326 nm $\pm Ca^{2+}$. 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. 4*D*).

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. 5*C*). 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,

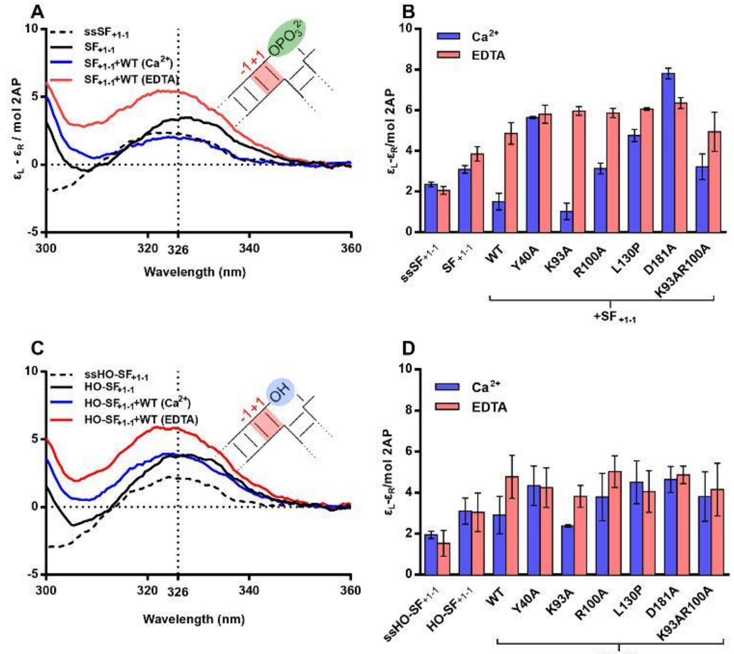


FIGURE 5. **ECCD monitored conformational change of single flap +1−1 2AP containing substrates upon binding hFEN1 and mutants.** All measurements were carried out at 20 °C and pH 7.5. *ss*, single strand. *A*, divalent metal ion-dependent reduction in 2AP exciton coupling signal occurred when single 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.

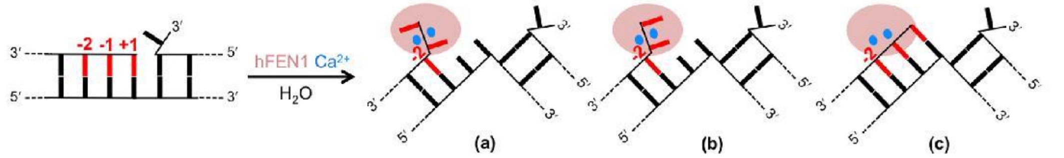


FIGURE 6. **Schematic model summarizing the responses of hFEN1-substrate complexes to addition of divalent metal ions based on ECCD results.** *Part (a)*, in the presence of divalent ions, unmodified substrates interacting with WT and K93A hFEN1s adopt an orientation of the −1 and −2 nt that is unstacked 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 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. 6*C*).

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. 4*B* and 6*C*). 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. 4*A* and 6*C*). 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. 3*C*), only a small change is observed in the +1 and −1 ECCD signal (Fig. 5*C*). 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. 2*C*), 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. 2*C*). However, several of these residues were implicated in active site substrate positioning by studies of the +1 and −1 ECCD signal (Figs. 5*B* 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. 4*D*). 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}

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

nature chemical biology

ARTICLE

PUBLISHED ONLINE: 15 AUGUST 2016 | DOI: 10.1038/NCHEMBIO.2148

Cellularly active *N*-hydroxyurea FEN1 inhibitors block substrate entry to the active site

Jack C Exell[1,8], Mark J Thompson[1], L David Finger[1], Steven J Shaw[1], Judit Debreczeni[2], Thomas A Ward[3], Claire McWhirter[2], Catrine L B Siöberg[4], Daniel Martinez Molina[4], W Mark Abbott[2], Clifford D Jones[5], J Willem M Nissink[6*], Stephen T Durant[7*] & Jane A Grasby[1,7*]

The structure-specific nuclease human flap endonuclease-1 (hFEN1) plays a key role in DNA replication and repair and may be of interest as an oncology target. We present the crystal structure of inhibitor-bound hFEN1, which shows a cyclic *N*-hydroxyurea bound in the active site coordinated to two magnesium ions. Three such compounds had similar IC$_{50}$ values but differed subtly in mode of action. One had comparable affinity for protein and protein–substrate complex and prevented reaction by binding to active site catalytic metal ions, blocking the necessary unpairing of substrate DNA. Other compounds were more competitive with substrate. Cellular thermal shift data showed that both inhibitor types engaged with hFEN1 in cells, and activation of the DNA damage response was evident upon treatment with inhibitors. However, cellular EC$_{50}$ values were significantly higher than *in vitro* inhibition constants, and the implications of this for exploitation of hFEN1 as a drug target are discussed.

Flap endonuclease-1 (FEN1) is the prototypical member of the 5′-nuclease superfamily[1,2], whose activities span a range of cellular pathways involved in DNA replication and genome maintenance[3,4]. FEN1 is a structure-selective metallonuclease essential for Okazaki fragment maturation through efficient removal of 5′ flaps resulting from strand displacement during lagging-strand synthesis[5,6]. This reaction produces nicked DNA suitable for ligation, thereby ensuring maintenance of genomic fidelity. FEN1 is also involved in long-patch base excision repair[7–9] (LP-BER), among other pathways.

Given its critical replicative function, it is not surprising that FEN1 is overexpressed in multiple cancer types[10–13]. It has been suggested both as a biomarker relating to prognosis and disease progression and as a potential therapeutic target. Target validation studies have focused either on chemosensitization[14,15] or on synthetic-lethal interactions[16–19] with established oncogenes.

Synthetic lethality arises when loss of function of either gene of an interacting pair is not cytotoxic, but mutation or inhibition of both does cause cell death; hence, targeting interacting partners of mutated genes in cancer offers potential for selective killing of cancer cells. Therapeutic interest in FEN1 arises from its known synthetic-lethal interactions with several genes frequently mutated in cancers[16,17,20]. FEN1 inhibition selectively impairs proliferation of colon cancer cells deficient in *Fbxw7* (*Cdc4*) and *Mre11a* (refs. 16,18), both frequently mutated in colorectal cancers.

FEN1 has also emerged as a potential chemosensitizing target. Via its role in LP-BER[17], FEN1 is critical for repair of methyl methanesulfonate (MMS)-induced alkylation damage[21], and its knockdown or inhibition increases sensitivity to temozolomide (TMZ) in glioblastoma[13] and colorectal cancer[14,16,18] cell lines.

This considerable interest in human FEN1 as a drug target has prompted development of high-throughput screening procedures[22,23] and the discovery of a series of *N*-hydroxyurea-based hFEN1 inhibitors[24]. We investigated the specificity and mode of action of these compounds and found they prevented access of the scissile phosphate diester of substrate DNA to catalytic metal ions. We also demonstrated cellular activity and target engagement in live cells, leading to activation of the DNA damage response and apoptosis.

RESULTS

N-Hydroxyurea hFEN1 inhibitors bind catalytic-site metals

Inhibitor **1** (ref. 24; **Fig. 1a**; see **Supplementary Note** for details of chemical synthesis) was cocrystallized with hFEN1–Mg^{2+} truncated after residue 336 (hFEN1-336Δ), which retains all catalytic features but lacks the flexible 44-residue C terminus[25,26]. The crystal structure of the hFEN1-336Δ–inhibitor complex (**Fig. 1b**) was solved at 2.84-Å resolution (**Supplementary Results**, **Supplementary Table 1** and **Supplementary Fig. 1**; PDB 5FV7). It resembles a kidney bean with the active site and requisite divalent metal ions residing at the indentation. The structure in the presence of the active site-bound inhibitor closely resembles that of hFEN1 in complex with proliferating cell nuclear antigen (PCNA)[27]. As with the PCNA-bound structure, no density was observed for the helical arch (α4 and α5) and α2-α3 loop regions, which are visible when the complex is cocrystallized with substrate or product DNA[2].

The inhibitor is situated in the protein's nuclease active site with the *N*-hydroxyurea moiety directly coordinating two Mg^{2+} ions positioned 4.5 Å apart (**Fig. 1b**), anchored by inner-sphere metal-coordinating contacts from carboxylates of Glu160, Asp179 and Asp181 and outer-sphere or water-mediated contacts from Asp34, Asp86, Glu158 and Asp233 (**Fig. 1c**). The thiophene ring of the inhibitor fills a small hydrophobic pocket formed by Met37, Tyr40 and Val133, and the sulfur of Met37 has a short-distance (4 Å) favorable contact to the electron-deficient pyrimidine-2, 4-dione ring of the ligand. The 2,3-dihydrobenzo[*b*][1,4]dioxine side chain contacts Met37 and Tyr40, though these contacts are

---

[1]Centre for Chemical Biology, Department of Chemistry, Krebs Institute, University of Sheffield, Sheffield, UK. [2]Discovery Sciences, Innovative Medicines and Early Development Biotech Unit, AstraZeneca, Cambridge, UK. [3]Bioscience, Oncology Innovative Medicines and Early Development Biotech Unit, AstraZeneca, Alderley Park, Cheshire, UK. [4]Pelago Bioscience AB, Solna, Sweden. [5]Chemistry, Oncology Innovative Medicines and Early Development Biotech Unit, AstraZeneca, Alderley Park, UK. [6]Chemistry, Oncology Innovative Medicines and Early Development Biotech Unit, AstraZeneca, Cambridge, UK. [7]Bioscience, Oncology Innovative Medicines and Early Development Biotech Unit, Cambridge, UK. [8]Present address: Department of Microbiology and Molecular Genetics, University of California, Davis, Davis, California, USA. *e-mail: j.a.grasby@sheffield.ac.uk, stephen.durant@astrazeneca.com or willem.nissink@astrazeneca.com

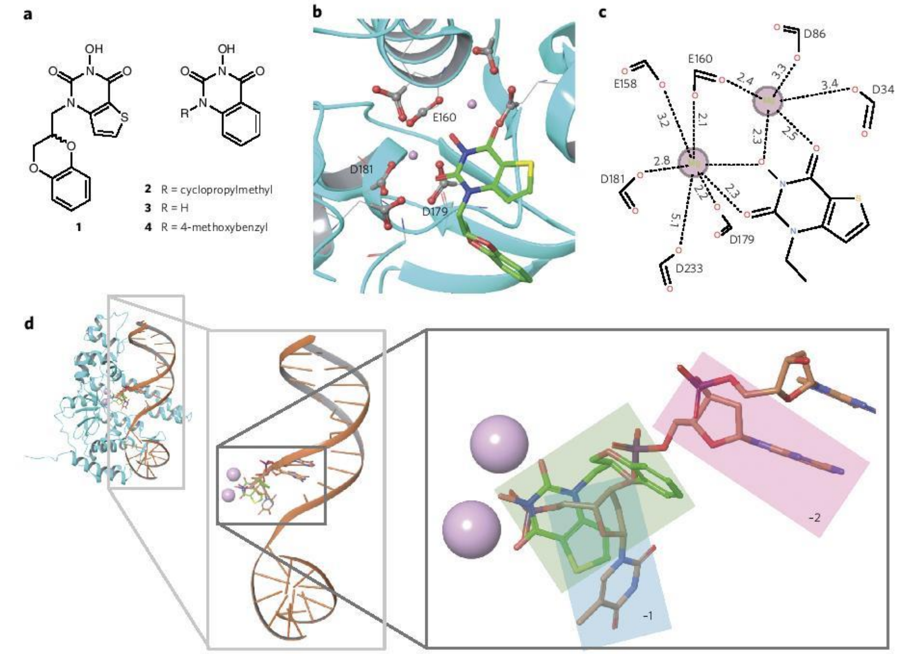


Figure 1 | Compounds used in this study and crystal structure of hFEN1-336Δ in complex with compound 1. (**a**) Schematic illustration of **1**–**4**, which are inhibitors of hFEN1 phosphate diester hydrolysis. (**b**) Structure of hFEN1-336Δ nuclease active site (PDB 5FV7) showing the seven highly conserved acidic residues (gray and red spheres represent carbonyl carbon and oxygen atoms, respectively), the two bound magnesium ions (pink spheres), and **1**. (**c**) Schematic representation of the metal-coordination spheres of the two magnesium ions in the active site, with distances reported in ångströms. (**d**) Structure of hFEN1-336Δ in complex with product DNA (PDB 3Q8K) superimposed with that of hFEN1-336Δ in complex with **1** (protein not shown) to show that the inhibitor and the terminal nucleotide of the product DNA interact with the divalent magnesium ions and share the same pocket in the protein. Metals are shown as pink spheres, the terminal 5′ nucleotide (−1) is highlighted in the cyan box, the penultimate nucleotide of the product DNA (−2) is highlighted in the pink box, and **1** is highlighted in the green box.

less directional and mostly hydrophobic in nature. It is evident that different binding poses in the active site are possible for the series of *N*-hydroxyurea-based inhibitors, which goes some way toward rationalizing the reported structure–activity relationship (SAR)[24]. The relatively weak nature of protein contacts with the side chain (N1-substituent) explains the modest improvement in IC_{50} values seen for compounds modified at this position[24]. It is also understandable how substitutions restricting the conformational freedom of the side chain—for example, introduction of a methyl group at the 7-position of the thieno[3,2-*d*]pyrimidine-2,4-dione system of 1—would significantly reduce binding affinity and therefore increase IC_{50}, as reported[24].

Inhibitor binding pose suggests a possible mode of action

Coordination of 1 to the metal ions that catalyze specific phosphodiester hydrolysis of the substrate suggested a mode of action for this inhibitor. We modeled ternary protein–inhibitor–DNA complexes using the present hFEN1-336Δ–inhibitor structure together with the published hFEN1-336Δ–product DNA complex[2] (**Fig. 1d**). Alignment of product-bound and ligand-bound structures indicated that the inhibitor and the phosphate monoester of the product DNA strand co-locate to bind the metal ions. Conversely, in the hFEN1-336Δ–substrate DNA complex[2], the scissile bond is not in contact with active site metal ions because the DNA is base-paired. Based on our previous work, it is assumed that a prereactive complex forms initially that requires the end of the DNA duplex to unpair and bind to metal ions as a prerequisite for cleavage[1,2,28]. Hence, we considered it plausible that substrate could bind in the presence of inhibitor, but that this would prevent DNA from accessing the catalytic metals as required for hydrolysis to occur (**Fig. 2a**). An alternative hypothesis was that the inhibitor precludes DNA binding, even though the compound is bound far from the other two main areas of protein–DNA interaction (the K^+/H2TH motif and the 3′ flap binding pocket). We undertook further work to characterize the hFEN1–inhibitor interaction and establish whether the *N*-hydroxyurea inhibitors compete with substrate DNA binding.

Inhibitor binding to hFEN1 requires magnesium ions

We used isothermal titration calorimetry (ITC) to quantify the interaction of 1, and related analogs 2 and 3 (ref. 22) bearing a smaller or no side chain (**Fig. 1a**), with the substrate-free protein (**Supplementary Table 2**). Similar dissociation constants (K_d) were obtained for 1 and 2 in the presence of Mg^{2+} with either hFEN1-336Δ (**Supplementary Fig. 2**) or full-length hFEN1 (**Supplementary Fig. 3a**,**b**), but the K_d of 3 was approximately ten-fold higher, suggesting that interactions between the side chains of 1 and 2 and the protein contribute to binding.

Ca^{2+} ions are often employed as a nonviable cofactor in biophysical measurements with hFEN1 because they facilitate accommodation of the substrate DNA and its required conformational changes[28,29], but they do not support catalysis. In fact, Ca^{2+} ions inhibit 5′-nuclease reactions by competing with Mg^{2+} (refs. 30,31), implying that the two ions occupy similar sites on the protein. However, K_d values were drastically increased upon replacement of Mg^{2+} with Ca^{2+} (**Supplementary Fig. 4**), showing that the latter did not support inhibitor binding. Thus, in accord with the crystal structure, interaction of **1** and **2** with hFEN1 was specific to the nuclease core domain and required Mg^{2+}. To estimate residence time, we probed the interaction of **1** with hFEN1-336Δ using surface plasmon resonance (**Supplementary Fig. 2d**). We obtained a dissociation constant similar to that obtained from ITC, and a residence time of 3 min.

Inhibitors bind to both protein and protein–DNA complex

Kinetic experiments were used to characterize hFEN1 inhibition by **1**, **2** and **4**. We measured rates of hFEN1-336Δ-catalyzed reaction with an optimal endonucleolytic double-flap substrate bearing a 6-carboxyfluorescein (FAM) label at the 5′ terminus[32] (DF1; **Fig. 2a** and **Supplementary Fig. 5a**). At substrate concentration close to K_m (100 nM), IC_{50} values for all three compounds were similar (**Table 1**), and a related exonucleolytic substrate gave similar IC_{50} results (**Supplementary Fig. 6a**,**b**). We determined the mode of inhibition by globally fitting rates of reaction at varying inhibitor and double-flap substrate concentrations to four inhibition models: competitive, uncompetitive, noncompetitive and mixed inhibition (equations (2–5) in Online Methods).

The uncompetitive model, where the inhibitor can bind only to enzyme–substrate complex, afforded a poor fit for **1**, which was unsurprising given the compound's high affinity for free protein. The competitive model, where binding of inhibitor and binding of substrate are mutually exclusive, also proved unsuitable, but the mixed and noncompetitive models produced acceptable fits (**Fig. 2b–d** and **Supplementary Fig. 7**). These models both assume the inhibitor can bind to DNA-free and DNA-bound forms of the enzyme, but the noncompetitive model (equation (4) in Online Methods) assumes both complexes have equivalent ligand dissociation constants. Allowing the dissociation constants to differ (mixed inhibition; **Fig. 2b** and equation (5) in Online Methods) produced a marginally better data fit, yielding near-equivalent dissociation constants for **1** (**Table 1**). Statistical model selection using Aikake's information criteria (AIC) overwhelmingly preferred the mixed inhibition model.

For **2**, only the competitive (equation (3) in Online Methods) and mixed inhibition models produced acceptable fits (**Fig. 2e** and **Supplementary Fig. 8**). The same statistical criteria (AIC) again favored the mixed model, but in this case the derived dissociation constants (K_{ic} and K_{iu}) differed by an order of magnitude (**Table 1**). For **4**, only the competitive model produced an acceptable fit (**Fig. 2f**, **Table 1** and **Supplementary Fig. 9**). Thus, whereas **1**, **2** and **4** all bound to hFEN1–Mg^{2+} with similar efficiency, only **1** showed notable affinity for the enzyme–substrate complex (hFEN1–Mg^{2+}–DNA), binding both DNA-free and DNA-bound forms of the enzyme with comparable dissociation constants.

Evidence for an hFEN1-Mg^{2+}-inhibitor-DNA complex

To verify formation of a quaternary complex of enzyme–Mg^{2+}–inhibitor–DNA (E–Mg^{2+}–I–DNA), we tested whether E–Mg^{2+}–I can form complexes with DNA without significant hydrolysis of the substrate occurring. High concentrations of **1** or **2** (100 μM) slowed the rate of Mg^{2+}-catalyzed reaction 10,000-fold under single-turnover conditions (**Table 1** and **Supplementary Fig. 10**), but appreciable substrate cleavage was still seen over the timescale required for biophysical measurements. Because Ca^{2+} did not support inhibitor binding (**Supplementary Fig. 4**), substituting it in place of Mg^{2+} as a nonviable cofactor was not applicable. Instead, we employed a previously characterized hFEN1 mutant, hFEN1-R100A. Arg100 is strictly conserved in FEN1 proteins, and its mutation to alanine slows reaction 7,000-fold[33]. The half-life of substrate with hFEN1-R100A–Mg^{2+} and inhibitors was sufficiently long to permit measurements without significant product formation (**Supplementary Fig. 11**), and ITC confirmed that the mutation did not affect inhibitor binding (**Supplementary Table 2**).

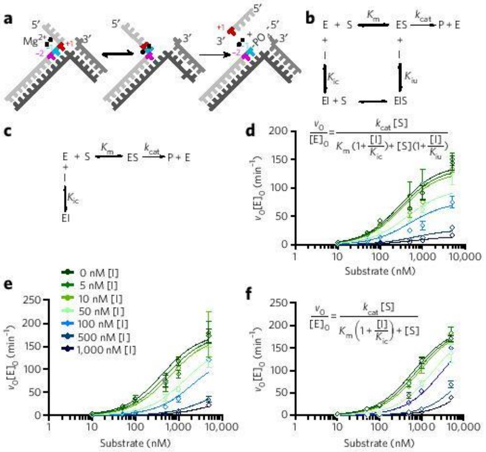


Figure 2 | Differences in inhibition characteristics of the compounds. (**a**) Schematic of hFEN1-catalyzed reaction showing double nucleotide unpairing at positions +1 and −1 (numbering relative to scissile phosphate). (**b**,**c**) Reaction schemes of mixed inhibition (**b**) and competitive inhibition (**c**) models. E, S, I and P represent enzyme, substrate, inhibitor and product, respectively. K_{ic} is the dissociation constant of I from free enzyme (competitive with substrate), and K_{iu} is the dissociation constant of I from ES complex (uncompetitive). (**d**-**f**) Nonlinear regression plots of normalized initial rates of reaction versus concentration of the substrate DF1 (open diamonds) at various concentrations of **1** (**d**; inset shows equation for mixed inhibition model), **2** (**e**; inset shows legend for inhibitor concentrations) or **4** (**f**; inset shows equation for competitive inhibition model). Error bars represent s.e.m. from global fitting of combined data from two triplicate experiments (fits to alternative models are shown in **Supplementary Figs. 7**–**9**).

Both **1** and **2** formed hFEN1-R100A–Mg^{2+}–I–DNA complexes as demonstrated by increases in anisotropy (r) of DF1 substrate upon titration with hFEN1-R100A–Mg^{2+}–I, with r reaching a common limiting value at high enzyme concentration (**Fig. 3a**). Data fitting to a simple binding isotherm revealed that in the absence of inhibitor, hFEN1-R100A and its wild-type equivalent (with which the use of noncatalytic Ca^{2+} ions was necessary to prevent reaction) had similar K_d values (**Supplementary Figs. 11a** and **12a–d**). However, substrate dissociation constants differed between quaternary complexes containing **1** or **2** (**Fig. 3a**): with **1**, hFEN1-R100A–Mg^{2+}–**1** had a K_d only three-fold greater than that of hFEN1-R100A–Ca^{2+}. In contrast, the substrate bound ten-fold more weakly to hFEN1-R100A–Mg^{2+}–**2**. These results were consistent with **1** having a K_{iu} value closer to its K_{ic} value than did **2**, again suggesting that **2** is more competitive than mixed in character. By competing away bound, FAM-labeled substrate with its unlabeled equivalent, we demonstrated specific interaction between hFEN1-R100A and this substrate (**Supplementary Figs. 5a**,**b** and **11h**).

DNA is bent in complexes with or without inhibitors

hFEN1 possesses two juxtaposed double-stranded DNA binding sites that accommodate double-flap substrate DNA in a conformation

Table 1 | Kinetic parameters in the absence and presence of inhibitors

Enzyme	Inhibitor	IC_{50}, nM[a]	k_{cat} (min^{-1})	K_m (nM)	K_{ic} (nM)	K_{iu} (nM)	k_{STmax} (min^{-1})	$t_{1/2}$ (min)	ΔAIC_c
hFEN1	None	NA	165 ± 9	20 ± 3	NA	NA	916 ± 49	7.57×10^{-4}	NA
hFEN1	**1**	ND	ND	ND	ND	ND	0.48 ± 0.04	1.43	NA
hFEN1	**2**	ND	ND	ND	ND	ND	1.52 ± 0.09	0.46	NA
hFEN1-336Δ	None	NA	160 ± 10	151 ± 16	NA	NA	755 ± 35	8.94×10^{-4}	NA
hFEN1-336Δ	**1**	46.4 ± 4.8	140 ± 9	297 ± 31	48 ± 5	117 ± 27	ND	ND	24.76[b]
hFEN1-336Δ	**2**	30.0 ± 6.0	182 ± 13	422 ± 50	17 ± 2	306 ± 125	ND	ND	10.21[c]
hFEN1-R100A	None	NA	ND	ND	ND	ND	0.087 ± 0.003	7.94	NA
hFEN1-R100A	**1**	ND	ND	ND	ND	ND	$\sim 4 \times 10^{-4}$	≤1750	NA
hFEN1-R100A	**2**	ND	ND	ND	ND	ND	$\sim 2 \times 10^{-3}$	≤360	NA
hFEN1-336Δ	**4**	16.9 ± 1.2	194.5 ± 11	630.8 ± 53	26 ± 2	NA	ND	ND	Amb.

k_{STmax} is maximal reaction rate under single-turnover conditions, used to calculate the substrate half-life ($t_{1/2}$). ΔAIC_c is the difference between second-order (corrected) Akaike information criteria values between models; if $\Delta AIC_c \geq 6$, the likelihood that the incorrect model was selected is $P < 0.0001$. NA, not applicable; ND, not determined; Amb., ambiguous.

[a] IC_{50} values were derived from rates at a substrate concentration close to K_m (100 nM). [b] ΔAIC_c for **1** compares noncompetitive with mixed-inhibition models. [c] ΔAIC_c for **2** compares competitive with mixed-inhibition models. Mixed-inhibition is preferred for both **1** and **2**. For **4**, competitive inhibition was the only model whose fit was not ambiguous.

with a 100° bend at the junction. To ascertain whether DNA bound similarly in the presence of inhibitor, we examined substrate bending using fluorescence resonance energy transfer (FRET). We labeled double-flap substrate with a rhodamine–FAM dye pair on its respective duplexes and verified that binding to hFEN1 produces an increase in FRET signal[34] (**Fig. 3b**, **Supplementary Figs. 5c–f, 13** and **14**). Titration of hFEN1-R100A–Ca^{2+} or hFEN1-R100A–Mg^{2+}–**1** into the labeled substrate produced comparable FRET-efficiency start and end values (**Fig. 3b**), confirming that the enzyme had engaged both DNA binding sites with or without inhibitor. The substrate K_d was raised by a factor of three in the presence of **1**, whereas substrate binding was much weaker with **2** present (**Fig. 3b** and **Supplementary Table 3**); hence, these results mirrored those obtained earlier by fluorescence anisotropy.

Inhibitors bound to catalytic metals block DNA unpairing

Unpairing of the reacting substrate duplex, which places the target phosphodiester onto active site metal ions, is a prerequisite for the hFEN1-catalyzed reaction one nucleotide into the double-stranded DNA (**Fig. 2a**)[28]. This metal ion-dependent conformational change may be monitored using substrates that contain a tandem 2-aminopurine (2AP) exciton pair at the −1 and −2 positions of the 5′ flap strand (DF3; **Supplementary Fig. 5g**) by measuring changes in the low-energy exciton-coupled CD spectrum resulting from the 2APs, usually in the presence of Ca^{2+} to prevent reaction[28].

In adopting the reactive conformation, the +1 and −1 nucleotides are assumed to become extrahelical, whereas the −2 nucleotide remains base-paired. In the absence of active site divalent ions (EDTA added), a strong maximum at 330 nm is observed from the hFEN1-R100A–DNA complex, resulting from the exciton pair and consistent with substrate remaining base-paired[28]. With hFEN1-R100A–Ca^{2+}–DNA, the DNA conformational change reversed the sign of the CD signal, producing a deep minimum at 310 nm (**Fig. 4a**). In the presence of **1** or **2**, the measured CD signal of hFEN1-R100A–Mg^{2+}–I–DNA did not differ substantially from that observed for hFEN1-R100A–DNA without divalent ions (**Fig. 4b,c**), even though the DNA was assumed to be fully bound under these conditions (10 μM DNA, 12.5 μM hFEN1-R100A). This demonstrated that the inhibitors prevented substrate conformational rearrangements necessary for hydrolysis (**Supplementary Fig. 15**).

N-Hydroxyurea FEN1 inhibitors also target EXO1

FEN1 is the prototypical member of the structure-specific 5′-nuclease superfamily, also comprising exonuclease-1 (EXO1), gap endonuclease-1 (GEN1) and xeroderma pigmentosum complementation group-G protein (XPG)[1]. ExoRNases XRN1 and XRN2 are also suggested members of the superfamily[1]. These nucleases all share a similarly folded nuclease domain with similar active site geometry and full conservation of essential catalytic residues[1,2]. Consequently, it has been hypothesized that the substrate selectivity of these proteins stems from strict recognition of their respective DNA substrate structures, followed by double nucleotide unpairing to initiate scissile phosphate diester hydrolysis[1].

It is known that hFEN1 inhibitors can exhibit limited but manageable promiscuity toward XPG[24]. However, testing against human EXO1-352Δ (nuclease domain of EXO1)[35] revealed that **1** and **2** both inhibited this target with IC_{50} values similar to those against hFEN1 (**Supplementary Figs. 5k** and **6a**,**e**). Differential scanning fluorimetry experiments[36] further confirmed binding of both compounds to both proteins in a divalent-metal-ion-dependent manner (**Supplementary Fig. 6g,h**). In contrast, inhibitor **1** was found ineffective against bacteriophage T5 FEN (**Supplementary Figs. 5l** and **6c**) and *Kluyveromyces lactis* XRN1 (**Supplementary Fig. 16**), both of which show a high level of active site conservation with the mammalian 5′-nuclease superfamily[1]. Similarly, **1** did not inhibit the structurally unrelated DNA-repair metallonuclease APE1 (**Supplementary Fig. 6f**).

When hFEN1 acts *in vivo* it is usually associated with the toroidal clamp PCNA. PCNA increases the stability of FEN1–DNA

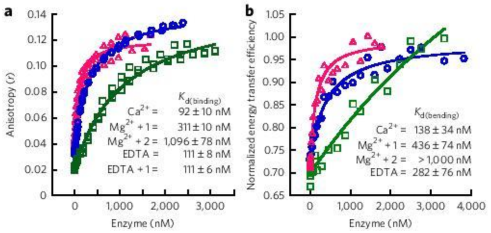


Figure 3 | Effect of inhibitors on substrate binding assessed by fluorescence anisotropy and FRET. (**a**) Typical fluorescence anisotropy titration data for hFEN1-R100A binding DF1 in the presence of 10 mM Ca^{2+} (magenta, open triangles), 8 mM Mg^{2+} plus 100 μM **1** (blue, open circles) or 8 mM Mg^{2+} with 100 μM **2** (green, open squares). (**b**) Typical normalized FRET binding data for DF1 and hFEN1-R100A. Colors and symbols are as in **a**. Plots in **a**,**b** are representative of experiments repeated independently three times.

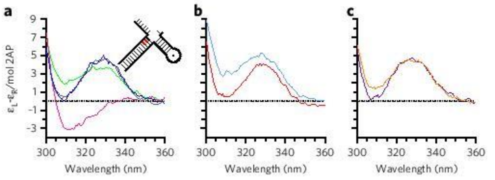


Figure 4 | *N*-Hydroxyurea inhibitors prevent FEN1 reaction by blocking substrate unpairing. (**a**–**c**) CD spectra recorded at pH 7.5 and 20 °C. (**a**) Tandem 2-aminopurine-containing substrate DF3 (illustrated schematically as inset and in **Supplementary Fig. 5g**) alone in the presence of 10 mM Ca^{2+} (blue) or 25 mM EDTA (gray), and the same substrate bound to hFEN1-R100A in the presence of 10 mM Ca^{2+} (magenta) or 25 mM EDTA (green). (**b**) DF3 bound to hFEN1-R100A in the presence of Mg^{2+} plus excess **1** (cyan) or EDTA plus excess **1** (red). (**c**) DF3 bound to hFEN1-R100A with excess **2** in the presence of Mg^{2+} (orange) or EDTA (purple). Full DNA sequences are shown in **Supplementary Tables 5** and **6** and **Supplementary Figure 5g**. Plots in **a**–**c** are representative of experiments repeated independently three times.

complexes[34], suggesting that association with PCNA might allow FEN1 to overcome inhibition. However, when we added human PCNA to hFEN1 reactions inhibited by 1 or 4, the slow rates of reaction we observed did not increase, implying the FEN1 interaction partner does not dramatically influence the IC_{50} of either compound (**Supplementary Fig. 6d**).

N-Hydroxyurea inhibitors engage with hFEN1 in live cells

Because of their contrasting inhibition modes, we selected 1 and 4 for additional cellular studies. We employed the cellular thermal shift assay (CETSA) technique[37] to establish whether they interacted with hFEN1 in SW620 colon cancer cells. CETSA detects changes in the stability of a protein upon engagement with a ligand, like a biochemical thermal-shift assay, but is performed with whole cells and a target-specific, label-free readout of engagement is obtained using a relevant antibody. In an isothermal concentration–response experiment, 1 and 4 stabilized hFEN1 (**Fig. 5a–c** and **Supplementary Fig. 17**) with similar EC_{50} values of 5.1 μM and 6.8 μM, respectively, despite their differing modes of inhibition. Interestingly, these micromolar-range values represented a substantial drop-off in observed binding affinity compared with observations in prior biochemical assays (IC_{50} = 46 nM and IC_{50} = 17 nM, respectively; **Table 1**), so we undertook a number of experiments to attempt to explain this. Cell permeability in MDCK and Caco-2 assays was not an issue (**Supplementary Table 4**); neither were other properties including solubility and chemical stability (data not shown). The compounds' affinity for free divalent metal ions in solution was insignificant, ruling out metal chelation as an explanation. Nonspecific protein binding may have contributed to the discrepancy between biochemical and phenotypic potency, but binding to other 5′-nuclease superfamily members represented the most obvious potential for off-target effects. Hence, we attempted further CETSA studies with 1 and 4 against human EXO1 (hEXO1), but we found this to be a nonviable CETSA target (with only fragments of the protein detected on the blots), perhaps reflecting instability of the protein under the assay conditions or its cellular context as a component of multiprotein complexes, which regulate its activity.

hFEN1 inhibition activates the DNA damage checkpoint

High concentrations of 1 proved cytotoxic toward SW620 cells, with an EC_{50} of 11 μM (**Fig. 5e**), but HeLa cells stably expressing a short hairpin RNA (shRNA) targeting hFEN1 were 70% viable at 20 μM 1 (**Fig. 5f**, orange). HeLa cells expressing a mock shRNA control were only 15% viable under the same conditions (**Fig. 5f**, black), showing a susceptibility to 1 similar to that of untransformed cells (**Fig. 5g**). Hence, a lack of hFEN1 conferred resistance to 1, suggesting that on-target activity was the primary cause of cytotoxicity.

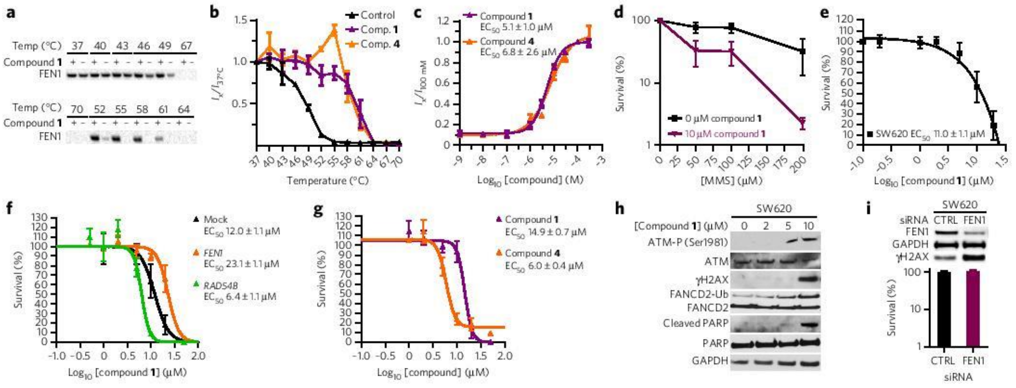


Figure 5 | Cellular engagement and activity of hFEN1 inhibitors 1 and 4. (**a**) Representative western blot data from a melt curve for **1**. Plus signs, treated sample; minus signs, control. (**b**) Melt and shift curves for FEN1 in intact SW620 cells with 100 μM **1** (purple), **4** (orange) or DMSO (control, black). (**c**) Ratio of hFEN1 protein isothermal shifts in cells exposed to 50 °C, plotted against concentration of **1** (purple) or **4** (orange). These curves were used to quantify target engagement of FEN1 in intact treated SW620 cells. (**d**) MMS sensitivity of SW620 cells treated with a continuous dose of 10 μM **1** (purple) or DMSO (control, black). (**e**) Dose-dependent sensitivity of SW620 cells to **1**. (**f**) Sensitivity to **1** of HeLa cells stably expressing *FEN1* (orange), *RAD54B* (green) or nontargeting (black) shRNA. (**g**) Dose-dependent sensitivity of HeLa cells to **1** and **4**. (**h**) Typical western blots showing that **1** induces a DNA damage response in a dose-dependent manner. (**i**) SW620 cells are insensitive to knockdown of FEN1 by siRNA, but accumulate DNA damage. **a**, **h** and **i** show blot images representative of three independent experiments. **b** and **c** show data from three independent triplicate experiments, fitted globally (i.e. N = 3, n = 9) with s.e.m. **d**–**g** and **i** show the means of three independent experiments ± s.e.m. Full images of cut gels used to prepare **h** and **i** are included in **Supplementary Figures 18** and **19**, respectively.

SW620 cells also showed increased sensitivity to MMS when cotreated with 1, in a dose-dependent manner (**Fig. 5d**), suggesting that the compound inhibits the LP-BER function of FEN1 in a cellular context. Enhanced toxicity of 1 toward HeLa cells expressing *RAD54B*-shRNA (**Fig. 5f**, green) was also observed, with an EC_{50} of 6.4 μM compared to 14.9 μM against untransformed cells (**Fig. 5e**,**g**), confirming the synthetic-lethal interaction between *FEN1* and *RAD54B* previously demonstrated by silencing of the former[18]. Inhibitor 4 also proved cytotoxic to HeLa cells ($EC_{50} = 6$ μM; **Fig. 5g**), appearing more potent than 1, whose EC_{50} of approximately 15 μM was in line with its toxicity against SW620 cells.

When treated with sublethal doses of 1, SW620 cells showed evidence of an induced DNA damage response (**Fig. 5h** and **Supplementary Fig. 18**) at concentrations consistent with the EC_{50} for target engagement observed by CETSA. The same compound effected a dose-dependent increase in ubiquitination of FANCD2, a marker for activation of the Fanconi anemia pathway recruited to stabilize stalled replication forks[38–40]. At higher doses, accumulation of phosphorylated ATM kinase and phosphorylated histone γH2AX was evident, indicating accumulation of unrepaired DNA double-strand breaks (DSBs). Cells treated with high concentrations of 1 also showed evidence of apoptosis, specifically the presence of cleaved PARP (**Fig. 5h**). Knockdown of hFEN1 by short interfering RNA (siRNA) activated a similar DNA damage response to that resulting from treatment with 1; these cells accumulated phosphorylated γH2AX but otherwise remained viable (**Fig. 5i** and **Supplementary Fig. 19**). DNA damage response activation and apoptosis were consistent with loss of hFEN1 function, because the consequences of unprocessed Okazaki fragments would include stalled or collapsed replication forks, replication errors and DSBs.

DISCUSSION

N-Hydroxyurea compounds 1, 2 and 4 prevented DNA cleavage with similar efficiency (**Table 1**), reflecting the SAR observed previously for similarly sized compounds[24] inasmuch as comparable IC_{50} values were obtained despite notable differences in side chain size and structure. These results were consistent with protein–inhibitor binding mediated primarily through interaction with active site Mg^{2+} ions, and a lack of strong contacts between the protein and inhibitor side chain, as seen in the structure of 1 bound to hFEN1 (**Fig. 1** and **Supplementary Table 2**). Although the metal-coordinating headgroup clearly provided the predominant binding contribution, the elevated K_d of 3 suggested interaction of the inhibitor side chain with the protein was nonetheless important for optimal affinity. Further studies revealed subtle differences in mode of action resulting from variation of the side chain structure.

Although the DNA substrate bound in its usual conformation in the presence of 1, hydrolysis was impaired by prevention of double nucleotide unpairing through steric blocking of the catalytic metals (**Figs. 1b–d**, **3** and **4**). These observations were reminiscent of the action of the HIV integrase inhibitor raltegravir[41]. Raltegravir and functionally related compounds bind to active site metal ions of the integrase–DNA complex, similarly obstructing access of the reacting phosphodiester bond to the metals. In contrast, 2 and 4, with altered side chains, proved mostly competitive in character and primarily acted to reduce affinity of the enzyme for its DNA substrate.

The micromolar EC_{50} values seen in CETSA experiments with 1 and 4 differed markedly from the compounds' nanomolar potency against purified protein, though they were consistent with phenotypic potency in DNA damage induction and cytotoxicity assays. A clear explanation for this was not found, but the elevated cellular EC_{50} values might reflect a high local concentration of hFEN1 in the nucleus during S phase, which could conceivably reach the micromolar range. The residence time of 1 on hFEN1 proved similar to that of raltegravir on its target (4.8 min)[42], although this is short compared to the median of 51 min for a representative set of marketed drugs[42]. The short residence time of 1 may necessitate a high local drug concentration in the vicinity of the target for effective inhibition in cells.

Although hEXO1 is likely inhibited alongside hFEN1, the cellular concentration of hEXO1 is not expected to be significantly higher, so this seems an unlikely explanation for the raised EC_{50} values. The results with hFEN1-deficient cells (**Fig. 5f**) did suggest some degree of target specificity. However, previous cellular studies assuming selective inhibition of hFEN1 by the *N*-hydroxyurea series of inhibitors must nonetheless be interpreted with caution considering the likelihood of parallel hEXO1 inhibition, since it is not possible to distinguish between phenotypes of hFEN1 and hEXO1 inhibition with this class of compounds. One such published[24] inhibitor, related to 1–4, has been employed to help validate a role for hFEN1 in homologous recombination (HR)[43], as treatment with the inhibitor induces deficient HR. However, hEXO1 is essential for competent HR[44–47], and the observed phenotype is explicable by inhibition of this enzyme alone. Although a role for hFEN1 in HR is otherwise supported in that study, we conclude that the *N*-hydroxyurea series should not be regarded as exclusive hFEN1 inhibitors.

The mixed inhibition mode of 1, which in theory permits 'dead-end' complexes of DNA and protein to form, did not confer any advantageous inhibition characteristics in cells. Unprocessed Okazaki fragments resulting from hFEN1 inhibition might be successfully repaired by the cell, with apoptosis resulting only when the DNA damage response is overwhelmed. Some support for this notion was seen in SW620 cells treated with 1, in which we observed dose-dependent activation of the Fanconi anemia pathway (**Fig. 5h**). Given that FANCD2 is recruited to stabilize stalled replication forks and initiate repair[38], treatment with 1 evidently did interrupt replication, prompting cells to activate other pathways to repair unprocessed Okazaki fragments directly. Failure to achieve this may cause collapse of replication forks into DSBs, and at higher doses of 1, we did see evidence for activation of DSB repair. These markers did not accumulate at lower doses, so the damage signal may be obvious only when the frequency of DSBs overwhelms the cell's DNA damage response. Accumulation of cleaved PARP, indicating early apoptosis, also suggested cells exposed to 1 were accumulating DNA damage associated with inhibition of hFEN1 and/or hEXO1 and were signaling for apoptosis.

Without exposure to inhibitor, both SW620 cells treated with *FEN1*-siRNA and HeLa cells stably expressing hFEN1-shRNA showed viability indistinguishable from untransformed controls, yet they constitutively initiated a DNA damage response (**Fig. 5i**). The *FEN1*-shRNA cells showed reduced sensitivity to 1, suggesting a degree of selectivity and on-target activity for the compound, since the DNA damage response remained competent. Our data suggest that removal of functional hFEN1 alone does not induce toxicity and that damage associated with its loss is successfully repaired until such mechanisms become overwhelmed. This result, alongside our other observations in human cells, suggests that targeting of hFEN1 in cancer will not prove effective as a monotherapy but could be useful in exploiting synthetic-lethal vulnerabilities. Synthetic-lethal interactions between *Fen1* and *Rad54b* (ref. 18), *Cdc4* (ref. 16) and *Mre11a* (ref. 16) are established, and other such interactions with potential clinical relevance are proposed[16,38]. We confirmed synthetic-lethal interaction with *Rad54b*, previously established using *Fen1* knockdown[18], through inhibition of hFEN1 by 1. Thus, hFEN1 inhibitors might prove beneficial as a component of targeted or personalized therapies, provided selectivity over hEXO1 and the other 5′-endonuclease superfamily members can be realized.

Received 1 July 2015; accepted 19 May 2016;

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Protein Data Bank: Coordinates and structure factors have been deposited in the Protein Data Bank under accession code 5FV7.

References

1. 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).
2. 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).
3. Kim, C.Y., Shen, B., Park, M.S. & Olah, G.A. Structural changes measured by X-ray scattering from human flap endonuclease-1 complexed with Mg^{2+} and flap DNA substrate. *J. Biol. Chem.* **274**, 1233–1239 (1999).
4. Zheng, L. *et al.* Functional regulation of FEN1 nuclease and its link to cancer. *Nucleic Acids Res.* **39**, 781–794 (2011).
5. Waga, S., Bauer, G. & Stillman, B. Reconstitution of complete SV40 DNA replication with purified replication factors. *J. Biol. Chem.* **269**, 10923–10934 (1994).
6. Bambara, R.A., Murante, R.S. & Henricksen, L.A. Enzymes and reactions at the eukaryotic DNA replication fork. *J. Biol. Chem.* **272**, 4647–4650 (1997).
7. Kim, K., Biade, S. & Matsumoto, Y. Involvement of flap endonuclease 1 in base excision DNA repair. *J. Biol. Chem.* **273**, 8842–8848 (1998).
8. Parikh, S.S., Mol, C.D., Hosfield, D.J. & Tainer, J.A. Envisioning the molecular choreography of DNA base excision repair. *Curr. Opin. Struct. Biol.* **9**, 37–47 (1999).
9. Beard, W.A. & Wilson, S.H. Structure and mechanism of DNA polymerase Beta. *Chem. Rev.* **106**, 361–382 (2006).
10. Mohan, V. & Srinivasan, M. in *New Research Directions in DNA Repair* (ed. Clark Chen) (InTech, 2013).
11. Singh, P. *et al.* Overexpression and hypomethylation of flap endonuclease 1 gene in breast and other cancers. *Mol. Cancer Res.* **6**, 1710–1717 (2008).
12. 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).
13. Nikolova, T., Christmann, M. & Kaina, B. FEN1 is overexpressed in testis, lung and brain tumors. *Anticancer Res.* **29**, 2453–2459 (2009).
14. Panda, H. *et al.* Amino acid Asp181 of 5′-flap endonuclease 1 is a useful target for chemotherapeutic development. *Biochemistry* **48**, 9952–9958 (2009).
15. Yoshimoto, K. *et al.* Complex DNA repair pathways as possible therapeutic targets to overcome temozolomide resistance in glioblastoma. *Front. Oncol.* **2**, 186 (2012).
16. 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).
17. Illuzzi, J.L. & Wilson, D.M. III. Base excision repair: contribution to tumorigenesis and target in anticancer treatment paradigms. *Curr. Med. Chem.* **19**, 3922–3936 (2012).
18. McManus, K.J., Barrett, I.J., Nouhi, Y. & Hieter, P. Specific synthetic lethal killing of RAD54B-deficient human colorectal cancer cells by FEN1 silencing. *Proc. Natl. Acad. Sci. USA* **106**, 3276–3281 (2009).
19. Durant, S.T. Telomerase-independent paths to immortality in predictable cancer subtypes. *J. Cancer* **3**, 67–82 (2012).
20. Hwang, J.-C. *et al.* The overexpression of FEN1 and RAD54B may act as independent prognostic factors of lung adenocarcinoma. *PLoS One* **10**, e0139435 (2015).
21. Shibata, Y. & Nakamura, T. Defective flap endonuclease 1 activity in mammalian cells is associated with impaired DNA repair and prolonged S phase delay. *J. Biol. Chem.* **277**, 746–754 (2002).
22. McWhirter, C. *et al.* Development of a high-throughput fluorescence polarization DNA cleavage assay for the identification of FEN1 inhibitors. *J. Biomol. Screen.* **18**, 567–575 (2013).
23. Dorjsuren, D., Kim, D., Maloney, D.J., Wilson, D.M. III & Simeonov, A. Complementary non-radioactive assays for investigation of human flap endonuclease 1 activity. *Nucleic Acids Res.* **39**, e11 (2011).
24. Tumey, L.N. *et al.* The identification and optimization of a N-hydroxy urea series of flap endonuclease 1 inhibitors. *Bioorg. Med. Chem. Lett.* **15**, 277–281 (2005).
25. Finger, L.D. *et al.* The wonders of flap endonucleases: structure, function, mechanism and regulation. *Subcell. Biochem.* **62**, 301–326 (2012).
26. Guo, Z. *et al.* Sequential posttranslational modifications program FEN1 degradation during cell-cycle progression. *Mol. Cell* **47**, 444–456 (2012).
27. Sakurai, S. *et al.* Structural basis for recruitment of human flap endonuclease 1 to PCNA. *EMBO J.* **24**, 683–693 (2005).
28. Finger, L.D. *et al.* Observation of unpaired substrate DNA in the flap endonuclease-1 active site. *Nucleic Acids Res.* **41**, 9839–9847 (2013).
29. 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).
30. Syson, K. *et al.* Three metal ions participate in the reaction catalyzed by T5 flap endonuclease. *J. Biol. Chem.* **283**, 28741–28746 (2008).
31. Tomlinson, C.G. *et al.* 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 (2011).
32. 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).
33. 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).
34. 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.* **42**, 1857–1872 (2014).
35. Orans, J. *et al.* Structures of human exonuclease 1 DNA complexes suggest a unified mechanism for nuclease family. *Cell* **145**, 212–223 (2011).
36. Niesen, F.H., Berglund, H. & Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat. Protoc.* **2**, 2212–2221 (2007).
37. Martinez Molina, D. *et al.* Monitoring drug target engagement in cells and tissues using the cellular thermal shift assay. *Science* **341**, 84–87 (2013).
38. Moldovan, G.L. & D'Andrea, A.D. How the fanconi anemia pathway guards the genome. *Annu. Rev. Genet.* **43**, 223–249 (2009).
39. Yeo, J.E., Lee, E.H., Hendrickson, E.A. & Sobeck, A. CtIP mediates replication fork recovery in a FANCD2-regulated manner. *Hum. Mol. Genet.* **23**, 3695–3705 (2014).
40. Schlacher, K., Wu, H. & Jasin, M. A distinct replication fork protection pathway connects Fanconi anemia tumor suppressors to RAD51-BRCA1/2. *Cancer Cell* **22**, 106–116 (2012).
41. Di Santo, R. Inhibiting the HIV integration process: past, present, and the future. *J. Med. Chem.* **57**, 539–566 (2014).
42. Dahl, G. & Akerud, T. Pharmacokinetics and the drug-target residence time concept. *Drug Discov. Today* **18**, 697–707 (2013).
43. Fehrmann, R.S.N. *et al.* Gene expression analysis identifies global gene dosage sensitivity in cancer. *Nat. Genet.* **47**, 115–125 (2015).
44. Bolderson, E. *et al.* Phosphorylation of Exo1 modulates homologous recombination repair of DNA double-strand breaks. *Nucleic Acids Res.* **38**, 1821–1831 (2010).
45. Desai, A., Qing, Y. & Gerson, S.L. Exonuclease 1 is a critical mediator of survival during DNA double strand break repair in nonquiescent hematopoietic stem and progenitor cells. *Stem Cells* **32**, 582–593 (2014).
46. Tomimatsu, N. *et al.* Phosphorylation of EXO1 by CDKs 1 and 2 regulates DNA end resection and repair pathway choice. *Nat. Commun.* **5**, 3561 (2014).
47. Tomimatsu, N. *et al.* Exo1 plays a major role in DNA end resection in humans and influences double-strand break repair and damage signaling decisions. *DNA Repair (Amst.)* **11**, 441–448 (2012).

Acknowledgments

This work was supported by U.K. Biotechnology and Biological Sciences Research Council grants BB/K009079/1 and BB/M00404X/1 (both to J.A.G.) and AstraZeneca. J.C.E. thanks the U.K. Engineering and Physical Sciences Research Council and AstraZeneca for a studentship. The authors thank C. Phillips for assistance with submissions of the crystallographic data and T. McGuire for synthetic support.

Author contributions

C.D.J. designed and synthesized inhibitors. J.C.E., M.J.T., L.D.F. and S.J.S. carried out kinetic and biophysical experiments. J.C.E., M.J.T., L.D.F., C.M., J.W.M.N., W.M.A. and J.A.G. designed experiments and analyzed this data. J.D. and J.W.M.N. obtained and analyzed structures. C.L.B.S. and D.M.M. performed the CETSA assays. T.A.W. carried out other cellular experiments, and T.A.W. and S.T.D. analyzed data. All authors contributed to the preparation of the manuscript.

Competing financial interests

The authors declare competing financial interests: details accompany the online version of the paper.

Additional information

Any supplementary information, chemical compound information and source data are available in the online version of the paper. Reprints and permissions information is available online at http://www.nature.com/reprints/index.html. Correspondence should be addressed to J.W.M.N., S.T.D. or J.A.G.

nature COMMUNICATIONS

ARTICLE

Received 18 Jan 2017 | Accepted 5 May 2017 | Published 27 Jun 2017 DOI: 10.1038/ncomms15855 OPEN

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).

The 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) 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 against the duplex DNA at the bend.

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

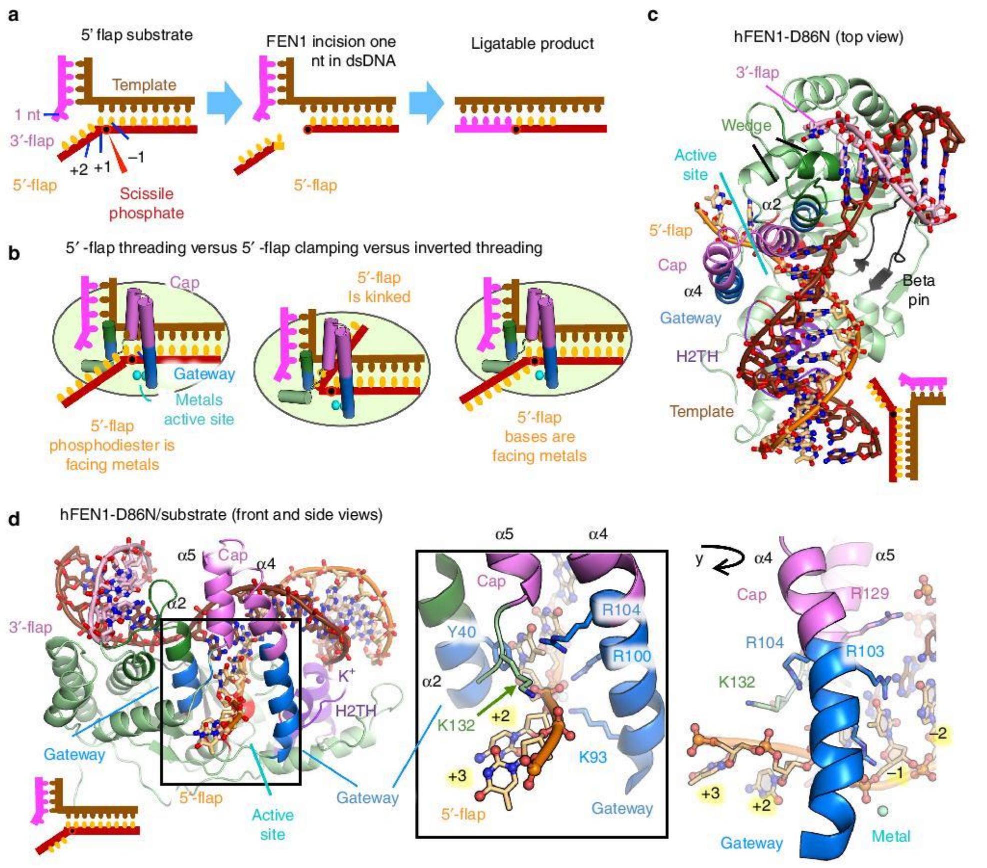


Figure 1 | Specificity and inverted threading of ss5′-flap in hFEN1 D86N substrate structure. (**a**) Schematic FEN1 incision on an optimal double-flap substrate, incising 1 nt into dsDNA to ensure a ligatable product. (**b**) Proposed models for ssDNA selection. (**c**) Top view of hFEN1-D86N crystal 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

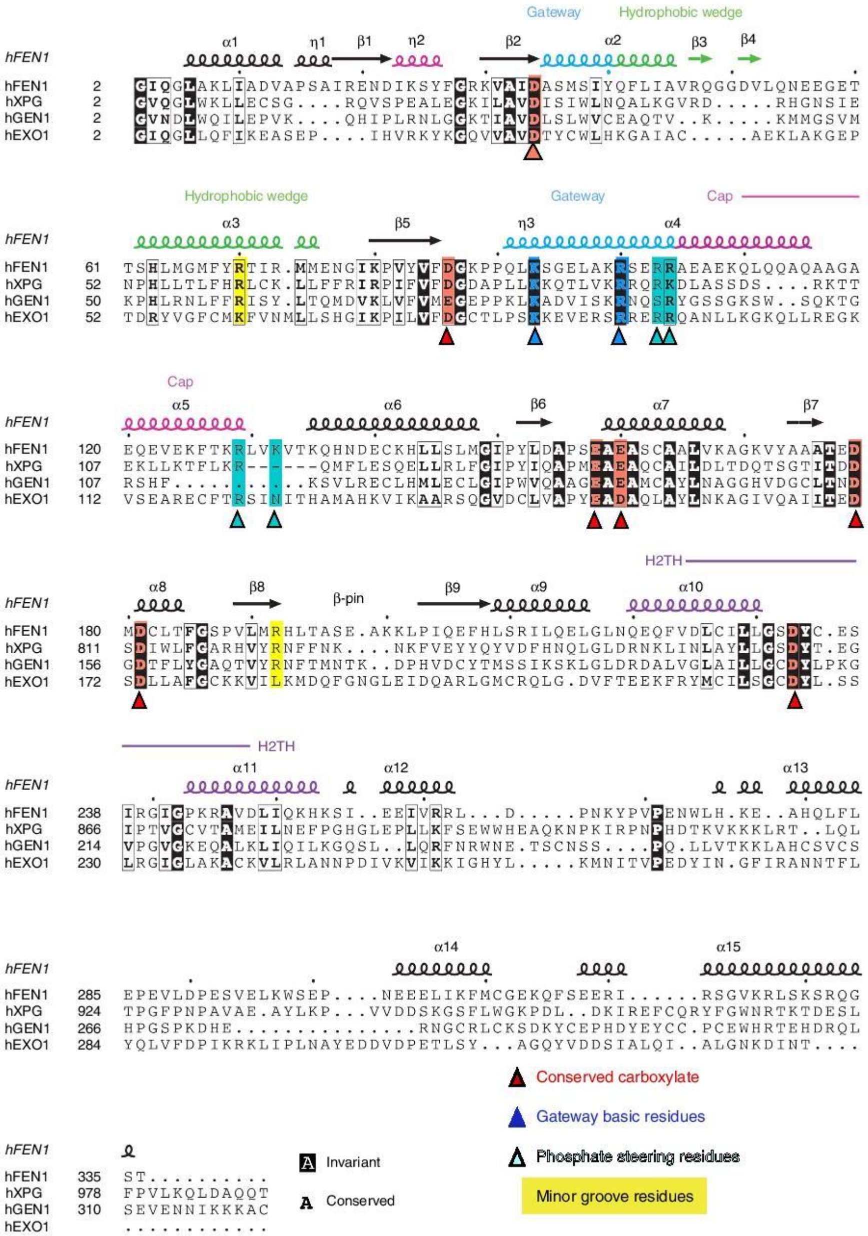


Figure 2 | FEN1 superfamily sequence and secondary structure alignment. Map of FEN1 secondary structure (PDB code 3q8k), structural elements, and mutants to a sequence alignment of FEN superfamily human members. XPG residues 117–766 were removed (dash) to facilitate alignment.

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			
a, *b*, *c* (Å)	105.8 105.8 100.7	105.2 105.25 104.1	105.2 105.2 104.5
α, β, γ (°)	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
Protein/DNA	3,492	3,519	3,545
Sm^{3+}, K^+	7	6	9
Water	36	173	625
B-factors		91	64
Protein/DNA	113.97	92	63
Sm^{3+}, K^+	144	91.5	93.3
Water	110	77.3	69
R.m.s. deviations			
Bond lengths (Å)	0.004	0.005	0.003
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

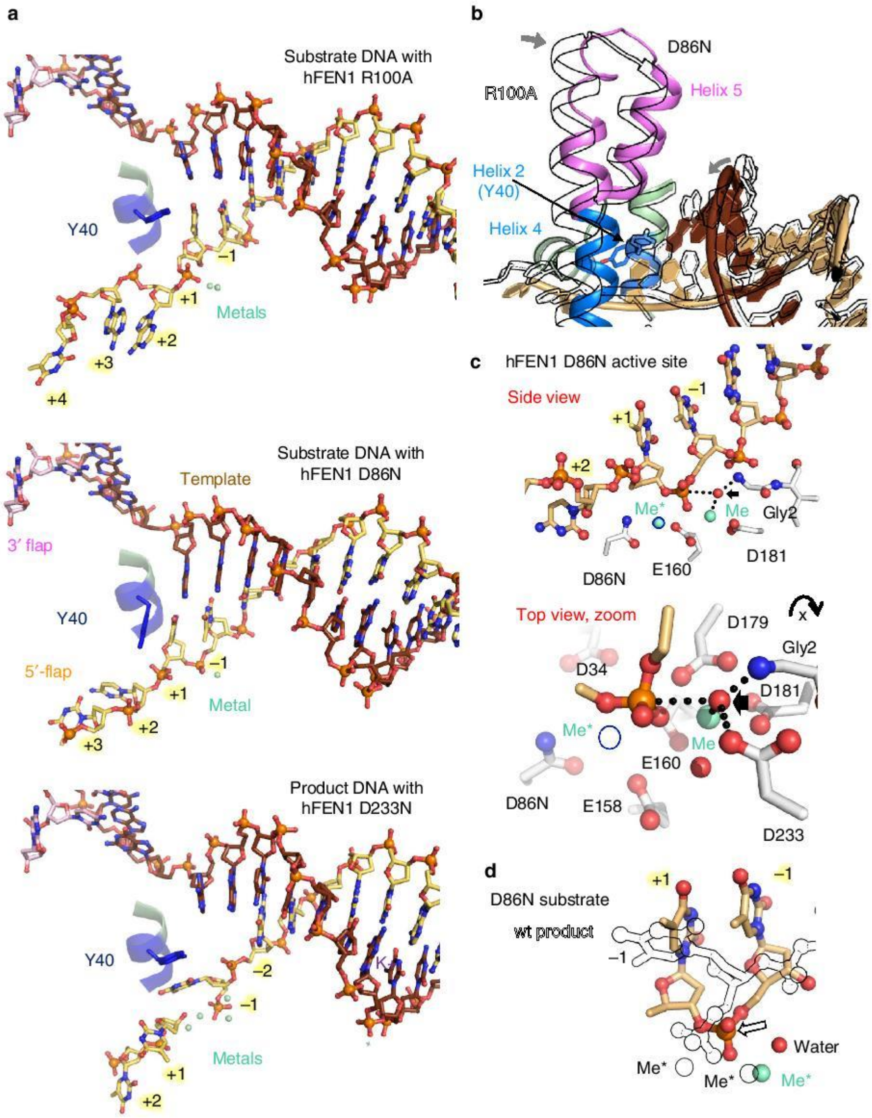


Figure 3 | Three FEN1 crystal structures show threading through the capped gateway. (**a**) DNA from hFEN1-D86N, hFEN1-R100A, and hFEN1-D233N structures showed threading not clamping. Tyr40 (stick model) changes its rotamer to track DNA movement through the gateway in the hFEN1-D86N structure. (**b**) 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,

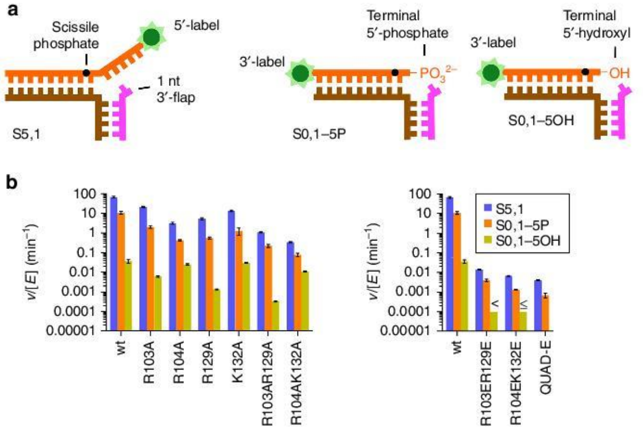


Figure 4 | Phosphate steering residue mutants show reduced activity. (**a**) Schematic of substrates used with fluorescent positions. (**b**) Comparison of multiple turnover rates for cleavage of each substrate at 50 nM. Reaction rates for glutamate mutants with S0,1-5OH were too slow to measure accurately, 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]).

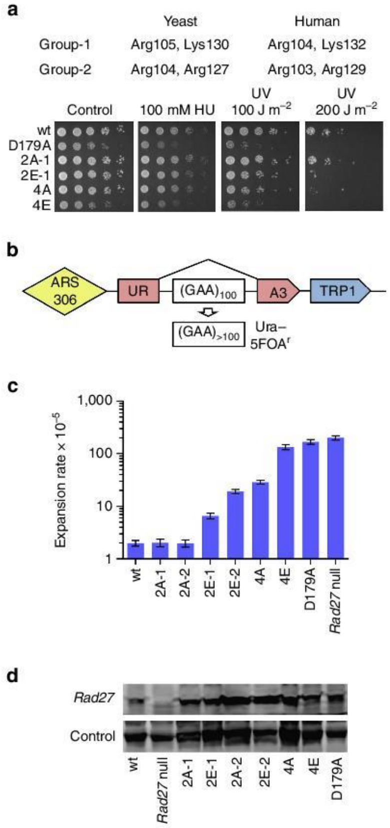


Figure 5 | Phenotypic and DNA repeat expansion defects of *Rad27* basic cap residue mutations in yeast *S. cerevisiae*. (**a**) Table of the tested basic residues in yeast (and their human counterpart) and spot-test (serial fivefold dilutions) for yeast growth with and without exposure to hydroxyurea (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 ($100\,\mathrm{J\,m^{-2}}$) was strongly deleterious to both strains, and higher-dose irradiation ($200\,\mathrm{J\,m^{-2}}$) 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

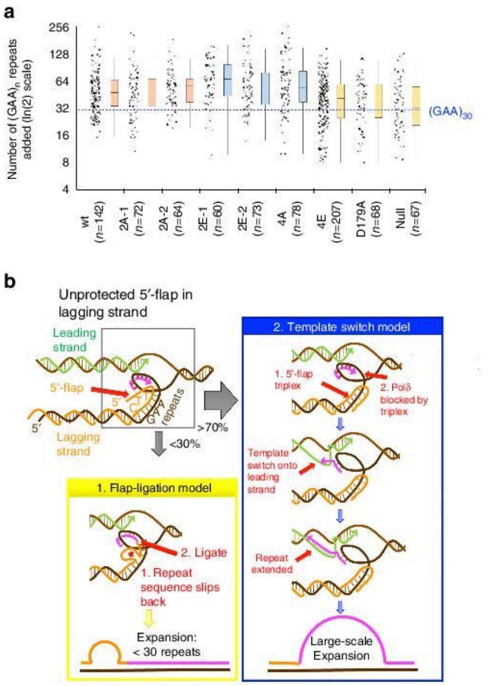


Figure 6 | FEN1 phosphate steering is essential for lagging strand precision at DNA repeats. (**a**) Graphed distributions of repeat expansion lengths shows that the majority of expansions in the wt, phosphate steering and D179A Rad27 mutants are >30 repeats. The numbers of added repeats in each strain 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

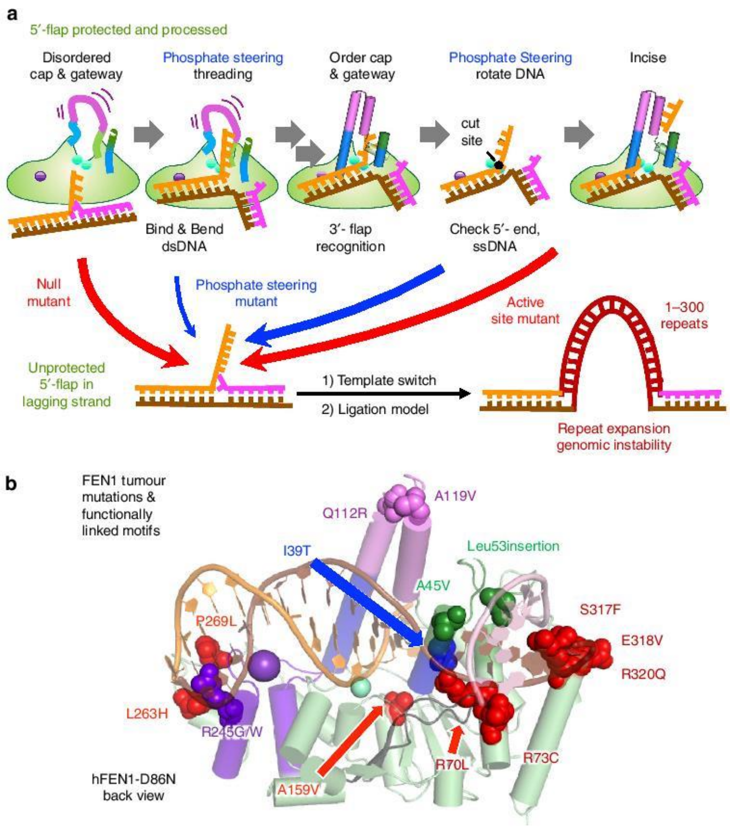


Figure 7 | Multiple motifs for FEN1 substrate recognition and hydrolysis ensure accurate incision activity and prevent genomic instability. (**a**) Schematic model of the FEN1 mechanism emphasizing the functional role of phosphate steering in the dynamic processes of 5′-flap inverted threading and shifting 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′-ggcggcttgccattaaagacatacacgggctt-3′ and 5′-aagcccgtgtatgtctttaatggcaagccgcc-3′; R103A, 5′-caaacgcagtgaggcgcgggctgaggca-3′ and 5′-tgcctcagcccgcgcctcactgcgtttg-3′; R103E, 5′-ccaaacgcagtgaggagcgggctgaggcag-3′ and 5′-ctgcctcagcccgctcctcactgcgtttgg-3′; R104A, 5′-ctctgcctcagccgcccgctcactgcgt-3′ and 5′-acgcagtgagcgggcggctgaggcagag-3′; R104E, 5′-acgcagtgagcgggaggctgaggcagag-3′ and 5′-ctctgcctcagcctcccgctcactgcgt-3′; R129A, 5′-ttagtgaccttcaccagcgccttagtgaatttttccacctc-3′ and 5′-gaggtggaaaaattcactaaggcgctggtgaaggtcactaa-3′; R129E, 5′-ttagtgaccttcaccagctccttagtgaatttttccacctc-3′ and 5′-gaggtggaaaaattcactaaggagctggtgaaggtcactaa-3′; K132A, 5′-cactaagcggctggtggcggtcactaagcagcac-3′ and 5′-gtgctgcttagtgaccgccaccagccgcttagtg-3′; K132E, 5′-gctgcttagtgacctccaccagccgcttagt-3′ and 5′-actaagcggctggtggaggtcactaagcagc-3′; R103ER104E, 5′-gcttctctgcctcagcctcctcctcactgcgtttggcca-3′ and 5′-tggccaaacgcagtgaggaggaggctgaggcagagaagc-3′; R129EK132E, 5′-gtgctgcttagtgacctccaccagctccttagtgaatttttccacc-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 × 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 SIGMA*FAST* 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 × 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 × 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% saturated 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 $(\text{ratio})_A$ 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.

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

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

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

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

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 with 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 (GraphPad Software, 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 $\mu g\,ml^{-1}$ 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′-CTCGATGTGCAGAACCTGAAGCTTGATCT-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 $\mu g\,ml^{-1}$ 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 (OD_{600} 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).

Acknowledgements

For financial support, we thank NCI (PO1CA92584 to J.A.T.); BBSRC (grants BB/K009079/1 and BB/M00404X/1 to J.A.G.); the Ministry of Higher Education Libya, EPSRC and the University of Sheffield (studentships to S.I.A., S.J.S. and E.J., respectively); NIH (R01GM060987 to S.M. and R01GM110387 to S.T.); and KAUST for core and CRG3 funding to S.M.H. and J.A.T. J.A.T. acknowledges support by a Robert A. Welch Chemistry Chair, the Cancer Prevention and Research Institute of Texas, and the University of Texas System Science and Technology Acquisition and Retention. The tumour mutation analysis includes data generated by the TCGA Research Network: http://cancergenome.nih.gov/. We thank the ALS, SSRL, the IDAT program, the DOE BER and the NIH project MINOS (R01GM105404) for X-ray data facilities. We thank James Holton for aiding X-ray data analysis.

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).

Biochemistry
Communication
pubs.acs.org/biochemistry

Human Exonuclease 1 Threads 5′-Flap Substrates through Its Helical Arch

Steven J. Shaw, L. David Finger, and Jane A. Grasby*

Centre for Chemical Biology, Department of Chemistry, Krebs Institute, University of Sheffield, Sheffield S3 7HF, U.K.

Supporting Information

ABSTRACT: Human exonuclease 1 (hEXO1) is a member of the 5′-nuclease superfamily and plays important roles in DNA repair. Along with acting as a 5′-exonuclease on blunt, gapped, nicked, and 3′-overhang DNAs, hEXO1 can also act as an endonuclease removing protruding 5′-single-stranded flaps from duplex ends. How hEXO1 and related 5′-nuclease human flap endonuclease 1 (hFEN1) are specific for discontinuous DNA substrates like 5′-flaps has been controversial. Here we report the first functional data that imply that hEXO1 threads the 5′-flap through a hole in the protein known as the helical arch, thereby excluding reactions of continuous single strands. Conjugation of bulky 5′-streptavidin that would "block" threading through the arch drastically slowed the hEXO1 reaction. In contrast, addition of streptavidin to a preformed hEXO1 5′-biotin flap DNA complex trapped a portion of the substrate in a highly reactive threaded conformation. However, another fraction behaves as if it were "blocked" and decayed very slowly, implying there were both threaded and unthreaded forms of the substrate present. The reaction of an unmodified hEXO1−flap DNA complex did not exhibit marked biphasic kinetics, suggesting a fast re-equilibration occurs that produces more threaded substrate when some decays. The finding that a threading mechanism like that used by hFEN1 is also used by hEXO1 unifies the mode of operation for members of the 5′-nuclease superfamily that act on discontinuous substrates. As with hFEN1, intrinsic disorder of the arch region of the protein may explain how flaps can be threaded without a need for a coupled energy source.

Human exonuclease 1 (hEXO1) is a member of the divalent metal ion-dependent 5′-nuclease superfamily whose enzymes process a wide variety of bifurcated nucleic acid structures.[1] Playing important roles in mismatch and double-strand break repair, hEXO1 catalyzes the processive exonucleolytic hydrolysis of DNA.[2] Substrates include nicked, gapped, 3′-overhang (Figure S1A), and blunt-ended DNA duplexes, which are degraded on the 5′-strand producing single-nucleotide products. In addition, hEXO1 also acts as an endonuclease removing 5′-single-stranded (ss) flaps from 5′-overhang or pseudo-Y (psY) DNAs (Figure S1B,C), in a manner analogous to that of its paralogue flap endonuclease 1 (FEN1). Indeed, this flap removal capability of hEXO1 has been suggested to initiate resection in double-strand break repair.[3]

Biochemical and more recent structural studies of FEN1 have shown that the 5′-flap portion of the substrate threads through a hole in the protein known as the helical arch (Figure 1A−C).[4−6] The helical arch of FEN1 is formed from two α helices ($\alpha4-\alpha5$). In turn, these helices can be divided into two regions, the first part of the "gateway" (base of $\alpha4$), which provides residues to the active site and is a universal feature of the 5′-nuclease superfamily, and the helical cap (rest of $\alpha4-\alpha5$).[7] Unlike the gateway, the cap is not present in superfamily members that do not enforce a specificity for discontinuous DNAs like GEN1 and XPG but is a feature exclusive to FEN1 and EXO1 (Figure 1D−F), yet the current thinking about the EXO1 mechanism differs from that of FEN1 and suggests that the cap region of the protein acts as a substrate clamp, rejecting the threading mechanism for substrate specificity.[8]

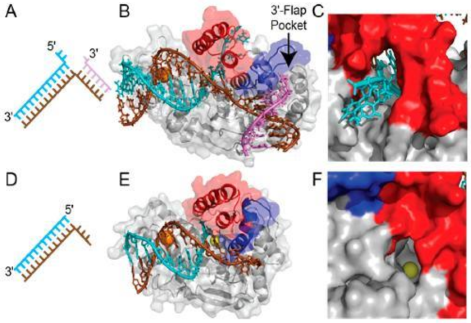


Figure 1. Similarities between 5′-nuclease superfamily members hEXO1−352 and hFEN1−336 in complex with DNA. (A) Double-flap and (D) 3′-overhang substrate DNAs crystallized with (B) hFEN1−336 (Protein Data Bank entry 5UM9) and (E) hEXO1−352 (Protein Data Bank entry 3QE9). The helical arch ($\alpha4-\alpha5$) is colored red, and the $\alpha2-\alpha3$ portion is colored blue. (C) The exit site of the 5′-flap through the back of the arch is shown in close-up for hFEN1, with the potential exit site of a flap from hEXO1 also illustrated (F).

To understand the mechanism of flap removal by hEXO1, we initiated a comparative study with hFEN1. We used a version of

Received: May 25, 2017
Revised: July 4, 2017
Published: July 6, 2017

DOI: 10.1021/acs.biochem.7b00507
Biochemistry XXXX, XXX, XXX−XXX

the hEXO1 protein truncated at residue 352 (termed hEXO1−352), which is sufficient for robust nuclease activity but lacks the disordered C-terminal tail that is the site of protein−protein interaction *in vivo*.[8] hEXO1−352 was crystallized with DNA previously, and the structure closely resembled that of hFEN1 except that it lacks the 3′-flap binding pocket of FEN1 (Figure 1B,E).[7,8] The 3′-flap pocket of FEN1 locks the DNA substrate into a particular orientation so that reaction of a double-flap substrate occurs at only one site, one nucleotide into double-stranded (ds) DNA.[9] One nucleotide into ds DNA was also the preferred reaction site of hEXO1−352 on a 3′-overhang DNA substrate (Figure 2A and Figure S1; all experimental details are available in the Supporting Information). In contrast, removal of the ss flap from a psY substrate by hEXO1−352 produced two major products that corresponded to reaction one nucleotide into the ss flap and at the ss−ds junction with minimal reaction within the ds DNA (Figure 2B and Figure S1). Nevertheless, because the ds DNA products of EXO1 action are also substrates until they reach a minimal duplex length, eventual reaction within the dsDNA will occur.

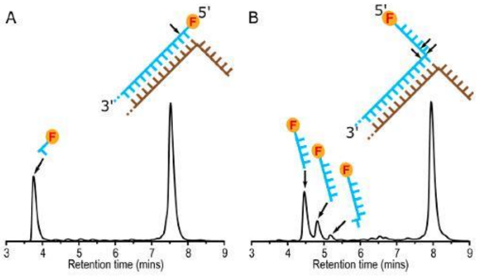


Figure 2. Products of hEXO1−352-catalyzed reactions as determined by high-performance liquid chromatography. Chromatograms of reaction of hEXO1−352 on (A) 3′-overhang and (B) psY substrates bearing 5′-fluorescein labels that produces (A) a single-nucleotide product and (B) three ss products corresponding to reaction one nucleotide into the duplex, at the ss−ds junction, and one nucleotide into the flap (six, five, and four nucleotides, respectively). Schematic representations of fluorescein-labeled products and substrates are shown.

To test the mode of interaction of hEXO1 with 5′-flapped DNAs, we used a 5′-biotinylated substrate that could form a very stable interaction with the 211.2 kDa streptavidin tetramer. The substrate was formed from a single 5′-fluorescent and biotinylated oligonucleotide that adopted a psY shape, abbreviated Bio-psY (Figure 3A and Figure S1). As all discontinuous 5′-ends can potentially be sites of reaction with hEXO1, this unimolecular substrate construction ensured only one binding orientation on hEXO1.

Three different types of complexes were constructed with Bio-psY (Figure 3A) and hEXO1 (Figure 3B; for materials and methods, see the Supporting Information). In the first complex, termed "pre-mixed", the biotinylated substrate and hEXO1 were preincubated together in the presence of the catalytically inert cofactor Ca^{2+} ions. Excess streptavidin was added to some of this premixed complex. If the 5′-ss DNA flap was threaded underneath the helical cap of hEXO1, the streptavidin would trap the DNA on the enzyme; therefore, this complex was termed "trapped". In the final complex, excess streptavidin was added to Bio-psY prior to its interaction with the enzyme in the presence of Ca^{2+} ions. The bulky streptavidin is large enough to prevent the 5′-ss flap from threading through the helical arch; thus, this mixture is termed "blocked". In contrast, if the arch instead acts as a clamp, then the 5′-streptavidin should not interfere with reaction.

To investigate which of the complexes was capable of fast hEXO1-catalyzed hydrolysis and therefore which complex was positioned correctly for reaction, catalysis was initiated by addition of excess Mg^{2+} ions. Reactions were monitored on the millisecond time scale using quenched flow rapid handling apparatus followed by analysis of samples using denaturing high-performance liquid chromatography with a fluorescence detector (Figure 3C). The premixed complex decayed at a rate of 0.43 s^{-1} (Table 1; note that the 1 mM catalytically inert Ca^{2+} ions added to form the complex are inhibitory to some extent in the presence of 8 mM Mg^{2+} ions). In contrast, the streptavidin "blocked" mixture decayed at a rate that was ∼240 times slower. Because prior streptavidin conjugation would not be expected to impede a clamping mechanism, this implies that a threaded structure is required for reaction on a biologically relevant time scale. The results closely mirror the properties of hFEN1 with similar 5′-streptavidin substrates.[5]

The most interesting behavior was observed with the hEXO1 streptavidin "trapped" substrate complex, which exhibited distinctly biphasic kinetics when reaction was initiated. A portion (∼20%) of this complex decayed at a rate of 2.05 s^{-1}, comparable to that of the "pre-mixed" complex (Figure 3C and Table 1). The other fraction decayed much more slowly with a rate that was similar to that of the "blocked" complex. These results have some resemblance to, but some differences from, those of an earlier streptavidin "trapping" experiment with hFEN1, in which close to 100% of the substrate decayed with a very fast rate constant.[5]

We interpreted the biphasic behavior exhibited by the hEXO1 "trapped" mixture as arising from one portion of the substrate being constrained by streptavidin in a position that is capable of fast reaction, with the 5′-ss flap threaded through the helical arch. The other fraction of the substrate had not adopted a threaded state, and when streptavidin was added, it became "blocked".

An important question was whether the substrate was saturated with hEXO1 in our experiment, because this could have been what prevented all the substrate from adopting a threaded conformation and being capable of fast reaction. To test this, the concentration of hEXO1 was increased and the amount of fast decaying species that arose in the streptavidin "trapping" experiment was determined (Figure 3D). Despite the change in the concentration of the enzyme, the amount of trapped fast-decaying substrate remained constant at the tested protein concentrations. This implied that the partition between threaded (fast-decaying) and unthreaded (slow-decaying) portions in the trapping experiment reflects an on-enzyme equilibrium.

The picture that emerges from our experiments is that EXO1, like FEN1, uses a threading mechanism and that, when present, passing the 5′-ss flaps of substrates underneath the helical cap is a prerequisite for fast reaction. However, unlike FEN1, which processes its optimal double-flap substrates at maximal diffusion-controlled rates,[7] hEXO1 is not evolutionarily optimized to thread 5′-flap substrates. Thus, even when flap substrates were fully saturated with enzyme, an on-enzyme partition between the threaded and unthreaded substrate

DOI: 10.1021/acs.biochem.7b00507

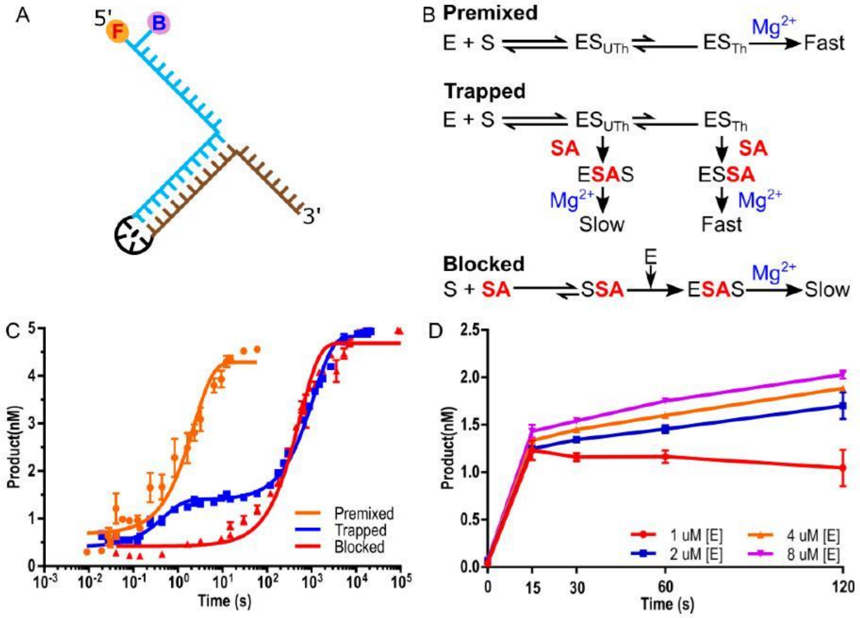


Figure 3. Streptavidin trapping and blocking experiments of EXO1 substrates reveal a role for threading in the hEXO1−352-catalyzed endonucleolytic reaction. (A) Illustration of the unimolecular pseudo-Y (Bio-psY) substrate with a 5′-fluorescein (F) label and a serinol-biotin TEG (B) used in the trapping experiments. (B) Reaction scheme for the formation of "premixed", "trapped", and "blocked" complexes involving enzyme hEXO1 (E), Bio-psY (S), and streptavidin (SA); UTh denotes unthreaded and Th threaded. (C) Decay of "pre-mixed" and streptavidin "trapped" and "blocked" complexes of Bio-psY with hEXO1 upon addition of Mg^{2+} ions. Rate constants from fits to single-exponential ("pre-mixed" and "blocked") or double-exponential ("trapped") decays are listed in Table 1. (D) Lack of variation in the proportion of fast-decaying species with concentrations of hEXO1 (1−8 μM) in a streptavidin trapping experiment.

Table 1. Single-Turnover Rates of the Various Bio-psY Complexes with hEXO1−352

state	k_{ST} (s^{-1})	amplitude
premixed	0.43 ± 0.05	100
trapped		
fast	2.05 ± 0.89	22.3
slow	0.0009 ± 0.00004	77.7
blocked	0.0018 ± 0.00010	100

favored the unthreaded noncatalytically proficient state. Nevertheless, the "pre-mixed" sample does not decay with the markedly fast (threaded) and slow (unthreaded) behavior of the "trapped" mixture. This appears to rule out the possibility that unmodified reactions of hEXO1 typically occur in both threaded and nonthreaded states (assuming the nonthreaded rate of reaction of an unmodified substrate is similar to that of a "blocked" one). The lack of a biphasic premixed profile would be expected if re-equilibration producing more threaded material occurs when some reacts, so long as the rate of re-equilibration is faster than, or similar to, the rate of reaction. Accordingly, our data are consistent with a model in which all flapped substrates typically react from a threaded state.

The threading mechanism was initially (sensibly) rejected for EXO1 on the basis of the perceived difficulties of passing DNA through a small hole in a protein without a coupled energy source.[8] A possible reconciliation between this and the data reported here may be provided by proposals for the mode of action of other 5′-nucleases in which regions of intrinsic disorder are suggested to be essential for enzyme function.[5,7,10] Passing DNA through a disordered arch region, which could then form secondary structure to permit reaction, would be expected to have a much less significant entropic barrier. Indeed, crystallographic *B*-factors for backbone amide nitrogens in structures of hEXO1 bound to DNA show large variations (Figure S2).[8] Residues in the arch and between $\alpha2$ and $\alpha3$ (Figure 1) possess very large (100−250) *B*-factors, implying a high degree of motion and lending weight to the plausibility of our disorder−thread−order proposal. Differences in the partition between the threaded and unthreaded state observed for hEXO1 compared to hFEN1 may reflect the tendency of the arch regions of the respective proteins to form secondary structure, although this requires further investigation.

The unified 5′-nuclease superfamily threading mechanism has important implications for EXO1 and FEN1 function during DNA replication and repair. The need to thread the 5′-portion of substrates through a hole in the protein confines reactions to discontinuous DNA ends that possess free 5′-termini. However, in contrast, reactions of continuous DNAs, such as the unbroken strand in a gapped DNA substrate, are prevented on a biological time scale by the threading mechanism. Thus, threading, likely together with intrinsic disorder, plays a vital role in 5′-nuclease reaction specificity.

DOI: 10.1021/acs.biochem.7b00507
Biochemistry XXXX, XXX, XXX−XXX

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.biochem.7b00507.

> Full details of materials and methods, oligonucleotide sequences and constructs (Table S1), DNA constructs used (Figure S1), and a plot of *B*-factors for hEXO1−352 (Figure S2) (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*Centre for Chemical Biology, Department of Chemistry, Krebs Institute, University of Sheffield, Sheffield S3 7HF, U.K. E-mail: j.a.grasby@sheffield.ac.uk.

ORCID

Jane A. Grasby: 0000-0003-3799-4136

Funding

This work was supported by an EPSRC studentship to S.J.S. and BBSRC Grant BB/M00404X/1 to J.A.G.

Notes

The authors declare no competing financial interest.

During the review of this work, a structure of hEXO1 with a threaded flap substrate was published.[11]

■ ABBREVIATIONS

ds, double-stranded; hEXO1, human exonuclease 1; hFEN1, human flap endonuclease 1; ss, single-stranded; psY, pseudo-Y (DNA structure).

■ REFERENCES

(1) Grasby, J. A., Finger, L. D., Tsutakawa, S. E., Atack, J. M., and Tainer, J. A. (2012) *Trends Biochem. Sci. 37*, 74−84.

(2) Keijzers, G., Liu, D., and Rasmussen, L. J. (2016) *Crit. Rev. Biochem. Mol. Biol. 51*, 440−451.

(3) Yang, S.-H., Zhou, R., Campbell, J., Chen, J., Ha, T., and Paull, T. T. (2013) *EMBO J. 32*, 126−139.

(4) Gloor, J. W., Balakrishnan, L., and Bambara, R. A. (2010) *J. Biol. Chem. 285*, 34922−34931.

(5) 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) *Nucleic Acids Res. 40*, 4507−4519.

(6) 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. S., Her, M. Z., Rashid, F., Hamdan, S. M., Mirkin, S. M., Grasby, J. A., and Tainer, J. A. (2017) *Nat. Commun. 8*, 15855.

(7) 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) *Cell 145*, 198−211.

(8) Orans, J., McSweeney, E. A., Iyer, R. R., Hast, M. A., Hellinga, H. W., Modrich, P., and Beese, L. S. (2011) *Cell 145*, 212−223.

(9) Finger, L. D., Blanchard, M. S., Theimer, C. A., Sengerova, B., Singh, P., Chavez, V., Liu, F., Grasby, J. A., and Shen, B. (2009) *J. Biol. Chem. 284*, 22184−22194.

(10) AlMalki, F. A., Flemming, C. S., Zhang, J., Feng, M., Sedelnikova, S. E., Ceska, T., Rafferty, J. B., Sayers, J. R., and Artymiuk, P. J. (2016) *Nat. Struct. Mol. Biol. 23*, 640−646.

(11) Shi, Y., Hellinga, H. W., and Beese, L. S. (2017) *Proc. Natl. Acad. Sci. U. S. A. 114*, 6010−6015.